Journal of Medical Internet Research

Journal Impact Factor (JIF) (2022): 7.4
Volume 17 (2015), Issue 9 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Letters to the Editor

Original Paper

Adherence to Technology-Mediated Insomnia Treatment: A Meta-Analysis, Interviews, and Focus Groups

Corine Horsch[1], MSc; Jaap Lancee[2], PhD; Robbert Jan Beun[3], PhD; Mark A Neerincx[1], PhD; Willem-Paul Brinkman[1], PhD

[1]Interactive Intelligence, Delft University of Technology, Delft, Netherlands

[2]Department of Clinical Psychology, University of Amsterdam, Amsterdam, Netherlands

[3]Interaction Technology, Utrecht University, Utrecht, Netherlands

Corresponding Author:
Corine Horsch, MSc
Interactive Intelligence
Delft University of Technology
EWI HB, 12th Floor
Mekelweg 4
Delft, 2628 CD
Netherlands
Phone: 31 152784145
Fax: 31 152787141
Email: c.h.g.horsch@tudelft.nl

Abstract

Background: Several technologies have been proposed to support the reduction of insomnia complaints. A user-centered assessment of these technologies could provide insight into underlying factors related to treatment adherence.

Objective: Gaining insight into adherence to technology-mediated insomnia treatment as a solid base for improving those adherence rates by applying adherence-enhancing strategies.

Methods: Adherence to technology-mediated sleep products was studied in three ways. First, a meta-analysis was performed to investigate adherence rates in technology-mediated insomnia therapy. Several databases were queried for technology-mediated insomnia treatments. After inclusion and exclusion steps, data from 18 studies were retrieved and aggregated to find an average adherence rate. Next, 15 semistructured interviews about sleep-support technologies were conducted to investigate perceived adherence. Lastly, several scenarios were written about the usage of a virtual sleep coach that could support adherence rates. The scenarios were discussed in six different focus groups consisting of potential users (n=15), sleep experts (n=7), and coaches (n=9).

Results: From the meta-analysis, average treatment adherence appeared to be approximately 52% (95% CI 43%-61%) for technology-mediated insomnia treatments. This means that, on average, half of the treatment exercises were not executed, suggesting there is a substantial need for adherence and room for improvement in this area. However, the users in the interviews believed they adhered quite well to their sleep products. Users mentioned relying on personal commitment (ie, willpower) for therapy adherence. Participants of the focus groups reconfirmed their belief in the effectiveness of personal commitment, which they regarded as more effective than adherence-enhancing strategies.

Conclusions: Although adherence rates for insomnia interventions indicate extensive room for improvement, users might not consider adherence to be a problem; they believe willpower to be an effective adherence strategy. A virtual coach should be able to cope with this “adherence bias” and persuade users to accept adherence-enhancing strategies, such as reminders, compliments, and community building.

(J Med Internet Res 2015;17(9):e214) doi:10.2196/jmir.4115

KEYWORDS

sleep initiation and maintenance disorders; patient compliance; meta-analysis; interview; focus groups; mobile apps; user-computer interface

Introduction

Overview

People who suffer from insomnia have difficulties with initiating sleep, maintaining sleep, or early-morning awakenings, and this sleep disturbance significantly impairs their daily functioning [1]. Having insomnia may lead to personal suffering, such as feeling tired after a night's sleep, reduced quality of life, and vulnerability to depression [2,3]. In addition, insomnia leads to societal costs that might include reduced productivity and more sick leave from work [2,4]. A review of the literature showed that about 9% to 15% of the western adult population suffers from insomnia symptoms and the daytime consequences thereof [5].

Although the consequences of insomnia may be severe and prevalence is substantial, only a few people seek treatment [6-8]. When help is sought, insomnia is most commonly treated with pharmacotherapy [7]. However, cognitive behavioral therapy for insomnia (CBT-I) is preferable, because CBT-I is equally effective in the short term and has more beneficial long-term effects than pharmacotherapy [9-11]. Generally, CBT-I consists of weekly sessions in which the focus lies on one or more of the following exercises: sleep restriction, stimulus control, relaxation, cognitive strategies, and sleep hygiene [12].

Although CBT-I is effective, there is a lack of knowledge and accessibility regarding this type of therapy [13]. General practitioners are often not aware of the existence of CBT-I, and neither is the general public [13]. In addition, there are too few sleep therapists to help all people with insomnia [14]. In order to increase the availability and accessibility of CBT-I, Espie et al [15] suggested a stepped model with Internet-based treatment as a first option. A meta-analysis about computerized CBT-I (CCBT-I) concluded that this therapy is a moderately effective self-help intervention for insomnia [16]. Nonetheless, adherence to insomnia and other technology-mediated treatments is often mentioned as a serious problem [17-19].

The World Health Organization (WHO) recognizes the importance of adherence to health regimes in general. They stated, "Adherence is a primary determinant of the effectiveness of treatment" [20]. In agreement with the WHO statement, Gould and Clum [21] found—in their meta-analysis of self-help treatments—that better adherence to a treatment improves the treatment effectiveness. They found that the effect size was three times higher for studies that had 75% to 100% adherence than for studies with adherence rates lower than 75%. The impact of adherence on treatment outcomes therefore warrants further investigation into how we could enhance adherence within an intervention in the context of insomnia therapy.

Various authors, for example, Beun [17] and Donkin et al [18], mention that treatment adherence is a problem for cognitive behavioral therapy (CBT) in general. Reports about adherence to various Internet-based interventions show mixed results. For example, Eysenbach [19] gives a few examples in his "law of attrition" of Internet-based interventions with adherence rates ranging from 1% to 35%. Interestingly, a meta-analysis about CCBT-I reported an average adherence rate of 78% for the six studies they included [16]. However, they did not make a distinction between treatment adherence and experimental compliance, that is, the proportion of the experimental assessments, such as questionnaires, that are completed. Thus, decisive conclusions on the exact adherence rates cannot be made.

The studies in this paper are conducted in the context of the Sleepcare project [22,23], which aims at the development of a virtual sleep coach that delivers personalized, automated sleep therapy via a mobile phone. A key challenge of this e-coach is to provide therapy support in such a way that the coachees really adhere to the regimen of the personal therapy plan. In this paper, we use the generic term coachee—instead of client, patient, user, etc—to refer to both patients and nonpatients who seek help to address their health issues. The first step in the development of a virtual sleep coach that meets this adherence challenge is the analysis of current adherence rates, current adherence-enhancing strategies, and coachees' willingness to accept those strategies. Therefore, we conducted a meta-analysis about adherence rates in technology-mediated sleep interventions; interviewed coachees about their adherence to existing sleep-supporting technology; and discussed adherence-enhancing strategies in a to-be-developed virtual sleep coach among focus groups with potential users, sleep experts, and coaches. This complementary analysis approach provided new insights on how a virtual coach can support coachees to adhere to sleep therapy (ie, the needs and constraints).

Study I: Meta-Analysis Adherence Rates

In order to determine whether a certain outcome is related to a treatment, adherence rates must be measured. Otherwise, it cannot be claimed that the outcome was caused by the intervention [21]. Capturing adherence data is relatively easy in technology-mediated interventions [18]. However, as there is currently no standard adherence measure [18,24,25], various measures are used. A review [18] of adherence in e-therapies found the following adherence measures: number of log-ins, completed modules, number of visits/posts to a forum, pages viewed/printed, and self-reported measures. Other measures that have been suggested are the usage time of the technology [26] and reports by a spouse or related others [24]. Different measures have different advantages and disadvantages. For example, time spent using the technology is an objective measure. However, time spent is presumably influenced by cognitive ability, reading speed, familiarity with the technology, etc [18]. Therefore, time spent does not necessarily represent treatment adherence. Moreover, there is a difference in passively using material (ie, reading, listening, watching) and actively applying this material (ie, performing the exercises) [21].

First, it is important to distinguish between at least two concepts: *treatment adherence* and *experiment compliance*. Treatment adherence refers to the extent a coachee processes and applies the content of the treatment (as provided by the coach), whereas experiment compliance refers to the coachees' completion of the experimental assessments. Other researchers have also made this distinction. For example, Christensen and colleagues [26] respectively use the terms adherence (experience content) and

dropout (research trial protocol), whereas Hebert and colleagues [27] respectively call it nonusage attrition and study attrition. Treatment adherence and experiment compliance might be related, but to our knowledge no information about this relationship has been reported in the literature.

Study II: User Adherence to Existing Sleep-Supporting Technology

After analyzing reported adherence rates to technology-mediated sleep treatment in the literature, the next step was to study coachees' reasons why they do or do not adhere to technology-mediated sleep interventions. To do so, interviews were conducted with people who (had) used a sleep product. The first step was to identify a sample of technology-mediated sleep products. The most familiar sleep product is probably the alarm clock. Besides alarms, there are many other sleep-supporting technologies on the market. For example, relaxation-supporting technologies, sleep-measuring apps and devices, and computerized therapies.

Study III: Focus Group Discussions—The Envisioned Sleep Coach

A limitation of the interviews from Study II, as will be discussed in more detail in the Results section, was that they were restricted to existing products, and did not include reflections on what might technically be possible regarding adherence-enhancing strategies. During the interviews, it also proved to be difficult for participants to think of additional functionality that could improve their adherence. To address the limitations of the interviews, focus groups were organized to discuss adherence-enhancing strategies of a to-be-developed sleep coach. The aim of study III was to gain insight into coachees' attitudes and beliefs toward these adherence-enhancing strategies, for which focus groups are particularly suited [28].

Methods

Study I

Overview

The meta-analysis was primarily performed to answer the question "How well do coachees adhere to technology-mediated insomnia interventions and diagnostic tools?" and, secondly, to answer the question "How does adherence relate to treatment outcome?" Various databases were queried—Web of Science, Scopus, PubMed, and PsychINFO—on July 8 and 14, 2014, to find studies that investigate insomnia regimes mediated by technology. The used query was: *insomnia* and *Internet-treatment*, *Internet-delivered*, *Internet-based*, *Internet-administered*, *Internet intervention*, *computerize*, *online treatment*, *Web application*, *Web-based*, *virtual*, *virtual reality*, *mass media intervention*, *smartphone*, *mobile phone*, *mobile technology*, *text message*, *handheld*, or *PDA* (personal digital assistant). In addition, the references from recent meta-analyses, and systematic reviews on self-help and computerized insomnia therapy [16,29,30] were screened for potentially relevant publications. Together, this resulted in 448 unique papers of which the abstracts were read and examined (by the first author, CH) for meeting the following exclusion criteria: no main focus on insomnia, no technology involved, treatment that does not include assignments at home, no experiment, or targeted at children. Studies on children were excluded because children's sleep problems often differ from those of adults. Besides, children's bedtimes are partly controlled by the parents. Therefore, interventions targeted at children have other characteristics than interventions for adults and were excluded. A total of 56 papers were read completely and the inclusion of those papers was discussed between the first and second author (CH and JL).

Figure 1 shows the flow diagram for inclusion and exclusion criteria, resulting in 21 papers from which data was retrieved. Due to a lack of reported adherence data in 3 of the papers, only 18 papers were used in the analysis. The papers selected for this meta-analysis can be found in Multimedia Appendix 1.

Description of Included Studies

Of the 18 included studies in this meta-analysis, 12 studies (67%) focused on CBT-I (Table 1 and Table 2). Oosterhuis and Klip [31] and Rybarczyk and colleagues [32] did include most of the CBT-I exercises in their intervention. Out of the 18 studies, 2 (11%) focused on sleep tracking by using an active sleep sampling device. The active sleep sampling device used by Riley and colleagues [33] mainly supported sleep restriction and stimulus control. The other standard CBT-I exercises were explained in an additional manual. Lawson and colleagues [34] used an active sleep sampling device inspired by Riley's device. They developed an active sleep sampling mobile phone app which focused on sleep tracking, but did not include the other CBT-I components. Lipschitz and colleagues [35] also developed a mobile phone app, offering sleep-focused, mind-body bridging exercises. The most important assumption of mind-body bridging for sleep is that the mind needs to be rested to sleep well. Haimov and Shatil [36] studied whether providing cognitive training, such as a memory game, affects sleep.

Table 1. Characteristics of included studies.

First author	Condition	Number of people	Number of females/males	Mean age	Sleep problem severity, measure, score
Oosterhuis [31]	Intervention	400	63% female	55	N/A[a]
Rybarczak [32]	Intervention	14	22/16	68	PSQI[b], 9.5
	CBT[c]	11			PSQI, 11.9
	Control	13			PSQI, 9.9
Ström [37]	Intervention	54	71/38	44	ISI[d], 18.08
	Waiting list	55			ISI, 18.11
Suzuki [38]	Intervention	21	16/25	40	N/A
	Waiting list	22			
Ritterband [39]	Intervention	22	34/10	N/A	ISI, ≥8
	Waiting list	23			ISI, ≥8
Van Straten [40]	Intervention	126	163/84	52	72% rated SQ[e]<6/10
	Waiting list	121			68% rated SQ<6/10
Vincent [41]	Intervention	59	79/39	N/A	N/A
	Waiting list	59			
Riley [33]	Intervention 1	24	52/38	49	ISI, 8-14 (25 people)
	Intervention 2	33			ISI, 15-21 (53 people)
	SMMT[f]	33			ISI, 22-28 (12 people)
Lancee [42]	CCBT-I[g]	216	520/103	52	Sleep-50, ≥19
	CBT-I[h]	202			
	Waiting list	205			
Ritterband [43]	Intervention	14	24/4	57	ISI, 17.1
	Waiting list	14			ISI, 15.9
Espie [15]	Intervention	55	120/44	49	Met DSM-5[i] criteria
	TAU[j]	54			Met DSM-5 criteria
	IRT[k]	55			Met DSM-5 criteria
Haimov [36]	Cognitive training (CogniFit)	34	29/22	72	Met AASM[l] criteria
	Active control[m]	17			Met AASM criteria
Lancee [44]	Low depression	198	316/163	47	ISI, 16.73
	Mild depression	182			ISI, 18.63
	High depression	99			ISI, 20.69
					Average ISI, 18.72
Lancee [45]	With support	129	197/65	48	ISI, 16.95
	Without support	133			ISI, 17.32
Lawson [34]	Intervention	36	21/5	34	N/A
Van Straten [46]	Intervention	59	83/35	49	PSQI, 12.4
	Waiting list	59			PSQI, 11.7
Holmqvist [47]	Intervention	39	55/18	N/A	ISI, 18.72
	CBT-I	34			ISI, 18.50
Lipschitz [35]	Intervention	37	27/10	37	ISI, 7.24

[a]Not applicable (N/A)
[b]Pittsburgh Sleep Quality Index (PSQI)
[c]Cognitive behavioral therapy (CBT)
[d]Insomnia Severity Index (ISI)
[e]Sleep quality (SQ)
[f]Self-monitoring minimal treatment (SMMT)
[g]Computerized cognitive behavioral therapy for insomnia (CCBT-I)
[h]Cognitive behavioral therapy for insomnia (CBT-I)
[i]Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5)
[j]Treatment as usual (TAU)
[k]Imagery relief therapy (placebo) (IRT)
[l]American Academy of Sleep Medicine (AASM)
[m]Active control consisted of word and paint training

Figure 1. Inclusion and exclusion criteria for papers in the meta-analysis.

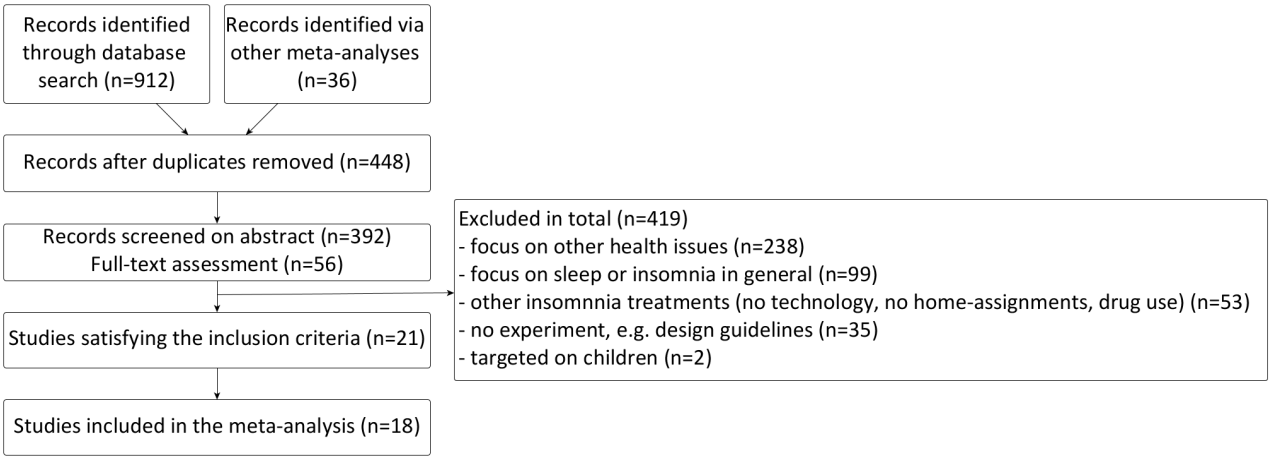


Table 2. Description of included studies.

First author	Intervention	Delivery	Treatment length	Follow-up length	Post[a]	Follow-up[b]	Adherence measure
Oosterhuis [31]	SE[c], SH[d], CTh[e], RX[f]	TV[g]	8 weeks	4.5 months	Q[h]	Q	N/A[i]
Rybarczyk [32]	RX, SC[j], SR[k], CTh, SH	Audiotape	6 weeks	4 months	Q&D[l]	Q&D	N/A
Ström [37]	CBT-I[m]	Internet	5 weeks	9 months	Q&D	D	N/A
Suzuki [38]	CBT-I	Internet	2 weeks	3 weeks	Q	Q	N/A
Ritterband [39]	CBT-I	Internet	9 weeks	6 months	Q&D	Q	N/A
Van Straten [40]	CBT-I	TV	6 weeks	None	Q&D	N/A	Self-report
Vincent [41]	CBT-I	Internet	5 weeks	4 weeks	Q&D	Q&D	Self-report
Riley [33]	ASS[n]/CBT-I	Device	6 weeks	6 weeks	Q&D	Q&D	N/A
Lancee [42]	CBT-I	Internet	6 weeks	4 weeks		Q&D	Self-report
Ritterband [43]	CBT-I	Internet	6-9 weeks	None	Q&D	N/A	Log
Espie [15]	CBT-I	Internet	6 weeks	8 weeks	N/A	Q&D	Log
Haimov [36]	CTr[o]	PC[p]	8 weeks	None	Q&D	N/A	N/A
Lancee [44]	CBT-I	Internet	6 weeks	4 weeks	N/A	Q&D	Self-report
Lancee [45]	CBT-I	Internet	6 weeks	6 months	Q&D	Q&D	Log
Lawson [34]	ASS	App[q]	7 days	None	Q	N/A	Log
Van Straten [46]	CBT-I	Internet	6 weeks	3 months	Q&D	Q&D	Log
Holmqvist [47]	CBT-I	Internet	6 weeks	8 weeks	Q&D	Q&D	N/A
Lipschitz [35]	MBB[r]	Internet	3 days	1 week	Q	Q	Self-report

[a]Postintervention measurement instrument
[b]Follow-up measurement instrument
[c]Sleep education (SE)
[d]Sleep hygiene (SH)
[e]Cognitive therapy (CTh)
[f]Relaxation (RX)
[g]Television (TV)
[h]Questionnaire (Q)
[i]Not applicable (N/A)
[j]Stimulus control (SC)
[k]Sleep restriction (SR)
[l]Sleep diary (D)
[m]Cognitive behavioural therapy for insomnia (CBT-I)
[n]Active sleep sampling (ASS) device
[o]Cognitive training (CTr) (CogniFit)
[p]Personal computer (PC)
[q]Mobile phone app (app)
[r]Mind-body bridging (MBB)

Study II

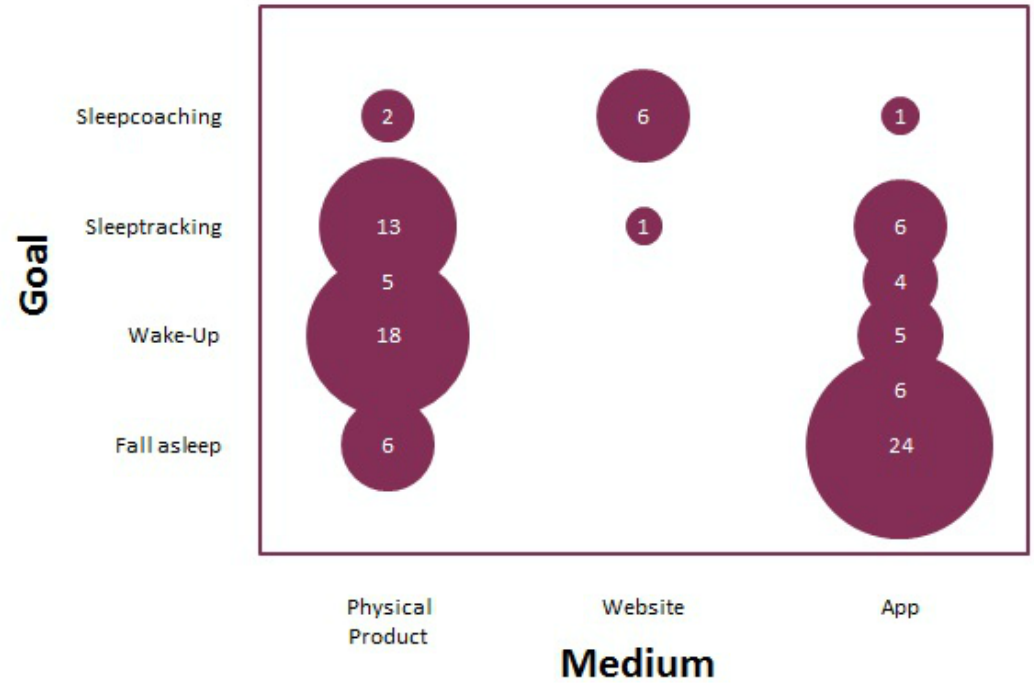
Participant Selection

In order to establish a purposive sample of users across sleep products, various sleep products were categorized. Based on their background knowledge and a media scan, the authors generated a list of 54 technologies over the course of a few months. This composed list was supplemented with apps because the goal of the Sleepcare project is to design a virtual sleep coach on a mobile phone. The first 25 Android apps and 25 iPhone apps found in Google Play and the iTunes store with the search word "sleep" on November 19, 2012, were added to the product list. A total of 7 apps were unrelated to sleep—3 games, 2 hypnosis apps, 1 unlock, 1 music timer—and were therefore discarded, resulting in a list of 97 sleep products. The categorization made in this paper aims to be simple and

objective. Sleep products were categorized based on their goal and the medium used. Figure 2 shows the distribution of the products across the two dimensions: goal and medium. The size of the bubbles shows how many products belong to the intersections of the categories.

After identifying the categories of existing sleep products, the next step was to learn more about the users' usage and adherence to the sleep products. Interviews were conducted with people who used a sleep product in each of the largest product-medium combinations (eg, apps that help people fall asleep).

Figure 2. A graph showing the relationship between the goal of sleep products and the medium used. The size of each bubble indicates how many products of the 97 identified sleep products belong to that category.



Participants

People registered as participants at the Sleepcare project website [48] were invited to participate in the interviews if they had ever used a technology-mediated sleep product. In addition, two sleep therapists were asked to invite people who used sleep coaching products, as none of the respondents to the call used a sleep coaching product. A total of 15 Dutch persons agreed to be interviewed—6 (40%) females and 9 (60%) males—their ages ranging from 22 to 65 years (mean 37.5, SD 14.8). The mean Pittsburgh Sleep Quality Index (PSQI) [49] score was 8.0 (SD 4.0), with 12 out of the 15 (80%) interviewees having a score above 5, which is the threshold for poor sleep quality classification.

Interviews

Besides adherence, the interviews covered other topics to gain insight regarding users' experiences with sleep products. Therefore, the semistructured interviews included both adherence-related questions and questions regarding the factors of the Unified Theory of Acceptance and Use of Technology (UTAUT) model [50]. The described results of the interviews in this paper, however, will only include adherence-related topics. The interviews were conducted in person, by Skype, or by telephone by the first author. The audio of the interviews was recorded. The study was approved by the Human Research Ethics Committee of Delft University of Technology.

Analyses

The first author (CH) performed the data analysis following the phases of thematic analysis as described by Braun and Clarke [51]. The first author familiarized herself with the data (phase 1) by conducting and transcribing the interviews, and reading the transcripts. While reading, the initial codes were generated bottom-up (phase 2). The first author coded the transcriptions and iteratively generated hierarchical codes and themes (phase 3). Short summaries of the codes and themes related to adherence were written down. The first and last author (CH and WB, respectively) discussed these summaries (phase 4) to form three final adherence-related categories (phase 5). In addition, an independent researcher applied the coding scheme to one of the interviews in order to minimize the threats to confirmability (known as objectivity in quantitative research). The independent coder confirmed the applicability and usefulness of the codes.

Study III

Overview

The envisioned coach would use different adherence-enhancing strategies during the entire coaching process. For example, different roles (eg, motivator and educator) could be played by different virtual characters to increase the effect of the to-be-developed sleep coach (ie, split-persona effect) [52]. Around 25 strategies were allocated to the coach ranging from strategies involving others (eg, peers or family members), helping with planning (eg, setting goals and making

commitments), and gaming strategies (eg, earning points and taking a quiz). These adherence-enhancing strategies were scripted explicitly in the scenarios in order to discuss them in the focus groups.

Materials

A total of 12 scenarios and 72 claims (see Multimedia Appendices 2 and 3) were written to evaluate the adherence-enhancing strategies. Scenarios consisted of stories about people and their activities, goals, and motivations regarding a system [53]. Claims stated important design decisions (eg, about the adherence-enhancing strategies) that needed to be evaluated in the focus group. Furthermore, three fictitious people varying in age, gender, family situation, and readiness-to-change were created to act in the scenarios (see Multimedia Appendix 4 for these personas).

Procedure and Participants

The scenarios and claims were discussed in six focus groups to evaluate the adherence-enhancing strategies. Two groups consisted of potential users, two groups consisted of coaches, and a further two groups consisted of sleep experts. Demographics of the Dutch participants can be found in Table 3. Each session lasted 2 hours and included a general introduction, an introduction round of the participants, and approximately four animated videos that represented the different scenarios. After watching one video, the participants were asked to individually rate their agreement with the claims on a 7-point Likert scale. Subsequently, participants were asked in turn to react to the claims and discuss their ideas. The sessions were videotaped for later analysis. The study was approved by the Human Research Ethics Committee of Delft University of Technology.

Analyses

The analysis was an iterative process of developing codes and themes in line with thematic analysis [51]. For that, the videotapes of the sessions were transcribed and summarized by the first author (CH). During that recapitulation, several codes emerged and an initial coding scheme of 12 codes was created. The first author coded the summaries according to this scheme. Additionally, a second coder, independent of the project, coded a sample of the summaries—48 of the 86 claims (56%). The second coder suggested eight additional codes. The two coders came together to discuss the coding scheme and agreed on a new scheme of 15 codes. The coding was improved (with this new scheme) by both coders, and within the sample a Cohen's kappa of .80 was reached. Next, the first author wrote short resumes per theme, making use of quotes.

Table 3. Demographics of the participants per focus group.

Focus groups	Participants, n (% female)	Age in years, mean, (SD)	Number of participants with a PSQI[a]>5, n (%)	Expertise
Potential users 1	8 (38)	35 (12)	3 (38)	N/A[b]
Potential users 2	7 (71)	48 (9)	5 (71)	N/A
Coaches 1	4 (75)	51 (8)	N/A	4 coaches (relationships, lifestyle, didactical)
Coaches 2	5 (80)	50 (6)	N/A	4 coaches (lifestyle, career), 1 psychologist
Sleep experts 1	3 (67)	50 (18)	N/A	1 psychologist, 1 therapist, 1 doctor
Sleep experts 2	4 (75)	47 (14)	N/A	3 researchers, 1 psychologist

[a]Pittsburgh Sleep Quality Index (PSQI).

[b]Not applicable (N/A).

Results

Study I

All analyses were completed with the Comprehensive Meta-Analysis statistical package, version 3, and were based on the random-effects model. In the analyses, a distinction was made between experimental compliance and treatment adherence. All studies reported experimental compliance, and most of them (10) also reported treatment adherence (see Table 1 and Table 2). Experimental compliance was typically determined based on the completion of questionnaires and sleep diaries that were part of the study protocol; for more information see Multimedia Appendix 5. Immediately after the intervention (ie, postmeasures), the experimental compliance for questionnaires was 78% (95% CI 70%-85%), and for sleep diaries 71% (95% CI 65%-77%). At the follow-up assessments, experimental compliance to questionnaires was 72% (95% CI 69%-76%), while for diaries it was 58% (95% CI 52%-64%). These aggregated numbers are displayed in Figure 3. In Multimedia Appendix 6, individual numbers per study, aggregated rates, heterogeneity statistics, and publication bias tests can be found. Generally, the analyses indicated a substantial heterogeneity in the data, which supports the choice for a random-effects model. The shapes of the funnel plots and the Egger test did not suggest a significant publication bias.

Treatment adherence was reported in various ways, which can roughly be classified into two groups, namely self-reports and logs. Self-reports refer to questions in which participants were

asked how well they adhered to the exercises. The five studies that used self-reports found that 41% (95% CI 36%-46%) of the participants met the adherence criteria set in that study. Logs refer to reports that show how many sessions were completed. A total of 5 studies used logs and found that 64% (95% CI 44%-79%) of the participants completed all sessions. If these two kinds of measures are taken together, an average treatment adherence of 52% (95% CI 43%-61%) is reached with reported adherence ranging from 28% [34] to 100% [35] across the 10 studies.

In Figure 3 the compliance and adherence rates and their 95% confidence intervals are shown; as can be seen, the self-reported treatment adherence is significantly different from the experimental compliance rates (nonoverlapping confidence intervals). Furthermore, two meta-regressions were run with studies that reported both experimental compliance and treatment adherence in order to discover a possible relationship between these two measures (experimental compliance and treatment adherence): one meta-regression with experimental compliance to postquestionnaires as the explanatory variable and logged treatment adherence as the outcome variable, and the other meta-regression with experimental compliance to follow-up questionnaires as the explanatory variable and self-reported treatment adherence as the outcome variable. These variables were chosen because most data were available for these combinations of variables. Both analyses did not reveal significant relationships between experimental compliance and treatment adherence (both had $P>.05$).

Lastly, the relationship between treatment adherence and the effect size of the individual treatments was explored. Multimedia Appendix 6, Figure 4, and Table 4 show the results of the meta-regression analysis. The analysis revealed a significant model ($Q_{model}=5.05$, df=1, $P=.03$), with a coefficient of 0.79 ($Z=2.25$, $P=.03$) for adherence. In other words, treatment adherence and treatment effect are positively correlated. For example, if adherence increases with 0.30 (30%), this would coincide with a 0.24 increase in effect size (Hedge's g) of the treatment, which is an increase of a small effect size of 0.20. The analysis also found that 75.4% ($I^2=75.4$, Q=48.87, df=12) of the total variance in effect size could be explained by the variation between the studies. Of this 75.4%, 40% ($R^2=.407$, $T^2_{total}=.059$, $T^2_{unexplained}=.035$) could be explained by treatment adherence.

Figure 3. Mean compliance and adherence rates and their 95% CIs. post: posttreatment measurement; follow: follow-up measurement; Qs: questionnaires; Ds: diaries; self-report: self-reported adherence with questions; logs: automatically logged behavior.

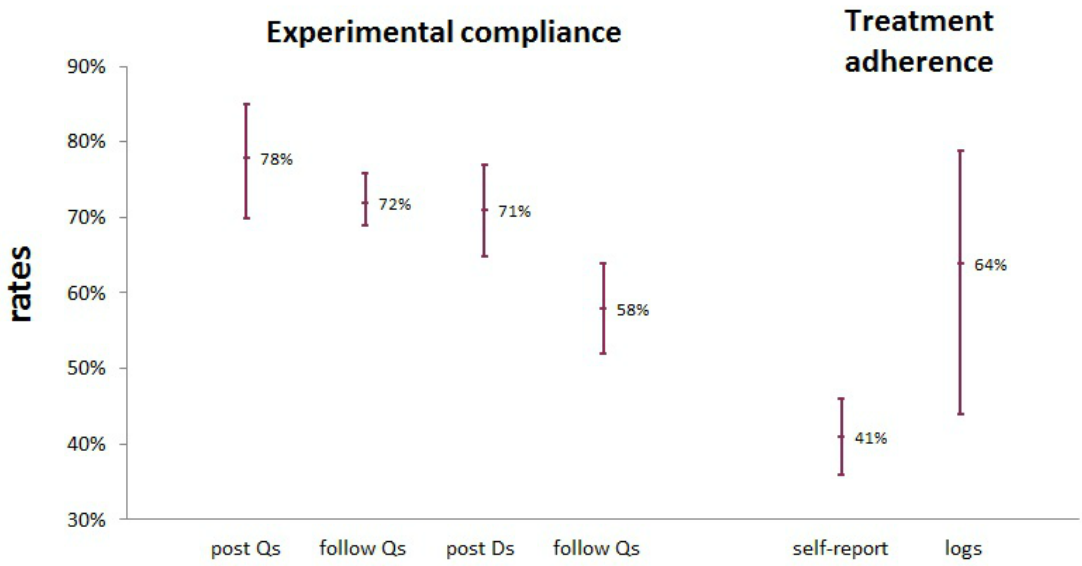
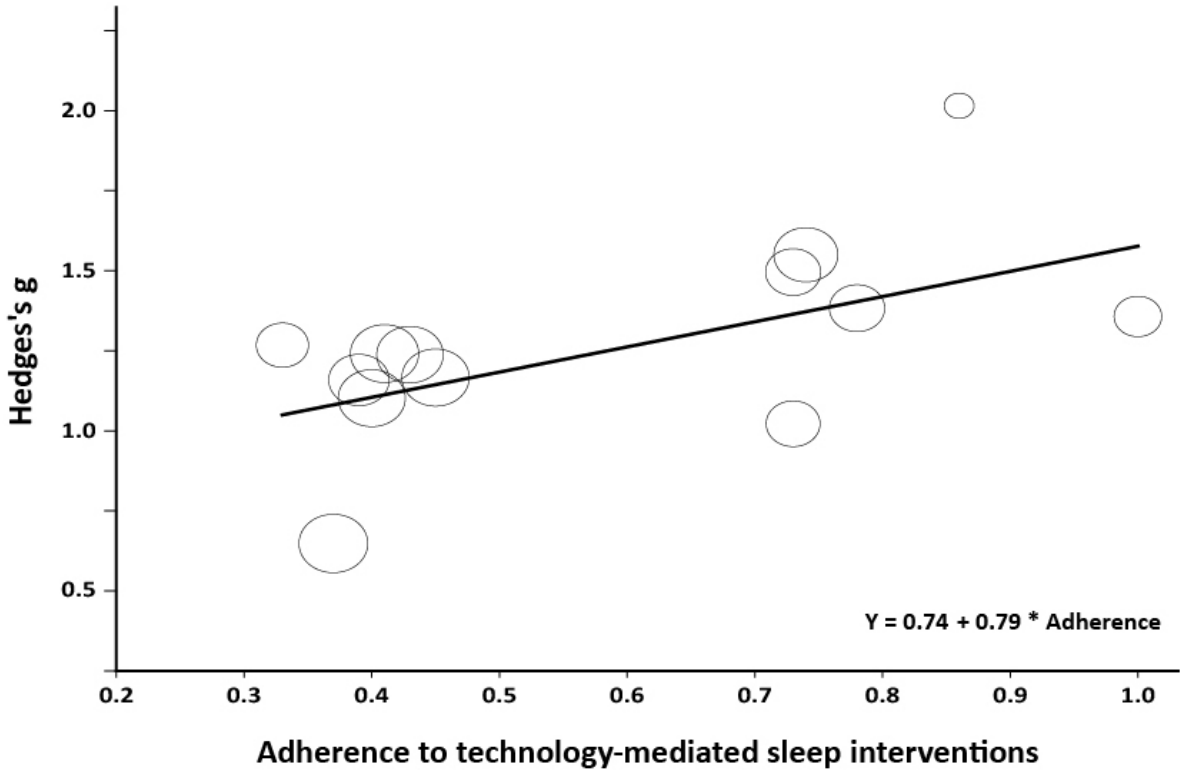


Table 4. Statistics of the meta-regression of adherence and effect size of the individual treatments.

Statistics meta-regression	Coefficient	Standard error	95% CI	*Z*	*P* (2-sided)
Intercept	0.74	0.20	0.35-1.13	3.69	<.001
Adherence	0.79	0.35	0.10-1.47	2.25	.03

Figure 4. Meta-regression of adherence on effect size of treatments. The circles represent the individual studies [15-46]. The circle size indicates the weight of the study. The effect size is given in standard difference in means.



Study II

Overview

The three main categories related to adherence are usage, effectiveness, and adherence (see Textbox 1).

Textbox 1. Main themes mentioned by participants in the interviews.

- Usage
  - Intention
  - Two reasons for usage
    - Overcome sleep problems
    - Interest in the product
  - When used
    - Sleep trackers, alarms, relaxation: used in the evenings
    - Sleep coaches: varying usage times
- Effectiveness
  - Per product type
    - Therapy-related products: no noticeable effect
    - Alarms: ambiguous effect, wake-up is okay, but not waking up better
- Adherence
  - Keep using
    - Therapy-related products: personal attitude
    - Consumer products: need functionality
  - Not using
    - Consumer products
      - No need for functionality (anymore)
      - Product does not work
      - Forget to use product

Usage

Two initial reasons for using a product emerged from the interviews. First, all interviewees used the product to overcome some of their sleeping troubles. Interviewees wanted to wake up better, initiate or maintain sleep, and/or increase insight into their sleep. Second, some interviewees used a product because they thought the product in itself was interesting. Above all, this holds for the sleep-tracking apps. Most products—alarms, automatic sleep trackers, and relaxation support—were used in the evening before going to sleep. Most of those products, however, were not used on the weekend. Sleep coach usage varied, depending on the kind of assignments included in the product (eg, diary, relaxation, sleep hygiene exercise, bedtime scheduling).

Participants' quotes regarding reasons to start using sleep products were as follows (translated from Dutch):

> *The reason was that in my opinion I was awake too often, and too long. I could not fall asleep anymore.* [Interview #11, online sleep therapy]
>
> *Friends of mine had the app and I wanted to try it as well.* [Interview #4, sleep tracking app]

Participants' quotes regarding usage of sleep products were as follows (translated from Dutch):

> *I turn it on in the evening when I am lying in bed and want to go to sleep* [Interview #1, relaxation app]
>
> *Every day I needed to get up I used it, but on the weekends, for example, when I don't need to get up I didn't do anything with the app.* [Interview #7, sleep tracking app]
>
> *Actually, I did it whenever it suited me [about filling in a sleep diary].* [Interview #10, online sleep therapy]

Effectiveness

One of the initial arguments for using a product was to overcome some kind of sleeping problem. However, the online sleep therapies were not perceived as having an effect on the interviewees' sleep problems. Additionally, interviewees mentioned that it was hard to determine if the therapy improved sleep in the long term because they tried several things. Nevertheless, most interviewees took some advice that worked for them and continued applying it. Furthermore, sleep tracking apps as well as online sleep coaches provided the interviewees with more insight into their sleep and habits. Products that wake interviewees up (ie, smart alarm apps and wake-up lights) were

assessed ambiguously. Both types of products did what was expected of them, namely wake the interviewee up. However, the effect of waking up better with the product was doubted. Moreover, smart alarms did not seem to fit into interviewees' daily lives (see quote below from interview #4, sleep tracking app).

Participants' quotes regarding the effectiveness of sleep products were as follows (translated from Dutch):

> *The goal to sleep better was not reached* [Interview #12, online sleep intervention]
>
> *I have no clue if it helped, because it is going better at the moment, but I did other things in that same time period.* [Interview #10, online sleep therapy]
>
> *Also getting out of bed when I am awake for more than 30 minutes. The advice has helped, yes.* [Interview #10, online sleep therapy]
>
> *It measures the sleep debt that you are building up, that was effective.* [Interview #6, sleep tracking app]
>
> *It did what it supposed to do, wake me up.* [Interview #7, sleep tracking app]
>
> *Still not very well, but it became a little bit better, a little bit more pleasant* [Interview #5, sunrise alarm]
>
> *Problem of the app [smart alarm] is that you do not know what time you will wake up exactly. If I have an appointment somewhere I need an hour to get ready. If you do not know how late your alarm will go, it is hard to plan.* [Interview #4, sleep tracking app]

Adherence

In general, interviewees perceived their own usage as sufficient. Interviewees especially perceived their own personal attitude, beliefs, and willpower as important for adherence. These personal characteristics were regarded as particularly important for adherence to therapy-related products. The usage of consumer products (eg, an alarm clock) was continued, because the interviewees needed the functionality. The main arguments for not using a consumer product were (1) no perceived need for the product, (2) a perceived lack of effectiveness, and (3) the interviewee forgot to use the product.

Participants' quotes regarding the satisfaction about adherence of sleep products were as follows (translated from Dutch):

> *It went well. I cannot remember not doing the exercises.* [Interview #13, online sleep therapy]
>
> *[about doing the exercises everyday] Well, that went ok.* [Interview #10, online sleep therapy]
>
> *I use it 3 or 4 times a week, depending on my needs.* [Interview #1, relaxation app]

Participants' quotes regarding the effect of personal attitude, beliefs, and willpower on adherence of sleep products were as follows (translated from Dutch):

> *I tried to keep myself to it as much as possible, and of course I missed a day now and then, but I tried really hard* [Interview #10, online sleep therapy]
>
> *You cannot just resign and accept your sleep problem.* [Interview #10, online sleep therapy]
>
> *I was really motivated, so that makes a difference.* [Interview #10, online sleep therapy]
>
> *[What dragged you through it?] My will. I intended to do it. I started it and I wanted to get a grip on my sleep problem, so I had to follow through. [So your own determination?] Yes, without discipline you will not succeed.* [Interview #11, online sleep therapy]
>
> *I felt like, I started it, so I should finish it.* [Interview #12, online sleep intervention]
>
> *You have to be serious about it. It is a therapy that you really have to complete, otherwise it will not have an effect. So, you have to believe in it.* [Interview #13, online sleep therapy]
>
> *If you do not recognize the need to change, you should not start it.* [Interview #13, online sleep therapy]

Participants' quotes regarding reasons for using and not using sleep products were as follows (translated from Dutch):

> *You have to set an alarm, anyway.* [Interview #6, sleep tracking app]
>
> *During the holidays there is no need for an alarm.* [Interview #5, sunrise alarm]
>
> *I did not have the impression that the app could change my sleeping pattern.* [Interview #6, sleep tracking app]
>
> *I simply forgot it.* [Interview #4, sleep tracking app]

Study III

Overview

The most obvious emerging themes in the focus groups were *users in control* and *doing it for your own sake*. In general, participants believed in the personal strengths and willpower of users to adhere to the proposed sleep coach. Furthermore, the adherence-enhancing strategies and motivation were discussed. See Textbox 2 for an overview of the results.

Textbox 2. Main themes mentioned in the focus groups with potential users, coaches, and sleep experts.

Users being in control

- Control increases commitment and motivation

Doing it for own sake

- Phrase that was strongly believed in was "I do it for my own sake"

Motivation: three conflicting ideas

- If coach is downloaded, then the user is motivated
- Downloading does not imply motivated usage
- Motivation can arise while using

Adherence-enhancing strategies

- Awarding points for progress
- Not seen as appropriate for sleep coach; however, awarding points can work against own expectations
- Giving compliments
  - Not too often, not for nonsignificant actions
  - Should contain context, and vary over time
- Providing reminders
  - Should not be necessary; however, they are practical
  - Reminders are perceived as positive when set by the users
- Provide rationale: two types of people
  - Type 1: first experience exercise, then explanation
  - Type 2: first explanation, then perform exercise
- I am not the only one
  - Provide a forum, stories from others, amount of app users, statistics

Users in Control

Potential users, coaches, and sleep experts agreed that the users should be in control. Different arguments were given. The coaches and sleep experts mainly argued that giving the user more control increases commitment and motivation. The potential users argued that they use the sleep coach for their own sake, so they want to be in control themselves. Another argument was that not being in control could lead to irritation. Aspects that participants believed the users should be able to control were the following: reminders, amount of information given by the app, scheduling exercises, decisions about motivation level, sharing therapy progress, sharing the outcome of questionnaires, and parameters shown in sleep diary overview.

Doing it for Your Own Sake

The other interesting theme was *doing it for your own sake*. In one scenario, there was an example exercise which entails making a list of people who can help you. In general, this exercise met resistance by the potential users. The idea that you have to solve your problems yourself was dominant. Users would feel ashamed to ask for support, and they believed the virtual sleep coach should help them. On the other hand, the coaches stated that thinking about social resources, such as family and peers, could really help people. The coaches mentioned that coachees usually consult a coach exactly because they try to solve their problems themselves, instead of asking their social resources for help. One potential user shared that only informing other people about her sleep problems and therapy already helped her a lot, even without asking for support. Nonetheless, the general mind-set was that people use such a coach for their own sake, and that they are and should be able to take responsibility for their own adherence.

Motivation

The claims underlying the envisioned usage scenarios stated that users should be motivated before they start sleep treatment, otherwise the probability of dropping out would be too high. The focus groups with the sleep experts manifested three different ideas about motivation. Some of the sleep experts argued that people will be motivated at least a little bit when they have downloaded the app, since that requires some effort. On the other hand, it was also argued that someone could show interest in the sleep coach, but he or she would not necessarily be motivated to use the sleep coach. Third, it was argued that

motivation could arise during different phases of a therapy; for example, after someone performs an exercise and experiences its effects. In that situation, users would not need to be highly motivated at the beginning of the therapy.

Adherence-Enhancing Strategies

Overview

Several adherence-enhancing strategies and ideas to increase motivation were scripted in the scenarios (eg, awarding points, compliments, reminders) and are discussed below.

Awarding Points for Progress

In general, participants reacted adversely to the idea of awarding points as described in the scenarios, mainly because the sleep coach was seen as a serious program for adults. Furthermore, it was believed that a point system is not appropriate for sleep exercises, but more for workout programs. Nevertheless, a few participants spoke up and said that they liked the idea of points. A few stories came up about how awarded points motivated participants in other domains against their own expectations. Thus, points might improve adherence, despite users' initial reluctance.

Compliments

Furthermore, both the coaches and the potential users made negative remarks about the compliments. In principle, both groups thought compliments could enhance a user's experience, but compliments should not be given too often, or for nonsignificant actions. They argued that compliments should contain context and should vary over time. Otherwise, compliments would not increase motivation.

Reminders

Reminders were embraced, as long as users are in control of those reminders. The users wanted to set the reminders themselves, because sometimes "you just forget to do something." On the other hand, some users stated they do not need reminders, since they are using the sleep coach for their own sake. Besides that, they argued that they are adults, are motivated, and have self-discipline. Both the coaches and the sleep experts agreed with those potential users and thought that reminders should not be necessary. However, from a practical point of view, they understood that people sometimes do forget to do therapy exercises.

Ideas Generated by the Participants

Other ideas to improve motivation mentioned by the participants were as follows: provide a rationale, show statistics, decrease the feeling of being alone, positive feedback, taking small steps, choosing your own coach, demanding a small investment before starting, and showing how much effort users have already invested.

According to the sleep experts, rationales for doing an exercise should be given before users start an exercise. However, the potential users and coaches mentioned there are two types of people: people who want to know how and why things are the way they are, and people who just want to experience an exercise and afterward gain an understanding of that exercise.

Secondly, different ideas were offered to ensure that users do not feel as if they are the only ones suffering from sleep problems. Ideas included a forum (suggested by users and coaches), reading stories from peers (suggested by coaches), and a measure that indicates how many people are using the app (suggested by sleep experts). The idea was that decreasing the feeling of being the only one with sleep problems could increase the motivation of users to adhere to the sleep therapy.

Discussion

Study I

The meta-analysis of adherence rates found a mean experimental compliance of at least 70%, except for the follow-up diaries. Filling out a diary every day for a full week a few months after the intervention requires quite some effort, which might explain a lower adherence rate (58%) to follow-up diaries than to the other experimental compliance measures. The average self-reported treatment adherence was 41%, whereas the average logged adherence was higher at 64%. This is surprising because the self-reported adherence was less "strict" than the logged adherence; for instance, users were categorized as adherent when they reported doing an exercise a certain number of times (eg, more than 4 times a week), while the logged adherence rate was based on doing all exercises. The average treatment adherence rate (logged and self-reported, combined) was 52%. Although self-reports and logs are not exactly the same, they both measure adherence and are similar enough to be combined. Nevertheless, this general adherence rate of 52% should be interpreted carefully.

Furthermore, this meta-analysis confirmed that treatment adherence is positively related to treatment effect when it comes to technology-mediated insomnia treatment. Moreover, this analysis showed that experimental compliance and treatment adherence are not related. In other words, the percentage of participants who filled out questionnaires after the intervention was not found to be an indication of how well people adhered to the treatment. Therefore, it seems important to distinguish between experimental compliance and treatment adherence.

The quality of the individual studies was not assessed using a predefined algorithm, which might be a limitation. However, the included studies were all published in peer-reviewed journals and proceedings, which warrant an acceptable level of quality. Besides, Glass and colleagues argue that all studies should be included [54]. According to them, all studies should be reviewed in context with each other regarding the topics at issue, not necessarily regarding the overall quality of each study. Since adherence is the main focus of this paper, instead of examining a possible relationship between general study quality and adherence [54], the methodological differences of measuring adherence were reviewed by differentiating between experimental compliance and treatment adherence, and self-reported and logged adherence.

Study II

The aim of the interviews was to gain more insight into the reasons why coachees adhere to technology-mediated sleep products. Surprisingly, interviewees were quite satisfied with

their own usage, which departs from the average 52% adherence rate found in the meta-analysis. The reasons why people started using a product were either out of interest or to overcome sleep problems. However, the products' effectiveness was doubted by the interviewees and was given as a reason to stop using a product. In interviewees' own opinions, they continued to use consumer products because they needed the functionality, whereas they adhered to therapy-related products because of their own attitudes, beliefs, and willpower. Previous research has also identified functionality as an important determinant for adherence in online sleep treatment [55]. Reasons for nonadherence were as follows: no need for the functionality, lack of effectiveness, or just forgetfulness.

Furthermore, it seemed challenging for interviewees to identify adherence-enhancing strategies in the products. It was also difficult for them to come up with an answer to the question of what could be added to the product to help them continue to use the product.

Study III

Focus groups were organized to discuss adherence-enhancing strategies. In addition to motivation, *users in control* and the awareness to *do it for your own sake* proved to be important for adherence. The focus groups provided insights into the up- and downsides of adherence strategies, such as awarding points, compliments, reminders, and community building.

General Discussion

Positive Attitudes Toward Adherence

The interviews and focus groups both revealed that people strongly believe willpower is an effective adherence strategy. Participants believed that their personal attitudes, beliefs, and motivation would ensure that they stick to their intentions of using a product. This result should be interpreted with caution because of three phenomena. First of all, sleep deprivation increases ego depletion [56]. In other words, when people are tired their willpower decreases and it will become more difficult to adhere to anything, including a virtual sleep coach. Second, the interviewees attributed their adherence to their own commitment and attitude, while nonadherence was attributed to malfunctioning of the product. This result should also be interpreted with caution because this phenomenon is in accordance with the self-serving bias. The self-serving bias states that successes are attributed to internal factors, while failure is attributed to external factors [57]. Therefore, the "good" adherence rates in the interviews were attributed to the interviewees' own willpower. Third, the participants in the focus groups were quite optimistic about their anticipated future adherence. Being optimistic about oneself and the future is one of the most robust biases (optimism bias) in psychology [58,59]. Several explanations for this unrealistic optimism has been offered, for example, ignoring everything that could go wrong [58], putting too much weight on current intentions [60], or having too much faith in willpower for future events [61]. These three phenomena provide reasons for treating participants' optimism toward adherence with caution.

Aversion to Adherence-Enhancing Strategies

Apart from relying on willpower for adherence, aversion to adherence-enhancing strategies emerged during the focus groups. Therefore, when designers implement adherence-enhancing strategies they should not assume that users would initially agree with the usefulness of these strategies.

Various design principles for a virtual sleep coach can be adopted from the interviews and focus groups. The first design principle covers functionality. During the first usage phase, the sleep coach should immediately tickle users' interest, for example, by providing automatic sleep tracking. In the interviews, it appeared that interest made coachees start using products. Next, the sleep coach can provide an already-needed functionality (eg, an alarm clock). According to the interviews, a needed functionality ensures that users keep using a product. Lastly, reminders need to be a part of the sleep coach. Reminders make sure that users do not simply forget to adhere to the coach. Both the participants in the interviews and focus groups indicated that sometimes they just forget to use a product. Participants in the focus groups showed a positive attitude toward reminders as long as the users were in control over the reminders. Therefore, including reminders in a sleep coach would be a good first step in future research to increase adherence.

A second design principle could be to withhold adherence support at the start of the intervention (ie, to postpone possible help by a virtual sleep coach). In this way, the coachees are acknowledged and respected as serious, motivated, and autonomous adults. Coachees can prove that they adhere to the assignments of the sleep coach; however, the virtual coach can detect when coachees fail to do their assignments, and then offer support. This support can take different forms (reminders, compliments, awarding points, etc) and can be varied over time based on the needs of the coachee.

A third design principle that can be applied is explaining why willpower does not guarantee success. After such an explanation, the understanding of the added value and acceptance of adherence-enhancing strategies might increase. On top of that, users could be given the control over the employment of adherence-enhancing strategies.

In the authors' opinion, the most important overall design principle is balance. Coachees should not feel overwhelmed with adherence-enhancing strategies, but appreciate some occasional support. Personalization of the virtual sleep coach can ensure that the perfect balance is reached for each and every user. For example, some users might need and appreciate reminders for filling out a sleep diary every day, while other users are more likely to forget to do their relaxation exercises.

Measuring Adherence

Lastly, we want to stress that studies should measure and report treatment adherence, and make a distinction between experimental compliance and treatment adherence. It is important that future studies measure and report adherence rates, since it is only by the adherence measure that it can be established whether the treatment actually induces the observed

outcome. The frequently made statement that adherence is important for the outcome of a treatment [17-19] seems to be supported by the findings of the meta-regression between effectiveness and treatment adherence. As a correlational analysis does not provide insight into the direction of a causal relationship, it remains unclear how effectiveness and adherence influence each other. Nevertheless, if coachees do not follow the treatment protocol (ie, adherence rates close to zero), the outcome could be attributed to other things outside the intervention [21], for example, to the waiting-list effect. Furthermore, it is important to make a distinction between experimental compliance and treatment adherence, since these seem to be two distinct constructs as the meta-analysis found no correlations. An earlier meta-analysis about the effectiveness of CCBT-I found a rather good "adherence rate" of 78% [16]. However, this rate would be considered as experimental compliance according to the definition used in this paper. Similar experimental compliance rates—79%, 72%, 70%, and 57%—were found by the meta-analysis, although treatment adherence was significantly lower. The average self-reported treatment adherence was 42%, whereas the logged treatment adherence was 64%. Although no significant difference between these two measures was found, it is important to consider how adherence is defined and measured. A study [62] that compared a paper diary with an electronic diary found a tremendous difference between self-reported adherence (90.5%) and logged adherence (10.9%) for a paper diary. Lastly, the question remains whether adherence in experimental settings resembles adherence in nonexperimental real-life settings. It could be that adherence rates in experiments are higher than in real-life situations. One possible explanation is the sunk-cost fallacy [63]. To illustrate, experiments demand more from participants regarding (pre-) measurements and participants might therefore be more committed to the intervention. When starting a treatment, they have already invested more time (ie, the sunk cost) compared to patients in nonexperimental settings, and are therefore less likely to drop out.

Research Quality

In order to review the quality of our research, it is helpful to know what we did to take care of the credibility, transferability, dependability, and confirmability of our studies [64]. Firstly, threats to all four concepts were minimalized by utilizing three different research methods—meta-analysis, interviews, and focus groups. Furthermore, the credibility of our findings is also consolidated by data source triangulation—literature, current users, potential users, coaches, and sleep experts. Additionally, honesty from our informants was reinforced by stating there are no right or wrong answers, and by allowing them the possibility to withdraw at any moment. We also had regular debriefing sessions between the executors and supervisors in order to strengthen credibility. The level of transferability to other application fields can only be judged by the readers, since they have the knowledge of these other domains [65]. Furthermore, future work can be done to replicate these findings in other fields. Transferability and dependability assessments are supported by descriptions of the research methods and Multimedia Appendices. Lastly, confirmability was addressed by audit trials and the second coders.

Conclusions

In conclusion, treatment adherence seems important for the effectiveness of technology-mediated insomnia treatments. Individuals expect that they will adhere well to such treatments and would not gain much from adherence-enhancing strategies. They believe willpower is an effective adherence strategy. The 52% average treatment adherence reported in this paper, however, suggests that there is room for improvement. A virtual coach should be able to cope with this “adherence bias,” and persuade users to accept adherence-enhancing strategies (eg, reminders, compliments, and community building). Future research is needed to test the four derived design principles for a virtual coach, which might help to realize a substantial improvement.

Acknowledgments

This research was supported by Philips and the Dutch Technology Foundation (STW), Nationaal Initiatief Hersenen en Cognitie (NIHC) under the Partnership program Healthy Lifestyle Solutions, which is partly financed by the Netherlands Organisation for Scientific Research (NWO).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Studies included in the meta-analysis.

[PDF File (Adobe PDF File), 6KB - jmir_v17i9e214_app1.pdf]

Multimedia Appendix 2

Precontemplation scenario.

[PDF File (Adobe PDF File), 8KB - jmir_v17i9e214_app2.pdf]

Multimedia Appendix 3

Claims in Dutch and English.

[PDF File (Adobe PDF File), 50KB - jmir_v17i9e214_app3.pdf]

Multimedia Appendix 4

Short summary of personas.

[PDF File (Adobe PDF File), 2KB - jmir_v17i9e214_app4.pdf]

Multimedia Appendix 5

General notes about the meta-analysis.

[PDF File (Adobe PDF File), 4KB - jmir_v17i9e214_app5.pdf]

Multimedia Appendix 6

Results of the meta-analysis.

[PDF File (Adobe PDF File), 456KB - jmir_v17i9e214_app6.pdf]

References

1. American Psychiatric Association, editor. Diagnostic and Statistical Manual of Mental Disorders. 5th edition (DSM-5). Arlington, VA: American Psychiatric Association Publishing; 2013.
2. Rosekind MR, Gregory KB. Insomnia risks and costs: health, safety, and quality of life. Am J Manag Care 2010 Aug;16(8):617-626 [FREE Full text] [Medline: 20712395]
3. Baglioni C, Battagliese G, Feige B, Spiegelhalder K, Nissen C, Voderholzer U, et al. Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. J Affect Disord 2011 Dec;135(1-3):10-19. [doi: 10.1016/j.jad.2011.01.011] [Medline: 21300408]
4. Daley M, Morin CM, LeBlanc M, Grégoire J, Savard J. The economic burden of insomnia: direct and indirect costs for individuals with insomnia syndrome, insomnia symptoms, and good sleepers. Sleep 2009 Jan;32(1):55-64 [FREE Full text] [Medline: 19189779]
5. Ohayon MM. Epidemiology of insomnia: what we know and what we still need to learn. Sleep Med Rev 2002 Apr;6(2):97-111. [Medline: 12531146]
6. Ancoli-Israel S, Roth T. Characteristics of insomnia in the United States: results of the 1991 National Sleep Foundation Survey. I. Sleep 1999 May 1;22 Suppl 2:S347-S353. [Medline: 10394606]
7. Benca RM. Diagnosis and treatment of chronic insomnia: a review. Psychiatr Serv 2005 Mar;56(3):332-343. [doi: 10.1176/appi.ps.56.3.332] [Medline: 15746509]
8. Morin CM, LeBlanc M, Daley M, Gregoire JP, Mérette C. Epidemiology of insomnia: prevalence, self-help treatments, consultations, and determinants of help-seeking behaviors. Sleep Med 2006 Mar;7(2):123-130. [doi: 10.1016/j.sleep.2005.08.008] [Medline: 16459140]
9. Morin CM, Gaulier B, Barry T, Kowatch RA. Patients' acceptance of psychological and pharmacological therapies for insomnia. Sleep 1992 Aug;15(4):302-305. [Medline: 1519003]
10. Vincent N, Lionberg C. Treatment preference and patient satisfaction in chronic insomnia. Sleep 2001 Jun 15;24(4):411-417. [Medline: 11403525]
11. Perlis ML, Smith MT, Cacialli DO, Nowakowski S, Orff H. On the comparability of pharmacotherapy and behavior therapy for chronic insomnia. Commentary and implications. J Psychosom Res 2003 Jan;54(1):51-59. [Medline: 12505555]
12. Morin CM, Espie CA. Insomnia: A Clinical Guide to Assessment and Treatment. New York, NY: Kluwer Academic/Plenum Publishers; 2003.
13. Morin CM. Insomnia: Psychological Assessment and Management. New York, NY: Guilford Press; 1993.
14. Morin CM, Beaulieu-Bonneau S, LeBlanc M, Savard J. Self-help treatment for insomnia: a randomized controlled trial. Sleep 2005 Oct;28(10):1319-1327. [Medline: 16295218]
15. Espie CA, Kyle SD, Williams C, Ong JC, Douglas NJ, Hames P, et al. A randomized, placebo-controlled trial of online cognitive behavioral therapy for chronic insomnia disorder delivered via an automated media-rich web application. Sleep 2012 Jun;35(6):769-781 [FREE Full text] [doi: 10.5665/sleep.1872] [Medline: 22654196]
16. Cheng SK, Dizon J. Computerised cognitive behavioural therapy for insomnia: a systematic review and meta-analysis. Psychother Psychosom 2012;81(4):206-216. [doi: 10.1159/000335379] [Medline: 22585048]
17. Beun RJ. Persuasive strategies in mobile insomnia therapy: alignment, adaptation, and motivational support. Pers Ubiquit Comput 2012 Jul 8;17(6):1187-1195. [doi: 10.1007/s00779-012-0586-2]

18. 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;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
19. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
20. Adherence to Long-Term Therapies: Evidence for Action. Geneva, Switzerland: World Health Organization; 2003. URL: http://apps.who.int/iris/bitstream/10665/42682/1/9241545992.pdf [accessed 2015-08-30] [WebCite Cache ID 6bBEJ6Lk1]
21. Gould RA, Clum GA. A meta-analysis of self-help treatment approaches. Clin Psychol Rev 1993 Jan;13(2):169-186. [doi: 10.1016/0272-7358(93)90039-O]
22. Beun R, Griffioen-Both F, Ahn R, Fitrianie S, Lancee J. Modeling interaction in automated e-coaching: a case from insomnia therapy. In: Proceedings of the Sixth International Conference on Advanced Cognitive Technologies and Applications. 2014 Presented at: Sixth International Conference on Advanced Cognitive Technologies and Applications; May 25-29, 2014; Venice, Italy.
23. Horsch C, Brinkman W, van Eijk R, Neerincx M. Towards the usage of persuasive strategies in a virtual sleep coach. In: Proceedings of the UK HCI 2012 Workshop on People, Computers & Psychotherapy.: BISL; 2012 Presented at: UK HCI 2012 Workshop on People, Computers & Psychotherapy; September 10, 2012; Birmingham, UK.
24. Edinger JD, Means MK. Cognitive-behavioral therapy for primary insomnia. Clin Psychol Rev 2005 Jul;25(5):539-558. [doi: 10.1016/j.cpr.2005.04.003] [Medline: 15951083]
25. Vermeire E, Hearnshaw H, Van Royen P, Denekens J. Patient adherence to treatment: three decades of research. A comprehensive review. J Clin Pharm Ther 2001 Oct;26(5):331-342. [Medline: 11679023]
26. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
27. Hebert EA, Vincent N, Lewycky S, Walsh K. Attrition and adherence in the online treatment of chronic insomnia. Behav Sleep Med 2010;8(3):141-150. [doi: 10.1080/15402002.2010.487457] [Medline: 20582757]
28. Kuniavsky M. Observing the User Experience: A Practitioner's Guide to User Research. San Francisco, CA: Morgan Kaufmann Publishers; 2003.
29. van Straten A, Cuijpers P. Self-help therapy for insomnia: a meta-analysis. Sleep Med Rev 2009 Feb;13(1):61-71. [doi: 10.1016/j.smrv.2008.04.006] [Medline: 18952469]
30. Ho FY, Chung K, Yeung W, Ng TH, Kwan K, Yung K, et al. Self-help cognitive-behavioral therapy for insomnia: a meta-analysis of randomized controlled trials. Sleep Med Rev 2015 Feb;19:17-28. [doi: 10.1016/j.smrv.2014.06.010] [Medline: 25104471]
31. Oosterhuis A, Klip EC. The treatment of insomnia through mass media, the results of a televised behavioral training programme. Soc Sci Med 1997 Oct;45(8):1223-1229. [Medline: 9381235]
32. Rybarczyk B, Lopez M, Benson R, Alsten C, Stepanski E. Efficacy of two behavioral treatment programs for comorbid geriatric insomnia. Psychol Aging 2002 Jun;17(2):288-298. [Medline: 12061413]
33. Riley WT, Mihm P, Behar A, Morin CM. A computer device to deliver behavioral interventions for insomnia. Behav Sleep Med 2010;8(1):2-15. [doi: 10.1080/15402000903425314] [Medline: 20043245]
34. Lawson S, Jamison-Powell S, Garbett A, Linehan C, Kucharczyk E, Verbaan S, et al. Validating a mobile phone application for the everyday, unobtrusive, objective measurement of sleep. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY: ACM; 2013 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 27-May 2, 2013; Paris, France p. 2497-2506.
35. Lipschitz DL, Landward R, Nakamura Y. An exploratory study of an online mind-body program for poor sleepers in a community sample. Eur J Integr Med 2014 Feb;6(1):48-55. [doi: 10.1016/j.eujim.2013.10.002]
36. Haimov I, Shatil E. Cognitive training improves sleep quality and cognitive function among older adults with insomnia. PLoS One 2013;8(4):e61390 [FREE Full text] [doi: 10.1371/journal.pone.0061390] [Medline: 23577218]
37. Ström L, Pettersson R, Andersson G. Internet-based treatment for insomnia: a controlled evaluation. J Consult Clin Psychol 2004 Feb;72(1):113-120. [doi: 10.1037/0022-006X.72.1.113] [Medline: 14756620]
38. Suzuki E, Tsuchiya M, Hirokawa K, Taniguchi T, Mitsuhashi T, Kawakami N. Evaluation of an internet-based self-help program for better quality of sleep among Japanese workers: a randomized controlled trial. J Occup Health 2008;50(5):387-399 [FREE Full text] [Medline: 18716392]
39. Ritterband LM, Thorndike FP, Gonder-Frederick LA, Magee JC, Bailey ET, Saylor DK, et al. Efficacy of an Internet-based behavioral intervention for adults with insomnia. Arch Gen Psychiatry 2009 Jul;66(7):692-698 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2009.66] [Medline: 19581560]
40. van Straten A, Cuijpers P, Smit F, Spermon M, Verbeek I. Self-help treatment for insomnia through television and book: a randomized trial. Patient Educ Couns 2009 Jan;74(1):29-34. [doi: 10.1016/j.pec.2008.07.050] [Medline: 18801639]
41. Vincent N, Lewycky S. Logging on for better sleep: RCT of the effectiveness of online treatment for insomnia. Sleep 2009 Jun;32(6):807-815 [FREE Full text] [Medline: 19544758]

42. Lancee J, van den Bout J, van Straten A, Spoormaker VI. Internet-delivered or mailed self-help treatment for insomnia?: a randomized waiting-list controlled trial. Behav Res Ther 2012 Jan;50(1):22-29. [doi: 10.1016/j.brat.2011.09.012] [Medline: 22055281]
43. Ritterband LM, Bailey ET, Thorndike FP, Lord HR, Farrell-Carnahan L, Baum LD. Initial evaluation of an Internet intervention to improve the sleep of cancer survivors with insomnia. Psychooncology 2012 Jul;21(7):695-705 [FREE Full text] [doi: 10.1002/pon.1969] [Medline: 21538678]
44. Lancee J, van den Bout J, van Straten A, Spoormaker VI. Baseline depression levels do not affect efficacy of cognitive-behavioral self-help treatment for insomnia. Depress Anxiety 2013 Feb;30(2):149-156. [doi: 10.1002/da.22004] [Medline: 23080373]
45. Lancee J, van den Bout J, Sorbi MJ, van Straten A. Motivational support provided via email improves the effectiveness of internet-delivered self-help treatment for insomnia: a randomized trial. Behav Res Ther 2013 Dec;51(12):797-805. [doi: 10.1016/j.brat.2013.09.004] [Medline: 24121097]
46. van Straten A, Emmelkamp J, de Wit J, Lancee J, Andersson G, van Someren EJ, et al. Guided Internet-delivered cognitive behavioural treatment for insomnia: a randomized trial. Psychol Med 2014 May;44(7):1521-1532. [doi: 10.1017/S0033291713002249] [Medline: 24001364]
47. Holmqvist M, Vincent N, Walsh K. Web- vs. telehealth-based delivery of cognitive behavioral therapy for insomnia: a randomized controlled trial. Sleep Med 2014 Feb;15(2):187-195. [doi: 10.1016/j.sleep.2013.10.013] [Medline: 24461370]
48. Sleepcare. URL: http://www.ikgalekkerslapen.nl/ [accessed 2014-11-17] [WebCite Cache ID 6U9ZBP8lv]
49. Buysse DJ, Ancoli-Israel S, Edinger JD, Lichstein KL, Morin CM. Recommendations for a standard research assessment of insomnia. Sleep 2006 Sep;29(9):1155-1173. [Medline: 17040003]
50. Venkatesh V, Morris M, Davis G, Davis F. User acceptance of information technology: toward a unified view. MIS Quarterly 2003 Sep;27(3):425-478. [doi: 10.2307/30036540]
51. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
52. Baylor A, Ebbers S. The pedagogical agent split-persona effect: when two agents are better than one. In: Proceedings of EdMedia: World Conference on Educational Media and Technology. Waynesville, NC: Association for the Advancement of Computing in Education (AACE); 2003 Presented at: EdMedia: World Conference on Educational Media and Technology; 2003; Honolulu, HI p. 459-462.
53. Carroll J. Five reasons for scenario-based design. Interact Comput 2000 Sep;13(1):43-60. [doi: 10.1016/S0953-5438(00)00023-0]
54. Glass GV, McGaw B, Smith ML. Meta-Analysis in Social Research. Beverly Hills, CA: Sage Publications; 1981.
55. Middlemass J, Davy Z, Cavanagh K, Linehan C, Morgan K, Lawson S, et al. Integrating online communities and social networks with computerised treatment for insomnia: a qualitative study. Br J Gen Pract 2012 Dec;62(605):e840-e850 [FREE Full text] [doi: 10.3399/bjgp12X659321] [Medline: 23211265]
56. Barnes CM, Schaubroeck J, Huth M, Ghumman S. Lack of sleep and unethical conduct. Organ Behav Hum Decis Process 2011 Jul;115(2):169-180. [doi: 10.1016/j.obhdp.2011.01.009]
57. Zuckerman M. Attribution of success and failure revisited, or: The motivational bias is alive and well in attribution theory. J Pers 1979 Jun;47(2):245-287. [doi: 10.1111/j.1467-6494.1979.tb00202.x]
58. Buehler R, Griffin D, Ross M. Exploring the "planning fallacy": Why people underestimate their task completion times. J Pers Soc Psychol 1994;67(3):366-381. [doi: 10.1037/0022-3514.67.3.366]
59. Weinstein ND. Unrealistic optimism about future life events. J Pers Soc Psychol 1980;39(5):806-820. [doi: 10.1037/0022-3514.39.5.806]
60. Koehler DJ, Poon CS. Self-predictions overweight strength of current intentions. J Exp Soc Psychol 2006 Jul;42(4):517-524. [doi: 10.1016/j.jesp.2005.08.003]
61. Helzer EG, Gilovich T. Whatever is willed will be: a temporal asymmetry in attributions to will. Pers Soc Psychol Bull 2012 Oct;38(10):1235-1246. [doi: 10.1177/0146167212448403] [Medline: 22675023]
62. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. Control Clin Trials 2003 Apr;24(2):182-199. [Medline: 12689739]
63. Arkes HR, Blumer C. The psychology of sunk cost. Organ Behav Hum Decis Process 1985 Feb;35(1):124-140. [doi: 10.1016/0749-5978(85)90049-4]
64. Lincoln YS, Guba EG. Naturalistic Inquiry. Beverly Hills, CA: Sage Publications; 1985.
65. Shenton AK. Strategies for ensuring trustworthiness in qualitative research projects. Education for Information 2004;22(2):63-75.

Abbreviations

AASM: American Academy of Sleep Medicine
ASS: active sleep sampling
CBT: cognitive behavioral therapy

CBT-I: cognitive behavioral therapy for insomnia
CCBT-I: computerized cognitive behavioral therapy for insomnia
CTh: cognitive therapy
CTr: cognitive training
D: sleep diary
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th Edition
IRT: imagery relief therapy (placebo)
ISI: Insomnia Severity Index
MBB: mind-body bridging
N/A: not applicable
NIHC: Nationaal Initiatief Hersenen en Cognitie
NWO: Netherlands Organisation for Scientific Research
PC: personal computer
PDA: personal digital assistant
PSQI: Pittsburgh Sleep Quality Index
Q: questionnaire
RX: relaxation
SC: stimulus control
SE: sleep education
SH: sleep hygiene
SMMT: self-monitoring minimal treatment
SR: sleep restriction
SQ: sleep quality
STW: Dutch Technology Foundation
TAU: treatment as usual
TV: television
UTAUT: Unified Theory of Acceptance and Use of Technology
WHO: World Health Organization

Edited by G Eysenbach; submitted 06.12.14; peer-reviewed by J Dizon, N Siriwardena; comments to author 21.04.15; revised version received 03.06.15; accepted 24.06.15; published 04.09.15.

Please cite as:
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;17(9):e214
URL: http://www.jmir.org/2015/9/e214/
doi:10.2196/jmir.4115
PMID:26341671

Original Paper

Use and Effectiveness of a Video- and Text-Driven Web-Based Computer-Tailored Intervention: Randomized Controlled Trial

Michel Jean Louis Walthouwer[1,2], MSc; Anke Oenema[1,2], PhD; Lilian Lechner[2,3], PhD; Hein de Vries[1,2], PhD

[1]Maastricht University, Department of Health Promotion, Maastricht, Netherlands

[2]School for Public Health and Primary Care (CAPHRI), Maastricht University, Maastricht, Netherlands

[3]Open University of the Netherlands, Department of Psychology, Heerlen, Netherlands

Corresponding Author:
Michel Jean Louis Walthouwer, MSc
Maastricht University
Department of Health Promotion
Peter Debyeplein 1
Maastricht, 6229 HA
Netherlands
Phone: 31 433882402
Fax: 31 433882402
Email: michel.walthouwer@maastrichtuniversity.nl

Abstract

Background: Many Web-based computer-tailored interventions are characterized by high dropout rates, which limit their potential impact.

Objective: This study had 4 aims: (1) examining if the use of a Web-based computer-tailored obesity prevention intervention can be increased by using videos as the delivery format, (2) examining if the delivery of intervention content via participants' preferred delivery format can increase intervention use, (3) examining if intervention effects are moderated by intervention use and matching or mismatching intervention delivery format preference, (4) and identifying which sociodemographic factors and intervention appreciation variables predict intervention use.

Methods: Data were used from a randomized controlled study into the efficacy of a video and text version of a Web-based computer-tailored obesity prevention intervention consisting of a baseline measurement and a 6-month follow-up measurement. The intervention consisted of 6 weekly sessions and could be used for 3 months. ANCOVAs were conducted to assess differences in use between the video and text version and between participants allocated to a matching and mismatching intervention delivery format. Potential moderation by intervention use and matching/mismatching delivery format on self-reported body mass index (BMI), physical activity, and energy intake was examined using regression analyses with interaction terms. Finally, regression analysis was performed to assess determinants of intervention use.

Results: In total, 1419 participants completed the baseline questionnaire (follow-up response=71.53%, 1015/1419). Intervention use declined rapidly over time; the first 2 intervention sessions were completed by approximately half of the participants and only 10.9% (104/956) of the study population completed all 6 sessions of the intervention. There were no significant differences in use between the video and text version. Intervention use was significantly higher among participants who were allocated to an intervention condition that matched their preferred intervention delivery format. There were no significant interaction terms for any of the outcome variables; a match and more intervention use did not result in better intervention effects. Participants with a high BMI and participants who felt involved and supported by the intervention were more likely to use the intervention more often.

Conclusions: Video delivery of tailored feedback does not increase the use of Web-based computer-tailored interventions. However, intervention use can potentially be increased by delivering intervention content via participants' preferred intervention delivery format and creating feelings of relatedness. Because more intervention use was not associated with better intervention outcomes, more research is needed to examine the optimum number of intervention sessions in terms of maximizing use and effects.

Trial Registration: Nederlands Trial Register: NTR3501; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=3501 (Archived by WebCite at http://www.webcitation.org/6b2tsH8Pk)

(J Med Internet Res 2015;17(9):e222) doi:10.2196/jmir.4496

KEYWORDS

intervention use; Web-based; computer tailoring; obesity; educational level; delivery strategy; matching

Introduction

Web-based computer-tailored interventions are increasingly being tested to target health-related behaviors, such as physical activity, dietary intake, and smoking [1,2]. These interventions are potentially cost-effective because they can reach many people with individualized information via the Internet for relatively low costs [1,3-5]. Unfortunately, these interventions often are not optimally used by the intended target groups. A rapid decline in use of intervention sessions in the first weeks after initial participation is seen, in particular among people with a low educational level [6-13]. As a result, many people will not be exposed to all essential intervention content if it is provided in multiple sessions over a longer period of time. This limits the potential impact of such interventions because evidence suggests that repeated intervention use is necessary to achieve sustainable behavioral changes [13-19]. Hence, many scholars have highlighted the necessity to increase research on strategies that can improve the (prolonged) use of Web-based computer-tailored interventions, particularly among people with a low educational level [6,11,20-22]. However, although research into this area is growing, thus far studies have yielded no or only modest improvements in intervention use [7,11].

Hence, the main aim of this study was to contribute to the required insight into how the use of Web-based computer-tailored interventions can be improved. This study is part of a randomized controlled trial (RCT) that has demonstrated that a video version of a Web-based computer-tailored obesity prevention intervention is more effective in reducing body mass index (BMI) and energy intake than a similar text version [23]. The first study aim was to examine if the video version resulted in more intervention use compared to the text version among people with a low educational level in particular. To provide insight into other possible ways to increase intervention use, the second study aim was to examine if the delivery of intervention content via persons' preferred intervention delivery format was associated with more intervention use. A related study aim was to examine if this match and more intervention use were related to better intervention effects. In addition, the final study aim was to examine demographic factors and intervention appreciation variables as potential predictors of intervention use.

Previous research has shown that there are multiple computer-mediated delivery modes that can be used to effectively communicate messages to people, such as text, video, and audio [24]. The idea to examine the additional effects of videos originated from the fact that most Web-based computer-tailored interventions merely consist of "dry" text-based information. Video-based messages are considered to be livelier and, therefore, more likely to be engaging and to stimulate revisits [25-32]. For example, a recent study has shown that providing video-tailored feedback can significantly increase the time spent with a Web-based physical activity intervention [30]. The use of videos, in particular, may be appropriate for persons with a low educational level because these individuals generally are less text-oriented [22,25,33]. In addition, matching the delivery format of intervention content with participants' preferred intervention delivery format may also lead to prolonged intervention use and eventually better health outcomes [34-38]. According to the Elaboration Likelihood Model, an adequate match between a person's preferences and educational materials will stimulate central route processing, which accordingly makes it more likely that positive changes are induced [39]. Conversely, a mismatch can reduce participation and may result in negative outcomes, such as dissatisfaction with the information and eventually poorer intervention effects [35,38].

The use of Web-based computer-tailored interventions can also be improved by gaining more insight into the factors that are predictive for intervention use. For example, prior research has suggested that interventions that are appreciated well are more likely to be used [21,40-42]. Based on Self-Determination Theory (SDT), it has been suggested that intervention use may be higher when an intervention is evaluated well on factors that can increase a person's intrinsic motivation, such as perceived autonomy and relatedness [43,44]. Research has also demonstrated that intervention use is influenced by demographic factors; men and people with a low educational level are more likely to discontinue a Web-based computer-tailored intervention, whereas usage is higher among women and older persons [11,17,45]. More insight into the demographic characteristics predictive for intervention use offers the possibility to encourage revisits among people who are less likely to revisit a Web-based computer-tailored intervention [11].

In conclusion, the main aim of this study was to provide insight into how the use and, relatedly, the effectiveness of Web-based computer-tailored interventions can be improved.

Methods

Ethical Approval

The Ethical Committee of the Open University Heerlen reviewed the study protocol and decided that there was no objection to performing the study. The study is registered in the Dutch Trial Register (NTR3501).

Study Design and Respondents

For this study, we used data from a RCT about the efficacy of the video and text version of the Web-based computer-tailored obesity prevention intervention. These 2 versions were compared to a waiting list control group. Baseline measurements (T0) took place between September 2012 and February 2013 and there was one follow-up measurement 6 months after baseline (T1). Participants were eligible to participate if they were at least 18 years of age, had a paid job, had a BMI between 18.5

and 30 kg/m^2, and had sufficient command of the Dutch language. People with a physical condition that influenced their dietary or physical activity pattern (eg, diabetes) were excluded from participation.

Participants were recruited via occupational health centers, but mainly directly through worksites and via advertisements in newspapers. Participants had to register at the study website, where they could read more information about the study and the intervention. After registration, participants were randomly assigned to 1 of the 3 study conditions (video version, text version, control group) in a computer-determined sequence after which they received a username and password by email. With their account, they could log in to the website and fill out the baseline questionnaire. Participants in the intervention conditions were given access to the intervention 2 weeks after completion of the baseline questionnaire. To decrease the likelihood of attrition, participants received 2 email reminders per questionnaire. Participants could further win 1 of 100 cash prizes of €100 if they completed all questionnaires (ie, total amount of prizes was €10,000).

Intervention

The video and text version of the Web-based computer-tailored intervention had the objective to prevent weight gain or achieve modest weight loss by guiding people in making and maintaining small changes in dietary intake and physical activity. Both versions had exactly the same content. In the video version, most educational content (ie, approximately 75%) was provided via videos, whereas the text version provided the educational content only via text without any visual elements. The text in the video version was used to give instructions about setting goals and making plans, delivering optional in-depth information (eg, about the small changes approach), and giving feedback about how to deal with many different barriers. In the videos, professional actors read the messages aloud by means of a news-driven format. The I-Change Model [46] and self-regulation theories [47,48] were used as the theoretical basis of the intervention. The intervention could be used for a maximum 3 months and consisted of 6 sessions, which each lasted approximately 15 minutes. Session 2 could be followed directly after session 1, but the subsequent sessions were weekly to monitor participants' progress over time. To decrease the likelihood of attrition, participants received 2 email reminders per session. Detailed information about the development of the intervention can be found elsewhere [25].

The aim of session 1 was to help participants set an appropriate weight goal (ie, maintain current weight or lose a little weight) and a behavior change goal (ie, improve dietary intake, physical activity, or both). For this purpose, participants received tailored feedback about their BMI, dietary intake, physical activity level, and sociocognitive beliefs toward making changes in diet and physical activity (ie, attitude, self-efficacy, and social influence).

Session 2 aimed to help participants make appropriate "if-then" action plans (ie, implementation intentions) [49]. For this purpose, tailored feedback was given to indicate which specific behavior changes participants could make to achieve their weight goal. Subsequently, participants had to specify when, where, and how they were going to perform the desired behavior change. After this session, participants could start with the planned behavior change.

The last 4 sessions could be accessed in the next weeks, with at least 1 week between each session. The main aim of these sessions was to indicate whether or not participants had achieved their goals. Session 3, for example, provided tailored feedback about participants' behavior change progress and offered the possibility to make coping plans. In addition, session 4 also consisted of narratives in which a role model told how his/her behavior change was going and how he/she dealt with difficult situations. In this session, participants could also change their goals and plans. Session 5 was similar to session 4, but also provided iterative feedback concerning participants' success in attaining their weight goal. Finally, session 6 was similar to the previous session, but also offered the possibility to set a long-term weight goal and make plans for achieving this goal.

Measurements

Outcome Variables

Intervention use was assessed by examining how many sessions were completed during the entire intervention period. Based on website tracking data, it was possible to assess whether or not participants had completed a particular intervention session. For each session, completion was scored as 1 and noncompletion as 0. These scores were summed, which resulted in a total score for intervention use ranging from 0 to 6 completed sessions.

Participants' dietary intake was assessed at both T0 and T1 by means of a food frequency questionnaire consisting of 66 items. This questionnaire was based on a validated questionnaire concerning fat intake [50]. Our questionnaire mainly assessed intake of energy-dense products originating from 6 different food categories (ie, dairy products, sandwiches and fillings, food at dinner, sweet and savory snacks, hot and cold beverages, and alcohol). For each food product, the frequency (ie, number of days per week) and quantity (ie, servings per day) were assessed and, when applicable, portion size and type of product (eg, use of skimmed, semiskimmed, or whole milk) were assessed as well. A score for the average daily intake of calories from energy-dense food products was calculated by combining these questions with the energy value of each food product [51].

At T0 and T1, physical activity was assessed using the Short Questionnaire to Assess Health-Enhancing Physical Activity (SQUASH) [52]. Research has shown that this is a reliable and valid questionnaire to estimate the level of physical activity among Dutch adults [52]. Per category (ie, commuting activities, leisure time activities, household activities, and activities at work), participants had to indicate on how many days per week they engaged in this activity, the average time per day spent in doing this activity, and the intensity of the activity (ie, light, moderate, or vigorous). The scores on these questions were used to calculate a total score for the average daily minutes of moderate and vigorous intensity physical activity.

To assess BMI, participants were asked to report their height in meters and their body weight in kilograms as measured in the morning without clothes and shoes at both T0 and T1. In addition, participants who had not completed the online

follow-up questionnaire at T1 were contacted by telephone to assess their body weight. In-line with the online questionnaire, these participants were asked to indicate their body weight in kilograms as measured in the morning without clothes and shoes.

Intervention Delivery Format Preference

At T0, intervention delivery format preference was measured by asking via which delivery format participants preferred to receive information in Web-based computer-tailored interventions (text only, videos only, combination of text and videos, or no preference). The answer to this question was combined with the assigned study condition to determine whether or not participants' preference matched with the delivery format of the allocated intervention condition. This resulted in 2 groups of participants: (1) participants with a matched preference and (2) participants with a mismatched preference. Participants in the video condition were considered to have a match when they preferred to receive intervention content via a combination of videos and text (as the video version consisted of both video and text). Participants in the text condition had a match when they had a text-only delivery format preference. Participants in the video condition were considered to have a mismatch when they preferred to receive intervention content via video only or text only. Participants in the text condition had a mismatch when they had a delivery format preference for video only or a combination of video and text. It should be noted that participants who indicated that they did not have a preference for a particular intervention delivery format were not included in this variable (n=320). In addition, participants who were allocated to the control condition were also not included in this variable because they did not receive the video or text version of the intervention.

Appreciation of Intervention

At T1, appreciation of the intervention was assessed by means of 8 concepts. First, participants were asked to indicate on a 5-point Likert scale (1=low and 5=high) how they appreciated the information and feedback messages in the intervention: interesting, useful, understandable, and fitting to own situation. In addition, participants were also asked to give an overall rating of their impression of the intervention on a scale ranging from 1 (very poor) to 10 (excellent). Finally, participants' perceptions of the intervention regarding autonomy, relatedness, and competence were assessed. These 3 concepts were derived from the SDT [43] and measured on a 5-point Likert scale (1=low and 5=high). First autonomy was assessed with 2 items by asking participants to which degree they experienced freedom in setting own goals and plans as well as in deciding which information they could read. Next, competence was assessed using 3 items. Participants had to indicate if the intervention had increased their confidence in their ability to manage their weight, dietary intake, and physical activity behavior. Finally, relatedness was assessed using 3 items by asking participants if they felt involved and supported by the intervention. A mean score was calculated for each of the 3 SDT concepts.

Demographics

Demographic characteristics were assessed at T0 and included gender (1=male; 2=female), age, and educational level (ie, the highest level of education completed). Educational level was classified into 3 categories: low (1=primary or basic vocational school), medium (2=secondary vocational school or high school), and high (3=higher vocational school or university) [53].

Statistical Analyses

All statistical analyses were conducted using SPSS 20.0 (IBM Corp, Armonk, NY, USA), applying a significance level of .05 for single variables and .10 for interaction terms [54]. At both T0 and T1, multiple imputation was used to replace missing values on demographics, sociocognitive variables, and the outcome variables. Based on the dropout rate and the amount of missing values, the number of imputations was set at 40.

Descriptive statistics and frequencies were used to describe the demographic characteristics of the study population at baseline as well as use of the different intervention sessions. Potential differences between the 3 study conditions at baseline were examined using analyses of variance (ANOVA) with Tukey post hoc tests for continuous variables and chi-square tests with Bonferroni correction for categorical variables.

Difference in use between the video and text intervention was assessed using ANCOVA. Another ANCOVA was performed to examine differences in use between the video and text intervention per educational level. Difference in use between participants who were assigned to an intervention condition that matched or mismatched their preferred intervention delivery format was also assessed with an ANCOVA.

Linear regression analyses with interaction terms were performed to examine whether the intervention effects were moderated by (1) intervention use and (2) matching or mismatching intervention delivery format. Moderation of intervention use was examined by comparing the effects of the intervention conditions to the control condition. For matching, moderation was examined by comparing the effects of the 2 intervention conditions with one another. The effect analyses were conducted for each outcome variable separately (ie, BMI, dietary intake, physical activity). The regression analyses were further adjusted for potential confounders (eg, baseline behavior and baseline differences). In addition, all moderation analyses were performed with both the multiple imputation as a completers-only dataset.

Finally, a linear regression analysis with the enter method was carried out to assess which demographics and intervention appreciation variables predicted intervention use.

Results

Study Sample

The CONSORT-EHEALTH flowchart [55] shows the use of the intervention and participation throughout the study per study condition (see Figure 1). In total, 1419 participants completed the baseline questionnaire; at 6-month follow-up, data were collected for 1015 (71.53%) participants. Of the participants who completed the baseline questionnaire, only 328 of 465 (70.5%) participants in the video condition and 364 of 491 (74.1%) in the text condition also completed the first

intervention session. Moreover, only 44 of 465 (9.5%) participants in the video condition and 60 of 491 (12.2%) participants in the text condition followed all intervention sessions. Overall, only 10.9% (N=956) of the participants completed all 6 sessions of the intervention.

Participants' mean age was 48.12 (SD 11.52) years and 831 of 1419 (58.56%) participants were female (see Table 1). The distribution of educational level between the 3 study conditions differed significantly (χ^2_4=10.3, *P*=.004) at baseline (see Table 1). The number of participants with a low educational level was significantly higher in the control condition compared to the text condition. Moreover, the number of participants with a medium educational level was significantly higher in the text and control condition in comparison to the video condition. In addition, compared to the control condition, significantly more participants in the video condition had a high educational level.

Table 1. Characteristics of the study sample and differences between the study conditions.

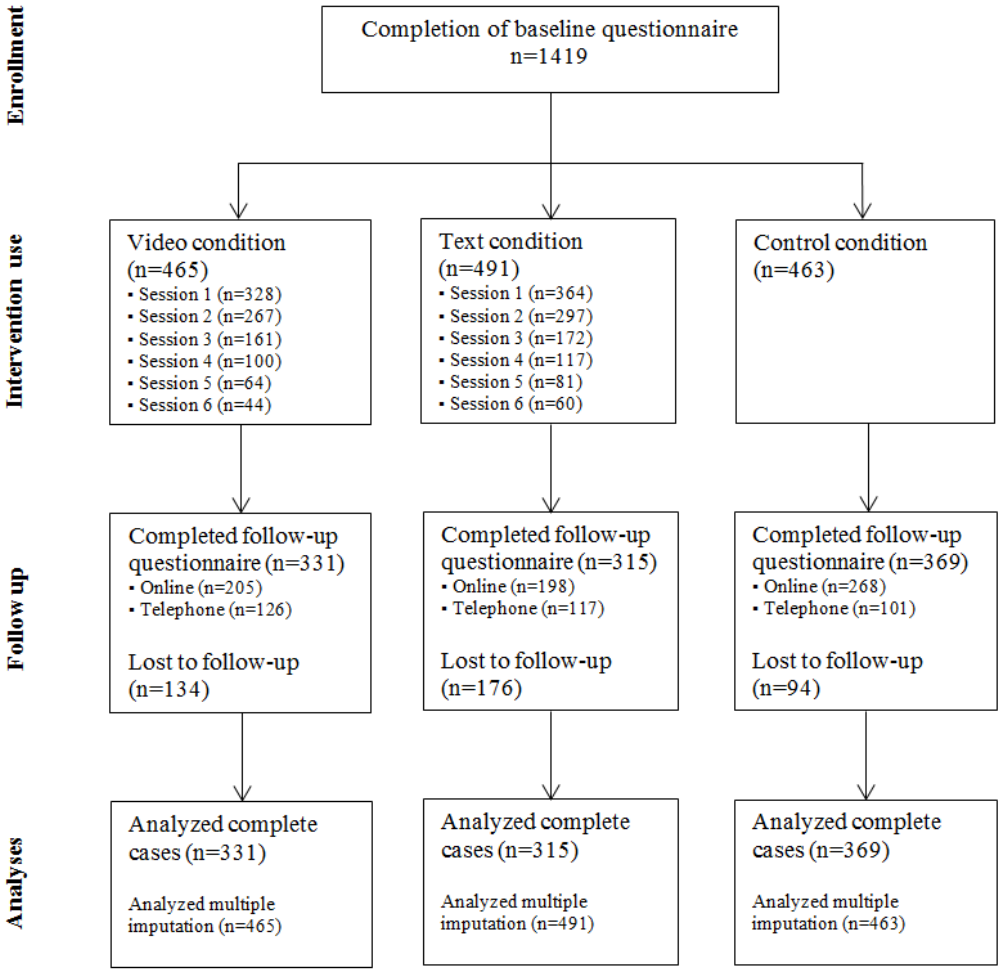
Baseline characteristics	Full sample (N=1419)	Video (N=465)	Text (N=491)	Control (n=463)	*F* (*df1*,*df2*)	χ^2 (*df*)	*P*
Baseline							
Gender (female), n (%)	831 (58.56)	273 (58.7)	284 (57.8)	274 (59.2)		0.2 (2)	.91
Educational level, n (%)						10.4 (4)	.004
Low	214 (15.08)	75 (16.1)	67 (13.6)[a]	72 (15.6)[a]			
Medium	436 (30.73)	118 (25.4)[a,b]	161 (32.8)[a]	157 (33.9)[b]			
High	769 (54.19)	272 (58.5)[a]	263 (53.6)	234 (50.5)[a]			
Age, mean (SD)	48.12 (11.52)	48.06 (12.05)	47.84 (11.58)	48.50 (10.92)	0.40 (2,2415)		.67
BMI, mean (SD)	26.42 (2.33)	26.43 (2.25)	26.45 (2.37)	26.37 (2.38)	0.13 (2,2348)		.88
Average daily minutes moderate and vigorous physical activity, mean (SD)	78.23 (83.40)	74.43 (73.27)	76.84 (81.11)	83.52 (94.51)	1.48 (2,2420)		.23
Average daily energy intake, mean (SD)	1296.91 (501.04)	1308.36 (490.37)	1314.70 (497.42)	1266.51 (515.07)	1.33 (2,2378)		.27
Intervention delivery format preference, n (%)						6.8 (4)	.34
Text only	579 (40.83)	194 (41.7)	206 (42.0)	179 (38.7)			
Video only	30 (2.12)	8 (1.7)	12 (2.4)	10 (2.2)			
Combination video/text	489 (34.48)	162 (34.8)	175 (35.7)	152 (32.8)			
No preference	320 (22.56)	101 (21.7)	97 (19.8)	122 (26.3)			
Matching intervention delivery format, n (%)							
Match	368 (48.61)	162 (44.5)[a]	206 (52.4)[a]	—		4.7 (2)	.03
Mismatch	389 (51.39)	202 (55.5)[a]	187 (47.6)[a]	—			
Follow-up	(n=1015)	(n=331)	(n=315)	(n=369)			
BMI, mean (SD)	25.99 (2.57)	25.87 (2.32)	25.96 (2.87)	26.12 (2.52)			
Average daily minutes moderate and vigorous physical activity, mean (SD)	114.87 (109.68)	114.82 (100.55)	113.77 (97.46)	115.52 (120.48)			
Average daily energy intake, mean (SD)	1078.97 (448.27)	1017.68 (434.46)	992.91 (460.40)	1157,55 (436.07)			

[a, b] Values within a row with identical letters were significantly different as determined by chi-square tests with Bonferroni correction.

Most participants preferred to receive information in Web-based computer-tailored interventions via text only (40.80%, 579/1419) followed by a combination of video and text (34.46%, 489/1419). Only 30 of 1419 (2.11%) participants preferred to receive information via video only and 320 of 1419 (22.55%) participants had no preference regarding the delivery format. In total, 368 of 757 (48.6%) participants were assigned to an intervention condition that matched their preferred intervention delivery format. For example, 206 of 491 (52.4%) participants who were assigned to the text condition also preferred to receive information in Web-based computer-tailored interventions via text only. The distribution of matching and mismatching intervention delivery format between the 2 intervention

conditions differed significantly (χ^2_3=4.7, *P*=.03). Significantly more participants in the text condition had a match compared to the video condition and vice versa. This result was also found when we included participants who had no preference for a particular intervention delivery format in the matching group (χ^2_3=8.3, *P*=.004).

Figure 1. Flowchart of the enrollment, allocation, and participation of respondents.



Intervention Use

The mean number of completed sessions was 2.07 (SD 1.91) in the video condition and 2.22 (SD 1.97) in the text condition, but this difference was not statistically significant ($F_{1,910}$=1.55, *P*=.21). In the analyses stratified by level of education also, no significant differences were found in the average number of completed sessions between the text and video condition: low ($F_{1,127}$=0.00, *P*=.84), medium ($F_{1,258}$=2.47, *P*=.12), and high ($F_{1,503}$=0.23, *P*=.63) educational level.

Yet, there was a significant difference in use between participants who were allocated to an intervention condition that matched their preferred intervention delivery format and those with a mismatch ($F_{1,910}$=4.58, *P*=.03). The mean number of completed sessions was 2.24 (SD 1.96) among participants with a match, whereas the mean was 1.95 (SD 1.88) among those with a mismatch.

Influence of Intervention Use and Matching/Mismatching Intervention Delivery Format

There were no significant video and text condition intervention use interaction terms for any of the outcome variables (see Table 2). This implies that intervention use did not have a moderating impact on the intervention outcomes. In addition, we also did not find significant condition intervention delivery format interaction terms for any of the outcome variables (see Table 2). This implies that the intervention outcomes were not influenced by whether or not participants were allocated to an intervention condition that matched their preferred intervention delivery format. These results were found with both the complete cases dataset as the multiple imputation dataset.

Table 2. Interactions terms regarding intervention use and matching/mismatching intervention delivery format for complete cases data.

Interaction terms[a,b]	β	*P*
BMI		
Condition*match/mismatch	.024	.56
Video condition*intervention use	–.090	.32
Text condition*intervention use	–.084	.33
Energy intake		
Condition*match/mismatch	.096	.29
Video condition*intervention use	–.041	.27
Text condition*intervention use	–.062	.10
Physical activity		
Condition*match/mismatch	–.090	.33
Video condition*intervention use	.022	.83
Text condition*intervention use	–.023	.81

[a] The moderation of intervention use was examined by comparing the intervention conditions to the control condition.

[b] The moderation of match/mismatch was examined by comparing the intervention conditions with each other.

Determinants of Intervention Use

The determinant analysis showed that participants with a higher BMI were significantly more likely to use the intervention more often (see Table 3). In addition, participants who felt involved and supported by the intervention (ie, feelings of relatedness) were also significantly more likely to use the intervention more often. The explained variance of the regression model was 16.0%.

Table 3. Determinants of intervention use (number of completed sessions) as assessed by multiple linear regression analysis.

Determinants	β	*P*
Study condition	.062	.37
Age	.090	.23
Gender	.133	.08
Educational level		
Low vs medium	.012	.92
Low vs high	.094	.45
BMI	.177	.02
Average daily minutes moderate and vigorous physical activity	.087	.24
Average daily energy intake	–.042	.56
The feedback messages fit to my own situation	–.082	.47
The feedback messages were understandable	.015	.88
The feedback messages were useful	.117	.27
The feedback messages were interesting	–.131	.24
Overall grade intervention (from 1 to 10)	–.082	.49
Feelings of autonomy	.148	.10
Feelings of relatedness	.291	.047
Feelings of competence	.018	.90

Discussion

Principal Findings

The main aim of this study was to examine how the use and effectiveness of Web-based computer-tailored interventions can be improved. For this purpose, we first examined if the use of a Web-based computer-tailored obesity prevention intervention can be increased by using videos as a delivery format. Secondly, we examined if the delivery of intervention content via participants' preferred delivery format can increase intervention

use. The third study aim was to examine if this match as well as more intervention use were related to better intervention effects. The final study aim was to identify which sociodemographic factors and intervention appreciation variables predict intervention use.

Intervention use (ie, number of completed intervention sessions) declined rapidly over time in both versions of the intervention. Contradicting our hypothesis, the video version was not used more often than the text version by the total study population or among participants with a low educational level. However, the intervention was used more often among participants who received intervention content via their preferred intervention delivery format. Our results further indicate that more intervention use and a matching intervention delivery format had not resulted in better intervention effects. In general, the intervention was more likely to be used more often by participants with a high BMI and participants who felt involved and supported by the intervention.

No support was found for the hypothesis that providing intervention content via (mainly) videos would result in more intervention use. This is an interesting finding because a previous study into the efficacy of this intervention has shown that the video version was appreciated significantly better than the text version [23]. Hence, a better appreciation does not necessarily lead to more intervention use. This suggestion should, however, be nuanced in-line with the fact that the appreciation of an intervention is also influenced by many other factors, such as the usability of an intervention and participants' motivation to change [41]. In addition, our finding also contradicts a recent study that has shown that video-tailored feedback can result in more time spent on a Web-based computer-tailored physical activity intervention [30]. However, the findings of our study are in-line with 2 other studies that also concluded that the use of videos as delivery format has no effect on intervention adherence [22,56]. Hence, results thus far indicate that the use of videos may not be the most optimal solution to increase the use of Web-based computer-tailored interventions.

Our results further show that the use of these interventions can be slightly increased by delivering intervention content via users' preferred intervention delivery format. Although videos did not increase intervention use, using this or another delivery format to match it with participants' delivery format preference may increase intervention use. However, the potential of matching remains ambiguous because our study and a recent similar study has concluded that a matching intervention delivery format does not result in better intervention outcomes [56]. Therefore, future research should first provide a better indication about whether or not future interventions should offer participants a delivery format choice.

A possible explanation for the absence of moderation effects of intervention use and matching is the fact that the most important information relevant for achieving a successful behavior change was included in the first 2 intervention sessions. This information may have been sufficient to achieve behavior changes. Further, it has been suggested that the relationship between intervention use and health outcomes is curvilinear instead of linear, implying that there is a saturation point after which no further benefit will be obtained [57]. More is not always better and sometimes increasing requirements for participants can even have iatrogenic effects, such as lowered engagement [58]. However, there is also evidence that people need to be exposed to educational content multiple times before intervention effects can be expected [18,59-62]. For example, prolonged intervention use is necessary for learning and practicing skills over time. Hence, more research is needed to identify what the optimum number of intervention sessions is in terms of maximizing use and effects.

Overall, intervention use was low in both the video and text version. The steep decline in intervention use can possibly be explained by the fact that the intervention consisted of 6 information-rich sessions which required a high level of active involvement (eg, making plans and answering questions). For example, in session 2, participants had to answer approximately 25 questions and also make an action plan, which requires much cognitive effort. This probably was too demanding for participants and may have resulted in an overload and premature dropouts [11,41,63]. Another explanation could be the fact that the most important intervention content was included in the first 2 sessions. It is possible that the content of these sessions was sufficient for participants to enable them to change their behavior successfully. Hence, not all participants may have needed to use the last intervention sessions in which their behavior and weight goals were evaluated. These explanations are confirmed by previous research that has shown that people are primarily interested in a simple comparison of their behavior against the relevant guidelines. There is a lack of interest in behavior change counseling sessions that require a high level of active involvement [17,64]. As in all Web-based interventions, dropout can also be the cause of technical problems, such as errors on the website and slow video buffering [65-67]. For example, a study has shown that participants will quit an intervention when it takes more than 2 seconds to load a video, with each incremental delay of 1 second resulting in a 5.8% increase in dropout rate [67]. Finally, although the use of videos and tailoring can be considered sophisticated, recent technical developments, such as gamification and mHealth, may have raised users' expectations of new products [68-71]. Hence, the video intervention possibly did not consist of sufficient innovative and attractive characteristics. In conclusion, these findings imply that still more research is needed into strategies that can increase the use of Web-based computer-tailored interventions.

Our determinant analysis shows that intervention use can possibly be increased by creating feelings of relatedness. Participants who felt involved and supported by the intervention (ie, feelings of relatedness) were more likely to use the intervention more often. According to the SDT, high feelings of relatedness will increase people's intrinsic motivation to change and consequently make behavior changes more likely [43]. A recent study of the Web-based computer-tailored intervention has shown that the video version was evaluated significantly better on feelings of relatedness compared to the text version [23]. Hence, the use of videos is a potentially effective strategy to increase feeling of relatedness. Possibly,

participants feel more involved and supported by a video delivery format because a person is actually talking to them in the videos and because it is easier to show empathy via spoken words.

Strengths and Limitations

Our study is characterized by several limitations. The most important limitation of this study is the fact that intervention use was assessed by the number of completed sessions. Although it has been suggested that there is a high correlation between number of completed sessions and time spent on the intervention [12], this does not give any information about the exposure to and engagement with the intervention content. Other measures, such as use of specific pages and amount of information read, may give a better indication of actual intervention use [45,72]. Our measure for intervention use may not have been sensitive enough to find an effect of usage on the outcome measures. Hence, it is strongly recommended to include a more extensive measurement of intervention use in future studies examining Web-based computer-tailored interventions [7,21]. A second limitation concerns the measurement of preferred intervention delivery format. In contrast to directly asking participants for their preference, it has been suggested that it may be better to assess the preference strength on a scale ranging from low to high [73]. Third, it may have been better to examine the influence of matching/mismatching intervention delivery format by first stratifying for intervention delivery format preference before randomizing people to study conditions. Fourth, people with a low educational level were underrepresented in the study sample. However, this is a common finding in intervention studies because these people are difficult to recruit. In the statistical analyses, we have further corrected for this by including educational level as a covariate. Fifth, because of our applied randomization procedure participants were aware of the study condition to which they were allocated prior to completing the baseline questionnaire. This may have influenced participants' responses to the baseline measurement.

Despite these limitations and the fact that we did not find support for all our hypotheses, this study provides a valuable contribution to the required research into this area. For example, an important strength is that we used a relatively new strategy (ie, use of tailored videos) to examine if the use of Web-based computer-tailored interventions can be improved among people with a low educational level in particular. Another strength is the fact that the analyses with the multiple imputation data resulted in exactly the same findings as the analyses with the complete cases data. In addition, we also corrected for potential confounding variables by including differences at baseline and predictors of attrition as covariates in the statistical analyses.

Conclusions

The use of videos as delivery format of intervention content is not the solution to improve the use of Web-based computer-tailored interventions. Nevertheless, the use of these interventions can potentially be increased by providing intervention content via participants' preferred intervention delivery format and ensuring that participants feel involved and supported by the intervention. The finding that more intervention use was not associated with better intervention outcomes implies that an intervention does not necessarily have to consist of many information-rich sessions. It may be sufficient to develop only 2 sessions that include the most important information necessary for achieving a successful behavior change. However, because only a few participants completed 3 or more sessions, more research is needed to identify what the optimum number of intervention sessions is in terms of maximizing use and effects. Until these strategies have been identified, it is recommended to minimize the number of sessions in future Web-based computer-tailored interventions and include the most important information for achieving a successful behavior change in the first sessions.

Acknowledgments

The study was funded by ZonMw, the Netherlands Organization for Health Research and Development (grant number: 200110001).

Conflicts of Interest

HdV is the scientific director of Vision2Health, a company that licenses evidence-based innovative computer-tailored health communication tools. The other authors declare that they have no competing interests.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V1.6.2 [55].

[PDF File (Adobe PDF File), 145KB - jmir_v17i9e222_app1.pdf]

References

1. Brug J, Oenema A, Campbell M. Past, present, and future of computer-tailored nutrition education. Am J Clin Nutr 2003 Apr;77(4 Suppl):1028S-1034S [FREE Full text] [Medline: 12663313]
2. Tate DF, Finkelstein EA, Khavjou O, Gustafson A. Cost effectiveness of internet interventions: review and recommendations. Ann Behav Med 2009 Aug;38(1):40-45 [FREE Full text] [doi: 10.1007/s12160-009-9131-6] [Medline: 19834778]

3. 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]
4. Smit ES, Evers SM, de Vries H, Hoving C. Cost-effectiveness and cost-utility of Internet-based computer tailoring for smoking cessation. J Med Internet Res 2013;15(3):e57 [FREE Full text] [doi: 10.2196/jmir.2059] [Medline: 23491820]
5. Schulz DN, Smit ES, Stanczyk NE, Kremers SP, de Vries H, Evers SM. Economic evaluation of a web-based tailored lifestyle intervention for adults: findings regarding cost-effectiveness and cost-utility from a randomized controlled trial. J Med Internet Res 2014;16(3):e91 [FREE Full text] [doi: 10.2196/jmir.3159] [Medline: 24650860]
6. Magnée T, Burdorf A, Brug J, Kremers SP, Oenema A, van Assema P, et al. Equity-specific effects of 26 Dutch obesity-related lifestyle interventions. Am J Prev Med 2013 Jun;44(6):e57-e66. [doi: 10.1016/j.amepre.2012.11.041] [Medline: 23683991]
7. Bennett GG, Glasgow RE. The delivery of public health interventions via the Internet: actualizing their potential. Annu Rev Public Health 2009;30:273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]
8. Spittaels H, De Bourdeaudhuij I, Vandelanotte C. Evaluation of a website-delivered computer-tailored intervention for increasing physical activity in the general population. Prev Med 2007 Mar;44(3):209-217. [doi: 10.1016/j.ypmed.2006.11.010] [Medline: 17197015]
9. Leslie E, Marshall AL, Owen N, Bauman A. Engagement and retention of participants in a physical activity website. Prev Med 2005 Jan;40(1):54-59. [doi: 10.1016/j.ypmed.2004.05.002] [Medline: 15530581]
10. Glasgow RE. eHealth evaluation and dissemination research. Am J Prev Med 2007 May;32(5 Suppl):S119-S126. [doi: 10.1016/j.amepre.2007.01.023] [Medline: 17466816]
11. Schulz DN, Schneider F, de Vries H, van Osch LA, van Nierop PW, Kremers SP. Program completion of a web-based tailored lifestyle intervention for adults: differences between a sequential and a simultaneous approach. J Med Internet Res 2012;14(2):e26 [FREE Full text] [doi: 10.2196/jmir.1968] [Medline: 22403770]
12. Wangberg SC, Bergmo TS, Johnsen JK. Adherence in Internet-based interventions. Patient Prefer Adherence 2008;2:57-65 [FREE Full text] [Medline: 19920945]
13. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
14. Hansen AW, Grønbæk M, Helge JW, Severin M, Curtis T, Tolstrup JS. Effect of a Web-based intervention to promote physical activity and improve health among physically inactive adults: a population-based randomized controlled trial. J Med Internet Res 2012;14(5):e145 [FREE Full text] [doi: 10.2196/jmir.2109] [Medline: 23111127]
15. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
16. Lenert L, Muñoz RF, Stoddard J, Delucchi K, Bansod A, Skoczen S, et al. Design and pilot evaluation of an internet smoking cessation program. J Am Med Inform Assoc 2003;10(1):16-20 [FREE Full text] [Medline: 12509354]
17. Verheijden MW, Jans MP, Hildebrandt VH, Hopman-Rock M. Rates and determinants of repeated participation in a web-based behavior change program for healthy body weight and healthy lifestyle. J Med Internet Res 2007;9(1):e1 [FREE Full text] [doi: 10.2196/jmir.9.1.e1] [Medline: 17478410]
18. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of Web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]
19. Strecher VJ, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored web-based smoking cessation program: randomized controlled trial. J Med Internet Res 2008;10(5):e36 [FREE Full text] [doi: 10.2196/jmir.1002] [Medline: 18984557]
20. Schneider F, van Osch L, Schulz DN, Kremers SP, de Vries H. The influence of user characteristics and a periodic email prompt on exposure to an internet-delivered computer-tailored lifestyle program. J Med Internet Res 2012;14(2):e40 [FREE Full text] [doi: 10.2196/jmir.1939] [Medline: 22382037]
21. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
22. Stanczyk NE, Crutzen R, Bolman C, Muris J, de Vries H. Influence of delivery strategy on message-processing mechanisms and future adherence to a Dutch computer-tailored smoking cessation intervention. J Med Internet Res 2013;15(2):e28 [FREE Full text] [doi: 10.2196/jmir.2153] [Medline: 23388554]
23. Walthouwer M, Oenema A, Lechner L, de Vries H. Are there differences in effects and appreciation between a video and text web-based computer-tailored obesity prevention intervention? A randomized controlled trial. J Med Internet Res 2015 (forthcoming). [doi: 10.2196/jmir.4083]
24. Bos N, Olson J, Gergle D, Olson G, Wright Z. Effects of four computer-mediated communications channels on trust development. ACM 2002:135-140. [doi: 10.1145/503376.503401]

25. Walthouwer MJ, Oenema A, Soetens K, Lechner L, De Vries H. Systematic development of a text-driven and a video-driven web-based computer-tailored obesity prevention intervention. BMC Public Health 2013;13:978 [FREE Full text] [doi: 10.1186/1471-2458-13-978] [Medline: 24138937]
26. Vandelanotte C, Mummery WK. Qualitative and quantitative research into the development and feasibility of a video-tailored physical activity intervention. Int J Behav Nutr Phys Act 2011;8(1):70 [FREE Full text] [doi: 10.1186/1479-5868-8-70] [Medline: 21718546]
27. Alley S, Jennings C, Persaud N, Plotnikoff R, 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]
28. Lee JA. Effect of Web-based interactive tailored health videos on users' attention, interactivity, overall evaluation, preference and engagement. Proc Am Soc Info Sci Tech 2012 Jan 11;48(1):1-3. [doi: 10.1002/meet.2011.14504801317]
29. Stanczyk NE, Bolman C, Muris JW, de Vries H. Study protocol of a Dutch smoking cessation e-health program. BMC Public Health 2011;11:847 [FREE Full text] [doi: 10.1186/1471-2458-11-847] [Medline: 22059446]
30. Soetens KC, Vandelanotte C, de Vries H, Mummery KW. Using online computer tailoring to promote physical activity: a randomized trial of text, video, and combined intervention delivery modes. J Health Commun 2014 Dec;19(12):1377-1392. [doi: 10.1080/10810730.2014.894597] [Medline: 24749983]
31. Meis T, Gaie M, Pingree S, Boberg E, Patten C, Offord K, et al. Development of a tailored, Internet-based smoking cessation intervention for adolescents. Journal of Computer-Mediated Communication 2002;7(3). [doi: 10.1111/j.1083-6101.2002.tb00151.x]
32. Justin D, Sara K, Hui Z, Yuqing R, Robert E, Joseph A. Watching together: integrating text chat with video. ACM 2007:877-886. [doi: 10.1145/1240624.1240756]
33. Paasche-Orlow MK, Parker RM, Gazmararian JA, Nielsen-Bohlman LT, Rudd RR. The prevalence of limited health literacy. J Gen Intern Med 2005 Feb;20(2):175-184 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.40245.x] [Medline: 15836552]
34. Williams-Piehota P, Schneider TR, Pizarro J, Mowad L, Salovey P. Matching health messages to information-processing styles: need for cognition and mammography utilization. Health Commun 2003 Oct;15(4):375-392. [doi: 10.1207/S15327027HC1504_01] [Medline: 14527864]
35. Kiesler DJ, Auerbach SM. Optimal matches of patient preferences for information, decision-making and interpersonal behavior: evidence, models and interventions. Patient Educ Couns 2006 Jun;61(3):319-341. [doi: 10.1016/j.pec.2005.08.002] [Medline: 16368220]
36. Dijkstra A, Conijn B, de Vries H. A match-mismatch test of a stage model of behaviour change in tobacco smoking. Addiction 2006 Jul;101(7):1035-1043. [doi: 10.1111/j.1360-0443.2006.01419.x] [Medline: 16771895]
37. Swift JK, Callahan JL. The impact of client treatment preferences on outcome: a meta-analysis. J Clin Psychol 2009 Apr;65(4):368-381. [doi: 10.1002/jclp.20553] [Medline: 19226606]
38. King M. The effects of patients' and practitioners' preferences on randomized clinical trials. Palliat Med 2000 Nov;14(6):539-542. [Medline: 11219886]
39. Dijkstra A. The psychology of tailoring-ingredients in computer-tailored persuasion. Social Pers Psych Compass 2008 Mar;2(2):765-784. [doi: 10.1111/j.1751-9004.2008.00081.x]
40. Brouwer W, Oenema A, Crutzen R, de Nooijer NJ, de Vries NK, Brug J. An exploration of factors related to dissemination of and exposure to internet-delivered behavior change interventions aimed at adults: a Delphi study approach. J Med Internet Res 2008;10(2):e10 [FREE Full text] [doi: 10.2196/jmir.956] [Medline: 18417443]
41. Brouwer W, Oenema A, Crutzen R, de Nooijer J, de Vries N, Brug J. What makes people decide to visit and use an internet ‐ delivered behavior ‐ change intervention? Health Education 2009 Oct 16;109(6):460-473. [doi: 10.1108/09654280911001149]
42. Peels DA, de Vries VH, Bolman C, Golsteijn RJ, van Stralen MM, Mudde AN, et al. Differences in the use and appreciation of a web-based or printed computer-tailored physical activity intervention for people aged over 50 years. Health Educ Res 2013 Aug;28(4):715-731 [FREE Full text] [doi: 10.1093/her/cyt065] [Medline: 23784076]
43. Deci E, Ryan R. Self-determination theory. In: Lange PH, editor. Handbook of Theories of Social Psychology: Collection: Volumes 1 & 2 (Sage Social Psychology Program). Thousand Oaks, CA: Sage Publications Ltd; 2011:416-433.
44. Ryan R, Patrick H, Deci E, Williams G. Facilitating health behaviour change and its maintenance: interventions based on self-determination theory. European Health Psychologist 2008;10(1):2-5.
45. Brouwer W, Oenema A, Raat H, Crutzen R, de Nooijer J, de Vries NK, et al. Characteristics of visitors and revisitors to an Internet-delivered computer-tailored lifestyle intervention implemented for use by the general public. Health Educ Res 2010 Aug;25(4):585-595 [FREE Full text] [doi: 10.1093/her/cyp063] [Medline: 19897515]
46. de Vries H, Mudde A, Leijs I, Charlton A, Vartiainen E, Buijs G, et al. The European Smoking Prevention Framework Approach (EFSA): an example of integral prevention. Health Educ Res 2003 Oct;18(5):611-626 [FREE Full text] [Medline: 14572020]
47. Leventhal H, Leventhal EA, Contrada RJ. Self-regulation, health, and behavior: A perceptual-cognitive approach. Psychology & Health 1998 Jul;13(4):717-733. [doi: 10.1080/08870449808407425]

48. Zimmerman B. Attaining self-regulation: A social cognitive perspective. In: Boekaerts M, Pintrich PR, Zeidner M, editors. Handbook of Self-Regulation. San Diego, CA: Academic Press; 2000:13-39.
49. Gollwitzer PM. Implementation intentions: strong effects of simple plans. American Psychologist 1999;54(7):493-503. [doi: 10.1037/0003-066X.54.7.493]
50. van Assema P, Brug J, Ronda G, Steenhuis I. The relative validity of a short Dutch questionnaire as a means to categorize adults and adolescents to total and saturated fat intake. J Hum Nutr Diet 2001 Oct;14(5):377-390. [Medline: 11906579]
51. RIVM. National Institute for Public Health and the Environment, Ministry of Health, Welfare and Sport. 2013. Dutch Food Composition Dataset (NEVO) URL: http://www.rivm.nl/en/Topics/D/Dutch_Food_Composition_Database [accessed 2015-09-09] [WebCite Cache ID 6bPvZJU2E]
52. Wendel-Vos GC, Schuit AJ, Saris WH, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. J Clin Epidemiol 2003 Dec;56(12):1163-1169. [Medline: 14680666]
53. UNESCO Institute for Statistics. International Standard Classification of Education. Montreal: UNESCO Institute for Statistics; 2011.
54. Rosnow RL, Rosenthal R. Statistical procedures and the justification of knowledge in psychological science. American Psychologist 1989;44(10):1276-1284. [doi: 10.1037/0003-066X.44.10.1276]
55. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
56. Vandelanotte C, Duncan MJ, Plotnikoff RC, Mummery WK. Do participants' preferences for mode of delivery (text, video, or both) influence the effectiveness of a Web-based physical activity intervention? J Med Internet Res 2012;14(1):e37 [FREE Full text] [doi: 10.2196/jmir.1998] [Medline: 22377834]
57. 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;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
58. Glasgow RE, Nelson CC, Kearney KA, Reid R, Ritzwoller DP, Strecher VJ, et al. Reach, engagement, and retention in an Internet-based weight loss program in a multi-site randomized controlled trial. J Med Internet Res 2007;9(2):e11 [FREE Full text] [doi: 10.2196/jmir.9.2.e11] [Medline: 17513282]
59. Dijkstra A, de Vries H, Roijackers J, van Breukelen G. Tailoring information to enhance quitting in smokers with low motivation to quit: three basic efficacy questions. Health Psychol 1998 Nov;17(6):513-519. [Medline: 9848801]
60. Brug J, Glanz K, van Assema AP, Kok G, van Breukelen GJ. The impact of computer-tailored feedback and iterative feedback on fat, fruit, and vegetable intake. Health Educ Behav 1998 Aug;25(4):517-531. [Medline: 9690108]
61. 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;12(4):e52 [FREE Full text] [doi: 10.2196/jmir.1430] [Medline: 21087922]
62. An LC, Perry CL, Lein EB, Klatt C, Farley DM, Bliss RL, et al. Strategies for increasing adherence to an online smoking cessation intervention for college students. Nicotine Tob Res 2006 Dec;8 Suppl 1:S7-12. [Medline: 17491165]
63. Prochaska JO. Multiple Health Behavior Research represents the future of preventive medicine. Prev Med 2008 Mar;46(3):281-285. [doi: 10.1016/j.ypmed.2008.01.015] [Medline: 18319100]
64. Verheijden MW, Jans MP, Hildebrandt VH. Web-based tailored lifestyle programs: exploration of the target group's interests and implications for practice. Health Promot Pract 2008 Jan;9(1):82-92. [doi: 10.1177/1524839906289515] [Medline: 16816030]
65. Booth AO, Nowson CA, Matters H. Evaluation of an interactive, Internet-based weight loss program: a pilot study. Health Educ Res 2008 Jun;23(3):371-381 [FREE Full text] [doi: 10.1093/her/cyn007] [Medline: 18349032]
66. Kypri K, McAnally HM. Randomized controlled trial of a web-based primary care intervention for multiple health risk behaviors. Prev Med 2005;41(3-4):761-766. [doi: 10.1016/j.ypmed.2005.07.010] [Medline: 16120456]
67. Krishnan SS, Sitaraman RK. Video stream quality impacts viewer behavior: inferring causality using quasi-experimental designs. IEEE/ACM Trans Networking 2013 Dec;21(6):2001-2014. [doi: 10.1109/TNET.2013.2281542]
68. Blohm I, Leimeister JM. Gamification. Wirtschaftsinf 2013 Jun 14;55(4):275-278. [doi: 10.1007/s11576-013-0368-0]
69. Bert F, Giacometti M, Gualano MR, Siliquini R. Smartphones and health promotion: a review of the evidence. J Med Syst 2014 Jan;38(1):9995. [doi: 10.1007/s10916-013-9995-7] [Medline: 24346929]
70. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
71. PwC. Emerging mHealth: Paths for Growth. Delaware: PwC; 2012. URL: https://www.pwc.com/gx/en/healthcare/mhealth/assets/pwc-emerging-mhealth-full.pdf [accessed 2015-09-09] [WebCite Cache ID 6bPxWmRLY]
72. Danaher BG, Seeley JR. Methodological issues in research on web-based behavioral interventions. Ann Behav Med 2009 Aug;38(1):28-39 [FREE Full text] [doi: 10.1007/s12160-009-9129-0] [Medline: 19806416]
73. Swift JK, Callahan JL, Vollmer BM. Preferences. J Clin Psychol 2011 Feb;67(2):155-165. [doi: 10.1002/jclp.20759] [Medline: 21120917]

Abbreviations

BMI: body mass index
RCT: randomized controlled trial
SDT: Self-Determination Theory
SQUASH: Short Questionnaire to Assess Health-Enhancing Physical Activity

Edited by G Eysenbach; submitted 02.04.15; peer-reviewed by C Vandelanotte, Y Wang; comments to author 07.05.15; revised version received 23.06.15; accepted 08.07.15; published 25.09.15.

Please cite as:
Walthouwer MJL, Oenema A, Lechner L, de Vries H
Use and Effectiveness of a Video- and Text-Driven Web-Based Computer-Tailored Intervention: Randomized Controlled Trial
J Med Internet Res 2015;17(9):e222
URL: http://www.jmir.org/2015/9/e222/
doi:10.2196/jmir.4496
PMID:26408488

Original Paper

Designing and Testing an Inventory for Measuring Social Media Competency of Certified Health Education Specialists

Julia M Alber[1], MPH, PhD; Jay M Bernhardt[2], MPH, PhD; Michael Stellefson[3], PhD; Robert M Weiler[4], MPH, PhD; Charkarra Anderson-Lewis[5], MPH, PhD; M David Miller[6], PhD; Jann MacInnes[6], PhD

[1]Center for Health Behavior Research, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[2]Center for Health Communication, Moody College of Communication, University of Texas, Austin, TX, United States

[3]Center for Digital Health and Wellness, Department of Health Education and Behavior, University of Florida, Gainesville, FL, United States

[4]Department of Global and Community Health, George Mason University, Fairfax, VA, United States

[5]University of Southern Mississippi, Department of Public Health, Hattiesburg, MS, United States

[6]School of Human Development and Organizational Studies in Education, University of Florida, Gainesville, FL, United States

Corresponding Author:
Julia M Alber, MPH, PhD
Center for Health Behavior Research
Perelman School of Medicine
University of Pennsylvania
110 Blockley Hall
423 Guardian Drive
Philadelphia, PA, 19104
United States
Phone: 1 215 573 9894
Fax: 1 215 573 5315
Email: alberj@upenn.edu

Abstract

Background: Social media can promote healthy behaviors by facilitating engagement and collaboration among health professionals and the public. Thus, social media is quickly becoming a vital tool for health promotion. While guidelines and trainings exist for public health professionals, there are currently no standardized measures to assess individual social media competency among Certified Health Education Specialists (CHES) and Master Certified Health Education Specialists (MCHES).

Objective: The aim of this study was to design, develop, and test the Social Media Competency Inventory (SMCI) for CHES and MCHES.

Methods: The SMCI was designed in three sequential phases: (1) Conceptualization and Domain Specifications, (2) Item Development, and (3) Inventory Testing and Finalization. Phase 1 consisted of a literature review, concept operationalization, and expert reviews. Phase 2 involved an expert panel (n=4) review, think-aloud sessions with a small representative sample of CHES/MCHES (n=10), a pilot test (n=36), and classical test theory analyses to develop the initial version of the SMCI. Phase 3 included a field test of the SMCI with a random sample of CHES and MCHES (n=353), factor and Rasch analyses, and development of SMCI administration and interpretation guidelines.

Results: Six constructs adapted from the unified theory of acceptance and use of technology and the integrated behavioral model were identified for assessing social media competency: (1) Social Media Self-Efficacy, (2) Social Media Experience, (3) Effort Expectancy, (4) Performance Expectancy, (5) Facilitating Conditions, and (6) Social Influence. The initial item pool included 148 items. After the pilot test, 16 items were removed or revised because of low item discrimination (r<.30), high interitem correlations (P>.90), or based on feedback received from pilot participants. During the psychometric analysis of the field test data, 52 items were removed due to low discrimination, evidence of content redundancy, low R-squared value, or poor item infit or outfit. Psychometric analyses of the data revealed acceptable reliability evidence for the following scales: Social Media Self-Efficacy (alpha=.98, item reliability=.98, item separation=6.76), Social Media Experience (alpha=.98, item reliability=.98, item separation=6.24), Effort Expectancy(alpha =.74, item reliability=.95, item separation=4.15), Performance Expectancy (alpha =.81, item reliability=.99, item separation=10.09), Facilitating Conditions (alpha =.66, item reliability=.99, item separation=16.04),

and Social Influence (alpha =.66, item reliability=.93, item separation=3.77). There was some evidence of local dependence among the scales, with several observed residual correlations above |.20|.

Conclusions: Through the multistage instrument-development process, sufficient reliability and validity evidence was collected in support of the purpose and intended use of the SMCI. The SMCI can be used to assess the readiness of health education specialists to effectively use social media for health promotion research and practice. Future research should explore associations across constructs within the SMCI and evaluate the ability of SMCI scores to predict social media use and performance among CHES and MCHES.

(J Med Internet Res 2015;17(9):e221) doi:10.2196/jmir.4943

KEYWORDS

social media; health education; professional competence

Introduction

Background

Social media, or "user-generated content utilizing Internet-based publishing technologies, distinct from traditional print and broadcast media," [1] has become popular for professional, personal, and promotional use. Social media is used to connect with and communicate bidirectionally with friends, coworkers, and family [1]. Social media offers an array of tools for connecting people and sharing content, such as social networking sites (eg, Facebook and Twitter), photo-sharing sites (eg, Flickr and Instagram), and video-sharing sites (eg, YouTube and Vimeo). Compared to other types of print and broadcast media, social media is unique in that it facilitates two-way communication that allows organizations to personalize content and engage with communities and the public. Tailoring and personalizing health messages through social media can increase both the relevance of the information distributed and attention paid to the communication by the recipients [2]. Such tailoring can result in a greater impact on the intended behavior [2]. As of 2014, 74% of adult Internet users report using social media sites [3]. Thus, social media has immense potential as a medium for organizations and individuals to reach a wide range of demographic groups based on age, gender, and race/ethnicity [4].

The Role of Social Media in Public Health Education

Social media is used to facilitate collaboration and engagement among health education professionals and the public in order to promote healthy behaviors [5-12]. Social media can engage and empower both communities and individuals to make healthier choices by helping to connect them to resources and facilitating collaboration between them to advocate for policies and programs that impact their health [13]. In a 2012 study, researchers found that approximately 60% of state health departments used at least one type of social media to meet their organizational objectives [14]. As an increasing number of health education organizations continue to take advantage of social media, use of these tools will generate numerous opportunities for influencing and changing health behavior [15-18].

Because health education specialists play a significant role in the dissemination of health information and the promotion of healthy behaviors [19], it is crucial for health education professionals to be able to capitalize upon the capabilities of different media to successfully distribute information and reach target populations [20]. The specific professional roles and duties of health education specialists are described in the document, *Seven Areas of Responsibility and Competencies for Health Education Specialists* [19]. Certified Health Education Specialists (CHES) and Master Certified Health Education Specialists (MCHES) are health education specialists who have successfully passed the CHES or MCHES examination. These examinations, administered by the National Commission for Health Education Credentialing, Inc (NCHEC), are competency-based assessments of the knowledge, application, and understanding of the *Seven Areas of Responsibility* [21]. The CHES examination reflects the entry-level sub-competencies of the *Seven Areas of Responsibility*, while the MCHES encompasses both entry- and advanced-level sub-competencies [22]. The *Seven Areas of Responsibility* provides a foundation of competencies that CHES and MCHES can use to effectively learn and apply social media technology for health education research and practice. Many of the responsibilities outlined in this document can be carried out through the use of social media. For example, *Area of Responsibility VI, Competency 6.1: Obtain and Disseminate Health-Related Information*, could be carried out by using Twitter or another social media platform to disseminate health information to a particular population [19]. Social media can be employed by health education specialists to not only provide access to reliable health information, but to also tailor and personalize health messages and content to individuals (Competency 7.2) [13,19]. Social media can also assist with empowering people to make healthier and safer decisions and facilitate participation (Competency 7.3) [13,19]. Social media can bring together members of communities (eg, diabetes patients), who may be dispersed across a city, a state, a nation, or the world, to provide mutual support and to work toward a common solution (Competency 2.1) [13,19]. One well-known application of social media for health promotion is the Centers for Disease Control and Prevention's (CDC) *Tips From Former Smokers* campaign. This campaign used Web-based videos, buttons and badges, images, and podcasts to share real-life stories of individuals living with smoking-related health issues. As a result of this campaign, free smoking cessation resources were disseminated to users, an estimated 1.64 million Americans attempted to stop smoking, and 6 million nonsmokers discussed the dangers associated with smoking with friends and family [23].

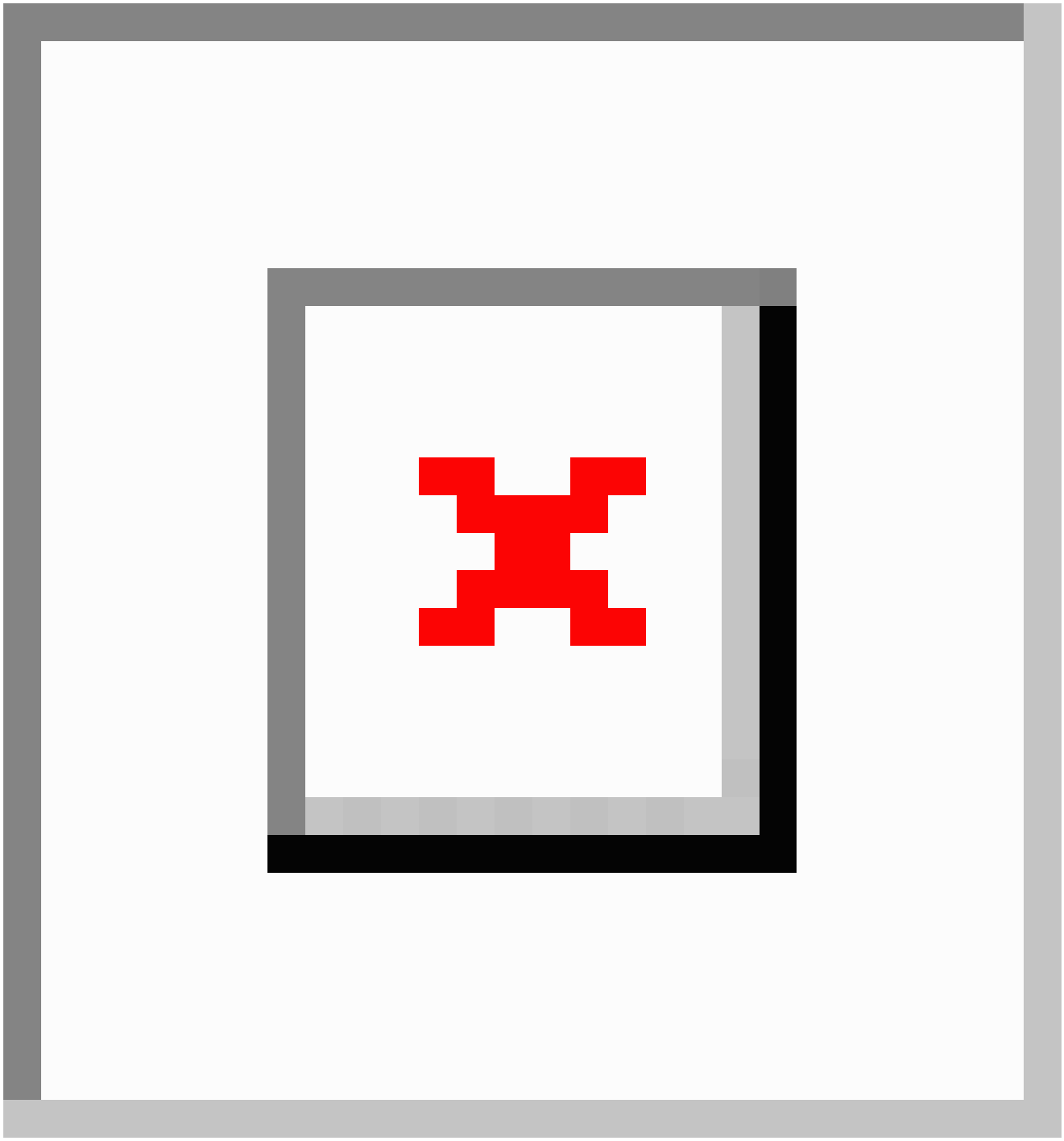
While there are commonly accepted principles that guide social media use and training for public health practitioners [13,24], there are no existing standardized measures of social media competency among health education specialists. Korda and Itani [15] stressed that social media implementation "require[s] careful application and may not always achieve their desired outcomes." Although public health research has illustrated promising applications of social media in practice [25], there are potential dangers or issues associated with using social media for health communication, such as sharing of misleading or inaccurate information or the violation of the privacy of clients or research participants [26]. Prior research has explored organizational uses of social media in health education settings [14,27,28]. Moreover, guidelines and best practices exist for planning, implementing, and evaluating social media activities in public health [5,13,24,29-31], but there is no research that has measured training or educational needs for health education professionals who are increasingly using social media to satisfy their occupational responsibilities. For this purpose, the objective of this study was to design, develop, and test a social media competency inventory for CHES and MCHES. The intended use of the inventory was for the assessment of workforce needs to inform the development of future training, educational programs, and organizational policies.

Methods

Overview

The design, development, and testing of the new measure for social media competency entailed three overarching phases: Phase 1, Conceptualization and Domain Specifications; Phase 2, Item Development; and Phase 3, Inventory Testing and Finalization. These phases and their corresponding steps were based on Crocker and Algina's 10-step process of test construction [32]. Approval from the lead researcher's (JA) university Institutional Review Board was obtained prior to beginning this study. Figure 1 depicts a sequential overview of the research activities that occurred within each of the three phases.

Figure 1. Outline of methods for designing, developing, and testing the Social Media Competency Inventory.



Phase 1: Conceptualization and Domain Specifications

Defining and Operationalizing Social Media Competency

The term "social media competency" was not previously defined in the literature; therefore, a review of the literature was conducted using Google Scholar, PubMed, the Cumulative Index to Nursing and Allied Health Literature (CINAHL), and the Education Resources Information Center (ERIC) using a combination of the following keyword search terms: competency, competency model, competence, competency framework, professional ability, successful use, performance, professional readiness, employee, information technology, social media, social network, Web-based technologies, new media, digital health, technology, Web 2.0, and eHealth.

Once social media competency had been defined and operationalized, observable behaviors that characterized the specific constructs to be measured in the inventory were identified using the literature and professional guidelines. More specifically, the following terms were searched on three databases (Google Scholar, PubMed, and CINAHL): health education, health promotion, health behavior, prevention, use, guides, practice, research, competency, ability, knowledge, attitudes, readiness, effective use, information technology, social media, social network, Web-based technologies, new media, digital health, technology, Web 2.0, and eHealth. Additionally, leading health organizations' websites were searched for guidelines and recommendations related to the responsibilities of health education specialists and evidence-based social media

practices. Organizations included the American Public Health Association (APHA), Centers for Disease Control and Prevention, National Institutes of Health (NIH), National Commission for Health Education Credentialing, Inc, the Society for Public Health Education (SOPHE), and the US Department of Health & Human Services (HHS). Using these sources, a list of potential and actual social media tasks completed by health education specialists was drafted. These tasks were compared with the entry-level sub-competencies outlined in the *Seven Areas of Responsibility*. Sub-competencies that could be met using social media were revised to incorporate social media. For example, one sub-competency is related to collecting primary health data (Sub-competency 1.3.1) [19]. Using this sub-competency, the following observable behavior was created: "Collect primary data (using survey or other method to collect data directly from social media) related to health using social media." As the inventory was intended for both CHES and MCHES, only entry-level sub-competencies were included in the development of the observable behaviors. The observable behaviors were organized according to each of the *Seven Areas of Responsibility for Health Education Specialists* [19]. This organization system was used to ensure that each of the seven areas would be adequately reflected in the final inventory, thus allowing the inventory to be directly linked to the key responsibilities of health education specialists.

Observable Behaviors: Expert Review and Revisions

A panel of four experts was asked to review the initial list of observable behaviors. The panel included three content experts and one measurement expert. The three content experts had worked in the field of health education for a minimum of five years, and had extensive knowledge on using social media technologies in health education research. One of the content experts is also an MCHES. The measurement expert was a research methodologist with vast experience in psychometrics and large-scale measurement and evaluation. The experts were sent a Web-based survey using Qualtrics survey software. The Web-based survey contained observable behaviors organized by the *Seven Areas of Responsibility for Health Education Specialists*. Experts were asked to indicate which behaviors should be kept or removed. Provided textboxes allowed experts to suggest additional observable behaviors for consideration.

After the initial expert review, the list of observable behaviors was revised and sent back to the experts. Experts were then asked to create a rank order of the remaining observable behaviors in each of the *Seven Areas of Responsibility for Health Education Specialists*. SPSS version 22.0 was used to calculate median ranks and interquartile ranges to determine the most important observable behaviors ranked by the experts.

Domain Specifications for Each Scale

Domain specifications were developed for each of the constructs identified as a function of the literature and expert review. For each construct, a table of specifications was developed to outline the content areas, the relevant learning domain levels (ie, stages of the affective domains, levels of the cognitive domain), and the representation of each of these elements across the scale. The domain specifications for each construct were again sent to the experts to review using a Web-based survey programmed into Qualtrics. Domain specifications were revised based on the experts' comments and feedback.

Phase 2: Item Development

Overview

Development of the items, item stems, and instructions were guided by recommendations from several survey methods resources [33-36]. An initial list of items was drafted using the list of observable behaviors and the representation of items outlined in the domain specifications. The number of items was based on the proportion of items in each cell of the table of specifications per construct, and by approximately doubling the number of items needed for the final scale for each construct [37].

Expert Review of Items and Item Revisions

Experts were asked to review items using a Web-based survey administered on Qualtrics. Within the survey, the experts were asked to evaluate the following characteristics by whether or not each item adequately reflected each characteristic (yes/no): brevity, focus, clarity, assurance, readability, and adequacy of response options [33,38]. Experts were provided with the definition of each of these characteristics. If an expert selected "no" for any item criteria, they were asked to explain why using a textbox provided below each item. Experts were also asked if any of the items could be perceived as biased or leading and were again given the opportunity to make comments in a textbox. Next, using the domain specifications for each construct as a reference, experts were asked to respond to the following question: "Overall, would you say the items in this section are representative of the universe of all possible questions related to [social media use in the health education construct]?" Finally, experts were provided the opportunity to make suggestions or propose changes related to the items and instructions included in the item pool. The items and the instructions were revised based on the feedback obtained by the experts.

Think-Aloud Sessions

Think-aloud sessions were conducted with CHES/MCHES using the revised item list. A think-aloud session is a type of cognitive interviewing method commonly employed for pilot-testing instruments to better understand the mental processes participants use to answer items [39]. A think-aloud session invites the participants to describe their thought processes aloud when responding to questions or reading instructions [39]. A purposive sample of CHES and MCHES (n=22) were invited to participate in the think-aloud sessions via LinkedIn [40], a professional social networking site [41-43]. To identify potential participants, the lead researcher (JA) entered "CHES" and "MCHES" into the LinkedIn search bar and invited all professionals (n=22) who appeared in the search results using either their listed email addresses (n=14) or the LinkedIn mailbox tool (n=8). Think-aloud sessions were completed over the phone (n=5) or in person (n=5). Participants were asked to open the Web-based survey containing the revised item list, read the items, and speak aloud as they responded to each item, describing how they came to the decision to answer each question. In addition, participants were asked if they had problems answering the items, if any of the questions were

frustrating or confusing, or if any questions could be perceived as offensive. Each session lasted approximately 45 minutes. While the sessions were not recorded, detailed notes were taken during each session. The participants did not receive an incentive for participating in a session. A thematic analysis [44] of the qualitative data from the think-aloud sessions was used to identify the reoccurring problems or issues experienced by participants when completing the assessment. The instrument was revised based on findings from the thematic analysis.

Pilot Test and Revisions

A random sample of CHES and MCHES (n=400) were emailed a link to the Web-based survey. The pilot-test data were analyzed using SPSS version 22.0 with listwise deletion used for handling missing data. Cronbach alpha values were calculated for each scale of the inventory to provide a measure of internal consistency for each construct [45]. Bivariate Spearman rank correlations (ρ) were calculated to assess associations between items in each scale to identify any extremely high correlations (ρ>.90). When an extremely high correlation was found between items, items were considered to be repetitive and unnecessary and thus removed from the pool [46]. The frequency of response options was explored to determine if each response option was being used in each scale. To examine item discriminations, corrected item-total correlations (r) were computed for each item within each scale [47]. The instrument was again revised using the results from Spearman rank and item-total correlation analyses.

Phase 3: Inventory Testing and Finalization

Field Test

Participants were recruited from NCHEC's database of CHES and MCHES (N=10,073). A random sample of CHES and MCHES (n=1000) from the database were sent the link to the instrument embedded in a Web-based survey via a mailed letter and in an email. Three emails were also sent to each participant reminding them to complete the survey at their earliest convenience. Participants were given US $1.00 in the mailed letter, as this has been shown to increase response rates in Web-based survey research [34]. The first 100 participants to complete the survey also received a US $10.00 Amazon gift card. The Web-based survey included the final items on the instrument, as well as demographic and organizational items. Demographic questions on age, sex, race and ethnicity, highest degree obtained, and household income were adapted from the Behavioral Risk Factor Surveillance System (BRFSS) questionnaire [48]. Organizational questions were adapted from items used in another study of CHES assessing Internet and social media access at work and years of experience in the health education profession [27].

Psychometric Analyses and Item Removal

The analyses of field data were conducted using three procedures: (1) classical test theory procedures, (2) factor analyses, and (3) Rasch analyses. Each construct in the instrument was analyzed separately.

Classical test theory procedures were executed using SPSS version 22.0. For each scale, bivariate interitem correlations, corrected item-total correlations, and Cronbach alpha statistics were computed [47]. Items with corrected item-total correlations below .30 were removed [49]. For each scale, Cronbach alpha was compared to the generally acceptable standard of .70 or higher [45].

Using Mplus Editor 7, categorical confirmatory factor analyses (CCFAs) were conducted for each construct's data to examine fit to a unidimensional model. CCFAs were conducted using weighted least-squares means and variance adjusted (WLSMV), an estimator suggested for noncontinuous data that is robust for nonnormal data [50]. Acceptable model fit is indicated when comparative fit index (CFI) values are greater than .90, root mean square error of approximation (RMSEA) values are .05 or less, chi-square test of model fit values are not statistically significant, and Tucker-Lewis index (TLI) values are less than .90 [51]. Items with low R^2 values (variance explained) or low parameter estimates were removed from the instrument. In this study, CCFA model fit was examined by (1) checking model fit indices (eg, CFI and TLI), (2) ensuring statistically significant (P<.05) parameter estimates for the path of the specified model, and (3) confirming that the magnitude of the parameter estimates are consistent with the theorized model [52].

Following CCFA, the data were analyzed under a Rasch framework, specifically the rating scale model (RSM), using the computer program jMetrik [53]. The RSM tests the probability that a person with a particular ability level will select a particular category (or response option) given a specific threshold and item difficulty level [54]. The assumptions of RSM are local independence of items, unidimensionality, and monotonicity [54]. For Likert-type data, RSMs are commonly applied [53], particularly when selecting higher response options is believed to correspond to higher ability, and that the probability of moving from one option response to the next is the same relative to item difficulty across the items [55]. RSMs can reduce the number of estimated parameters compared to less constrained models (eg, partial credit model), can assist in reducing a large number of items originally developed for a scale, and requires a lower sample size than some alternative models. As the instrument developed included six different constructs, six RSMs were fitted to the data: one RSM for each of the scales.

Local independence, monotonicity, item and category infit or outfit, item difficulty, item characteristic curves (ICCs), item reliability, item separation indices, and threshold values were all examined. Local independence was investigated by examining the correlations between the item residuals [56]. Bivariate correlations between item residuals are recommended to be below |.20|, however, these correlations should be considered relative to all correlations [57]. Monotonicity was assessed by examining ICCs and determining if the threshold values increased in order (ie, higher response options had higher threshold values). Threshold values were examined to see whether or not they increased in order, and whether or not the distance between each threshold was between 1.00 and 5.00 logits [58]. For each item's ICC, the curve that represented the lowest category was checked to ensure it was the furthest to the left, and the curve that represented the highest category was

furthest to the right. Each category (or response option) was examined to ensure it had the highest probability for being observed at some point on the latent continuum. Item reliability, which provides an estimate for the quality of the item placement within an order of items along the latent trait, should have a value of .80 or greater [59]. Item separation, which provides an estimate of the quality of being able to locate items on the latent trait, should be 2.00 or greater [53,55]. Items with outfit and infit values more than 0.50 logits outside of the recommended values of 0.50 to 1.50 logits were removed from the instrument [58].

Guidelines Development

The reliability and validity evidence obtained during the preceding steps was used to establish general guidelines for administering, analyzing, and interpreting the final inventory. An evaluation was conducted to examine evidence of construct validity, internal structure, response process validity, external validity, and predictive validity of the inventory. Scoring and interpretation guidelines were created using recommendations from Crocker and Algina [32] and Osterlind [60].

Results

Phase 1: Conceptualization and Domain Specifications

Overview

Based on the review of literature and professional guidelines, social media competency was defined, in the context of health education, as, "the user's potential to apply social media technologies to disseminate health information and messages, engage and empower individuals to make healthier decisions, and encourage conversation and participation related to the mission of their health organization." Six core constructs were identified as important for assessing social media competency: (1) Social Media Self-Efficacy, (2) Social Media Experience, (3) Effort Expectancy, (4) Performance Expectancy, (5) Facilitating Conditions, and (6) Social Influence. These constructs were identified using a technology competency model framework [61] and constructs from the integrated behavioral model (IBM) [62] and the unified theory of acceptance and use of technology (UTAUT) [63]. Social Media Self-Efficacy is an individual's confidence in their ability to use social media technologies, as a function of their employment, to meet their employer's needs as well as to reach and engage the public. Social Media Experience includes actions or tasks completed by the individual related to social media, social media websites, and tools that exist and are utilized for professional purposes in health education. Effort Expectancy is an individual's perceptions of the ease of using social media while at work. Performance Expectancy is one's beliefs about the impact of social media on their job performance. Facilitating Conditions refers to an individual's beliefs regarding the existence of technical and organizational infrastructure to support the use of social media in the workplace. Finally, Social Influence is an individual's beliefs about how those important to them at their workplace believe they should use social media.

Observable Behaviors

The list of observable behaviors initially developed (n=77) were based on behaviors described in the *Seven Areas of Responsibility and Competencies for Health Education Specialists* [19], and guidelines for social media use in health promotion [13,64]. Expert panelists commented on the wording of the behaviors (eg, changing "select" to "identify") and suggested behaviors that could be added to the list (eg, applying health literacy principles to social media campaigns). Based on their suggestions, the wording for 11 behaviors was modified and seven behaviors were added to the list.

Domain Specifications

Domain specifications were developed for each of the six constructs based on the literature and expert feedback. It was clear from the literature, as well as from expert feedback, that Social Media Self-Efficacy required the largest number of items (n=50) to adequately measure the content area and each level of the cognitive domain. Social Media Experience was viewed as the second-most important construct requiring the second-largest number of items (n=20). The content of both of these scales was represented in the *Seven Areas of Responsibility for Health Education Specialists*, and incorporated four levels of the cognitive domain—apply, analyze, evaluate, create—from revisions of Bloom's Taxonomy of the Cognitive Domain [65]. The domain specifications for the four other constructs—Effort Expectancy, Performance Expectancy, Facilitating Conditions, Social Influence—were organized according to the five stages—receiving, responding, valuing, organizing, characterizing—in Krathwohl's Affective Domain Taxonomy [66]. Scales for measuring each of these constructs were used in conjunction with expert feedback to conclude that 3 items could adequately measure each of these four constructs (ie, Effort Expectancy, Performance Expectancy, Facilitating Conditions, and Social Influence). In sum, the domain specifications across all scales represented 82 items.

Phase 2: Item Development

Expert Review of Items and Item Revisions

The initial pool had a total of 148 items (Social Media Self-Efficacy = 91 items, Social Media Experience = 40 items, Effort Expectancy = 5 items, Performance Expectancy = 4 items, Facilitating Conditions = 4 items, and Social Influence = 4 items). All experts selected "yes," indicating that all instructions and items were concise, clear, focused, and readable; had assurance; and displayed an appropriate number of response options. Experts also suggested edits for some of the items (n=74), which amounted to minor modifications in wording.

Think-Aloud Sessions

Think-aloud session participants (n=10) reported working in a diverse array of professional settings, including academia (5/10, 50%), nonprofits (3/10, 30%), a local health department (1/10, 10%), and a state health department (1/10, 10%). Five themes were identified from the qualitative data collected during think-aloud sessions: (1) definitions and terminology instruction, (2) item wording, (3) unintended thought process, (4) formatting and organization, and (5) suggested items. Identified themes informed revisions to items for clarification, revising instructions

to be consistent across the inventory, reducing the number of items that appeared on each page of the survey, and organizing the items in the Social Media Self-Efficacy and Social Media Experience scales by the *Seven Areas of Responsibility for Health Education Specialists*. Finally, the midpoint (ie, neither confident nor unconfident) of the Social Media Self-Efficacy scale was removed because some participants selected this option only when they were unfamiliar with a word or phrase.

Pilot Test

A total of 36 out of 400 (9.0%) participants completed the pilot test. A total of 16 items were removed or revised based on data from the pilot test. None of the response options appeared to be severely skewed in one direction, and all response options were used by pilot survey participants. Within the Social Media Self-Efficacy scale, 9 items were highly correlated ($\rho \geq .90$), suggesting that they measured similar concepts; therefore, these 9 items were removed. A total of 4 items were also removed from the Social Media Experience scale because of high correlations between items ($\rho \geq .90$). A total of 1 item was removed from the Effort Expectancy scale because of a low corrected item-total correlation ($r=.07$), while 2 items from the Social Influence scale were revised based on comments from participants.

Phase 3: Inventory Testing and Finalization

Field Test

A total of 353 individuals out of 1000 (35.30%) completed the Web-based survey during the field test. The demographic characteristics of field test participants can be found in Table 1. Approximately 16.1% (57/353) of field test participants did not provide demographic or organizational information on the survey. The majority of participants identified as female (263/353, 74.5%) with a mean age of 36.87 years (SD 11.58). A total of 60.9% identified as white (215/353), while 10.5% identified as black or African American (37/353) and 9.1% identified as multiple races (32/353). Over half of the participants (208/353, 58.9%) reported a household income of US $50,000 or more. Half of the participants reported having at least a master's degree (176/353, 49.9%), 22.1% reported having at least a bachelor's degree (78/353), and 11.9% reported earning a doctoral degree (42/353).

Organizational characteristics of field test participants can be found in Table 2. On average, participants had 10.03 years (SD 9.15) of experience in the health education field. Practice setting varied, with approximately one-quarter of participants indicating they worked in academia (88/353, 24.9%) and 15.6% reporting they worked for a nonprofit organization (55/353). Other settings included local government or health department (32/353, 9.1%), clinical (25/353, 7.1%), private or corporate (22/353, 6.2%), state government (17/353, 4.8%), federal government (21/353, 5.9%), health insurance (9/353, 2.5%), and K-12 education (3/353, 0.8%). The majority of participants (292/353, 82.7%) reported workplace access to the Internet, but less than half of participants (171/353, 48.4%) reported full access to all social media sites at their place of employment.

Table 1. Demographic characteristics of field test participants (n=353).

Demographics		n (%)
Sex		
	Male	33 (9.3)
	Female	263 (74.5)
	Missing	57 (16.1)
Race/ethnicity		
	White	215 (60.9)
	Black or African American	37 (10.5)
	Asian	5 (1.4)
	Pacific Islander	1 (0.3)
	American Indian or Alaska Native	2 (0.6)
	Hispanic, Latino, or Spanish	3 (0.8)
	Multiple Races/other	32 (9.1)
	Missing	58 (16.4)
Income (US $)		
	$24,999 or less	15 (4.2)
	$25,000 to $49,999	54 (15.3)
	$50,000 to $74,999	66 (18.7)
	$75,000 or more	142 (40.2)
	Don't know	10 (2.8)
	Missing	66 (18.7)
Highest degree earned		
	Bachelor	78 (22.1)
	Master	176 (49.9)
	Doctorate	42 (11.9)
	Missing	57 (16.1)

Table 2. Organizational information for field test participants (n=353).

Organizational information		n (%)
Access to Internet at work		
	Yes	292 (82.7)
	No	4 (1.1)
	Missing data	57 (16.1)
Access to social media at work		
	Full access	171 (48.4)
	Limited access	69 (19.5)
	No access	50 (14.2)
	Not sure	7 (2.0)
	Missing	56 (15.9)
Employer monitors/blocks websites		
	Yes	175 (49.6)
	No	89 (25.2)
	Don't know	32 (9.1)
	Missing	57 (16.1)
Setting		
	State government/health department	17 (4.8)
	Local government/health department	32 (9.1)
	Clinical	25 (7.1)
	Academia	88 (24.9)
	Nonprofit	55 (15.6)
	Private or corporate	22 (6.2)
	Federal government	21 (5.9)
	Health insurance	9 (2.5)
	K-12 education	3 (0.8)
	Other	26 (7.4)
	Missing	55 (15.6)

Psychometric Analyses and Item Removal

The initial classical test theory analyses revealed Cronbach alphas ranging from .64 to .99. Two alphas for data collected using the Facilitating Conditions and Social Influence scales were slightly below the .70 recommended value [45]. A total of 13 items that were highly correlated and measured similar content in two other scales were removed (Social Media Self-Efficacy = 12 items, Social Media Experience = 1 item removed). Further, 4 total items with low corrected item-total correlations (r<.30) were removed (Effort Expectancy=2 items, Facilitating Conditions =1 item, Social Influence=1 item). Analyses of the final inventory items revealed internal consistency ranging from .66 to .98, and corrected item-total correlations ranging from .41 to .86. Table 3 lists summary statistics generated from the final classical test theory procedures.

Table 3. Summary statistics from classical test theory procedures across final scales.

Scale	Cronbach alpha	Corrected item-total (r) range
Social Media Self-Efficacy	.98	.66-.86
Social Media Experience	.98	.75-.85
Effort Expectancy	.74	.51-.63
Performance Expectancy	.81	.60-.73
Facilitating Conditions	.66	.57-.70
Social Influence	.66	.41-.57

Initial CCFAs for the scales revealed statistically significant chi-square test of model fit indices across the scales. RMSEA values were above the recommended .05 level, aside from one scale—Social Influence—which had an RMSEA value of .04. However, many scales (n=4) were close to the cutoff value for mediocre fit (.10) [50]. With the exception of Effort Expectancy, all other TLI and CFI values were .90 or greater, indicating acceptable fit [50]. All standardized loadings were significant, ranging from .36 to .70. Only one scale (Facilitating Conditions) had standardized loadings below .50. CCFAs were conducted a second time for scales that had item(s) removed as a result of the RSM analyses. A summary of the CCFAs for each of the final scales is presented in Table 4.

Table 4. Summary statistics from CCFAs[a] across final scales.

Scale	χ^2 (df)	RMSEA[b] value (95% CI)	TLI[c]	CFI[d]	Standardized loading range	R^2 value range
Social Media Self-Efficacy	7376.1 (1595)[e]	.11 (.11-.11)	.91	.97	.72-.92[e]	.52-.85
Social Media Experience	1161.0 (170)[e]	.14 (.13-.15)	.96	.97	.80-.92[e]	.65-.85
Effort Expectancy	89.9 (1)[e]	.55 (.45-.65)	.63	.88	.70-.73[e]	.49-.54
Performance Expectancy	32.9 (1)[e]	.33 (.24-.43)	.94	.98	.85-.88[e]	.72-.77
Facilitating Conditions	7.1 (1)[e]	.14 (.06-.25)	.97	.99	.36-.70[e]	.13-.50
Social Influence	1.4 (1)	.04 (0-.16)	.99	.99	.52-.57[e]	.28-.32

[a]Categorical confirmatory factor analysis (CCFA).

[b]Root mean square error of approximation (RMSEA).

[c]Tucker-Lewis index (TLI).

[d]Comparative fit index (CFI).

[e]*P*<.001.

RSM analyses were conducted initially for all scale items remaining after classical test theory procedures. Items with item infit and outfit values drastically outside of the recommended range (ie, more than 0.50 logits outside of 0.50-1.50) were removed from the Social Media Self-Efficacy scale (n=1). A review of domain specifications and item fit statistics led to the removal of items measuring similar content with worst fit statistics (Social Media Self-Efficacy=9 items, Social Media Experience=8 items, Social Media Effort Expectancy= 2 items). Initial RSM analyses showed that almost all scales possessed appropriate category fit statistics, and acceptable threshold values that increased in the appropriate order. Only one scale, Effort Expectancy, revealed a noteworthy issue with regard to the category thresholds. While the threshold values for the Effort Expectancy scale ranged from -2.76 to 2.33, the values did not increase in order. The original categories were 0 (Strongly Disagree), 1 (Somewhat Disagree), 2 (Neither Agree or Disagree), 3 (Somewhat Agree), and 4 (Strongly Agree). The threshold for category 2 (Neither Agree or Disagree) was larger than for category 3 (Somewhat Agree). This result presented an issue with the scale as it indicated that higher response categories do not necessarily respond to higher ability level. A follow-up RSM analysis was completed to determine if merging the neutral category (Neither Agree or Disagree) with one of the other categories would cause the thresholds to increase monotonically with either of these changes [67]. Therefore, rescoring included recoding the value of 2 to 1 (ie, scoring sequence 01123) and then recoding the value of 2 to 3 (ie, scoring sequence 01223). This allowed for the neutral category to first become collapsed with Somewhat Disagree, and then, in the second analysis, become collapsed with Somewhat Agree. Items were reverse coded before the analysis to account for the collapsed categories in each analysis. Both analyses resulted in monotonically increasing thresholds. The differences in the increased thresholds for the first change (ie, collapsing with Somewhat Disagree) were more severe than for the second change (ie, collapsing with Somewhat Agree) as evidenced by the curves in the ICCs.

Table 5 lists summary statistics for the second RSM analyses conducted for the final scale items. Item reliabilities were above

the recommended value of .80 or greater [59]. Likewise, the item separation indices for each scale were above the cutoff value of 2.00 [53,59]. Almost all scales had item infit and outfit values within the recommended range of 0.50 to 1.50 logits [58], with the exception of Facilitating Conditions, which had values above 1.50 yet below 2.00 logits. Similarly, the Facilitating Conditions scale had some category infit and outfit values outside of the range of 0.50 to 1.50 logits. Some evidence of local dependence was observed across the scales with residual correlations above the recommended value of r=|.20|.

Table 5. Summary statistics from rating scale model analyses across scales of final inventory.

Scale	Item reliability	Item separation index	Item infit range	Item outfit range	Category infit range	Category outfit range
Social Media Self-Efficacy	.98	6.76	0.63-1.45	0.64-1.62	0.94-1.06	0.90-1.08
Social Media Experience	.98	6.24	0.77-1.45	0.74-1.43	0.88-1.25	0.83-1.24
Effort Expectancy	.95	4.15	0.86-1.15	0.89-1.14	0.86-1.15	0.89-1.14
Performance Expectancy	.99	10.09	0.85-1.35	0.74-1.29	0.93-1.10	0.87-1.15
Facilitating Conditions	.99	16.04	0.78-1.78	0.66-1.86	0.71-1.78	0.66-1.86
Social Influence	.93	3.77	0.88-1.16	0.86-1.10	0.83-1.11	0.78-1.08

Final Social Media Competency Inventory

The final Social Media Competency Inventory (SMCI) can be found in Multimedia Appendix 1. The scale consists of 82 items distributed across six scales: Social Media Self-Efficacy (n=50), Social Media Experience (n=20), Effort Expectancy (n=3), Performance Expectancy (n=3), Facilitating Conditions (n=3), and Social Influence (n=3). Guidelines for the administration, scoring, and interpretation of the SMCI can be found in Multimedia Appendix 2.

Discussion

Principal Findings

Through a multistage instrument-development process, the SMCI was designed to measure six core constructs: (1) Social Media Self-Efficacy, (2) Social Media Experience, (3) Effort Expectancy, (4) Performance Expectancy, (5) Facilitating Conditions, and (6) Social Influence. Using a random sample of CHES/MCHES, evidence of generalizability was provided. Furthermore, including both CHES and MCHES as study participants allowed the reliability and validity evidence to be expanded to a larger population of health education specialists. The demographic and organizational data reported in the field test was comparable to recent studies including samples of CHES/MCHES [27,68].

Overall, adequate reliability and validity evidence supported the utility of the SMCI for assessing health education specialists' use and access to social media technologies for health promotion research and practice. Furthermore, the use of think-aloud sessions during the pilot test provided response process validity evidence within the SMCI's intended population. Information from the think-aloud sessions assisted in determining that participants were interpreting the items and response options as intended. However, because the Effort Expectancy scale experienced disorder thresholds, additional research needs to further explore the thought process of public health education specialists when interpreting and using the Likert response options for this particular scale. It is possible that the neutral option was used as "I'm not sure" or other unintended thought processes.

Results from classical test theory, confirmatory factor analysis, and Rasch RSM procedures provided evidence to support the internal structure of the scales within the SMCI. However, two scales (Facilitating Conditions and Social Influence) revealed internal consistency values below the recommended cutoff values. Nevertheless, data collected using both of these scales generated acceptable item reliability values in the RSM analyses. Based on the CCFA and bivariate residual correlation analyses, the data collected from the Facilitating Conditions and Social Influence scales should be fitted to a more multidimensional model to determine if this allowance provides a better fit for each of the scales' data. Future research is needed to explore the external structure of the scales included in the inventory, as well as the predictive validity of the SMCI.

Understanding the Competency and Theoretical Frameworks of the Social Media Competency Inventory

Constructs within the SMCI were selected using a competency modeling framework and a theoretical framework based on the integrated behavioral model and the unified theory of acceptance and use of technology. According to the integrated behavioral model, there are four conditions under which a behavior is most likely to occur [62]. First, a person should have strong intention to participate in the behavior as well as the knowledge and skills to perform it. Second, there should not be any substantial environmental constraints that could prevent the behavior from being performed. Third, the behavior should be important to the person. And lastly, the person should have some prior experience performing the behavior. Similar to the health behavior theories from which the IBM was established (ie, theory of planned behavior and theory of reasoned action), intention is the most important predictor of behavior. Intention to participate in a behavior offers indication of the individual's "perceived likelihood of performing a behavior" [62]. An individual's behavioral intention is predicted by their personal agency, self-efficacy, and perceived norms associated with the behavior. Possessing the appropriate skills and knowledge is

crucial for a person to be able to successfully perform the behavior, and previous experience with the behavior can translate to the behavior becoming habitual. As with IBM, the unified theory of acceptance and use of technology also emphasizes the significance of behavioral intention, positing that behavior is predicted by behavioral intention as well as facilitating conditions [63]. Behavioral intention is the individual's intention to use the specific technology. Facilitating conditions refer to an individual's beliefs in the existence of technical and organizational infrastructure to support the use of the technology. Behavior intention is predicted by effort expectancy, performance expectancy, and social influence. By blending the theoretical constructs and relationships from these two frameworks, a model for assessing social media competency as well as their relationship to social media performance was created.

Marcolin et al [61] discussed different measures related to technology-competence modeling. They identified three main outcomes related to technology-related user competence: cognitive, skill-based, and affective. Cognitive outcomes refer to the individual's knowledge about the technology and how to use the technology. Skill-based outcomes represent the transition from knowledge to automaticity, which refers to the individual's ability to generalize his or her knowledge to new technology-related tasks. Affective outcomes refer to the motivations and attitudes of the individual as they both pertain to user competence. An instrument attempting to model competency should measure these three outcomes.

Social media competency can be explained as a person's intention in the sense that it indicates their readiness to access and use social media as a function of their employment. This capacity is influenced by their attitudes and beliefs related to social media: to be more specific, their beliefs on how social media use impacts their ability to perform as a health education specialist, how those important to them perceive social media use, the ease of learning how to use social media for health education, and the existing technical and organizational infrastructures for using social media at their place of employment. These perceptions related to four constructs from the UTAUT: (1) effort expectancy, (2) performance expectancy, (3) social influence, and (4) facilitating conditions. These beliefs may also correspond to behavioral and normative beliefs constructs from the IBM. Furthermore, an individual's previous experience using social media is likely to affect their capacity to use social media.

Limitations and Opportunities for Future Research

There are several limitations that should be addressed in future research. First, the list of CHES and MCHES from which the random samples were drawn for the pilot and field tests were not inclusive of all CHES and MCHES. Only CHES and MCHES who agreed to have their contact information distributed to researchers were included on the list. However, this contact list did contain more than 75% of all CHES and MCHES. Similarly, missing data related to demographic and organizational information limits the ability to generalize the findings from the field test to all CHES and MCHES. However, it should be noted that the majority of participants (84%) did provide this information in the field test. Similar missing data related to demographics and organizational information has been observed in other studies of CHES [68]. Not all health education specialists are CHES and MCHES; therefore, future research is needed to test the reliability of SMCI data among health education specialists who are not CHES or MCHES.

Second, data collection for both the pilot and field tests were conducted through a Web-based survey with self-report data. This may have impacted the representation of CHES and MCHES. Some invited CHES and MCHES may not have wanted to participate in a Web-based survey versus a paper-and-pencil or telephone survey. However, Web-based surveys do allow for anonymous surveys, which may have decreased socially desirable responses and offered greater privacy to participants [45]. Multiple methods for data collection on each scale should be conducted in the future to generate multitrait/multimethod matrix validity evidence. Wright [69] provides several advantages of Web-based surveys for research, including reduced time and costs. Because the pilot and field tests were international in scope, it would have been far more time consuming and expensive to have participants complete the inventory in person or by postal mail. Nevertheless, the data obtained from the Web-based surveys were self-reported, and there is no guarantee that individuals provided accurate information.

Evidence of local dependency among items in the SMCI scales' data was another limitation. Large residual correlations may suggest the possibility of multidimensionality [70]. While some research suggests that parameter estimates of item response theory (IRT) models can be somewhat robust to minor violations of unidimensionality or local dependency [71], additional research should be done to determine if multidimensionality exists for the data collected using each of the six different scales.

One last study limitation was that only some types of validity evidence were explored in this study. Types of validity evidence in need of further exploration include divergent, convergent, predictive, and multitrait/multimethod matrix. While it is important to explore convergent and divergent relationships among constructs as well as predictive validity, this was not feasible in this inceptive instrument-development study. Adding more scales to the SMCI would have made the Web-based survey even longer, and may have reduced completion rates. Nonetheless, for the purposes of interpretation, it is important to differentiate between competency and performance, and also understand the relationship between competency and performance. Future research should examine the relationship between social media competency and performance among CHES and MCHES.

Conclusions

The growth in the popularity and functionality of social media technologies corresponds to increasing potential for engaging and reaching specific populations for health promotion activities. While health education specialists widely use social media and general guidelines for social media use in public health are available, an assessment instrument for evaluating the potential of health education specialists to effectively use social media in the workplace was previously unavailable. The SMCI, which

was developed and tested in this study, provides a unique measure to assess the capacity of health education specialists to use social media technologies. The SMCI can be applied to identify gaps in confidence and experience, as well as professional development needs within health education organizations. This data can be used to inform the development of specific guidelines, training, and policies. More research is now needed to explore the dimensionality of data collected using the SMCI. Future studies should also examine the relationship among the six constructs within the SMCI, and the ability of the SMCI to predict social media use and performance among CHES and MCHES. While the results of this study do not offer absolute support for use of the SMCI in high-stake situations (eg, employment decisions), the SMCI can be used to obtain a general understanding of the readiness of health education professionals to use social media to engage populations and deliver relevant public health messages. This study provides the necessary foundation for future research that will help ensure that the health education field is sufficiently prepared to effectively use social media to promote and protect public health.

Acknowledgments

We gratefully recognize the efforts of Robert S Gold, PhD, DrPH, and J Don Chaney, PhD, MCHES, for providing expert feedback during the inventory-development process.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The Social Media Competency Inventory.

[PDF File (Adobe PDF File), 460KB - jmir_v17i9e221_app1.pdf]

Multimedia Appendix 2

Guidelines for administration, scoring, and interpretation of the Social Media Competency Inventory.

[PDF File (Adobe PDF File), 519KB - jmir_v17i9e221_app2.pdf]

References

1. Terry M. Twittering healthcare: social media and medicine. Telemed J E Health 2009;15(6):507-510. [doi: 10.1089/tmj.2009.9955] [Medline: 19659410]
2. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]
3. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2014. Social networking fact sheet URL: http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/ [accessed 2014-01-01] [WebCite Cache ID 6ZEogDwPV]
4. Duggan M, Brenner J. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2013 Feb 14. The demographics of social media users - 2012 URL: http://www.pewinternet.org/2013/02/14/the-demographics-of-social-media-users-2012/ [accessed 2015-06-12] [WebCite Cache ID 6ZEoo6ND9]
5. Neiger BL, Thackeray R, Van Wagenen SA, Hanson CL, West JH, Barnes MD, et al. Use of social media in health promotion: purposes, key performance indicators, and evaluation metrics. Health Promot Pract 2012 Mar;13(2):159-164. [doi: 10.1177/1524839911433467] [Medline: 22382491]
6. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256. [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
7. Vyas AN, Landry M, Schnider M, Rojas AM, Wood SF. Public health interventions: reaching Latino adolescents via short message service and social media. J Med Internet Res 2012;14(4):e99 [FREE Full text] [doi: 10.2196/jmir.2178] [Medline: 22789678]
8. Wójcicki TR, Grigsby-Toussaint D, Hillman CH, Huhman M, McAuley E. Promoting physical activity in low-active adolescents via Facebook: a pilot randomized controlled trial to test feasibility. JMIR Res Protoc 2014;3(4):e56 [FREE Full text] [doi: 10.2196/resprot.3013] [Medline: 25357008]
9. Capurro D, Cole K, Echavarría MI, Joe J, Neogi T, Turner AM. The use of social networking sites for public health practice and research: a systematic review. J Med Internet Res 2014;16(3):e79 [FREE Full text] [doi: 10.2196/jmir.2679] [Medline: 24642014]

10. Roman LA. Using social media to enhance career development opportunities for health promotion professionals. Health Promot Pract 2014 Jul;15(4):471-475. [doi: 10.1177/1524839914535213] [Medline: 24951604]
11. Bernhardt JM, Alber J, Gold RS. A social media primer for professionals: digital dos and don'ts. Health Promot Pract 2014 Mar;15(2):168-172. [doi: 10.1177/1524839913517235] [Medline: 24396119]
12. Parvanta C, Nelson DE, Parvanta SA, Harner RN. In: Riegelman R, editor. Essentials of Public Health Communication. 1st edition. Sudbury, MA: Jones & Bartlett Learning; 2011.
13. Centers for Disease Control and Prevention. 2013. CDC social media tools, guidelines & best practices URL: http://www.cdc.gov/socialmedia/tools/guidelines/ [accessed 2015-06-12] [WebCite Cache ID 6ZEp8Iuin]
14. Thackeray R, Neiger BL, Smith AK, Van Wagenen SB. Adoption and use of social media among public health departments. BMC Public Health 2012;12:242 [FREE Full text] [doi: 10.1186/1471-2458-12-242] [Medline: 22449137]
15. 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]
16. Wong CA, Merchant RM, Moreno MA. Using social media to engage adolescents and young adults with their health. Healthc (Amst) 2014 Dec;2(4):220-224 [FREE Full text] [doi: 10.1016/j.hjdsi.2014.10.005] [Medline: 25984444]
17. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
18. Maher CA, Lewis LK, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
19. The National Commission for Health Education Credentialing, Inc, Society for Public Health Education, American Association for Health Education. A Competency-Based Framework for Health Education Specialists - 2010. Whitehall, PA: National Commission for Health Education Credentialing, Inc; 2010.
20. Bensley R, Brookins-Fisher J. Community Health Education Methods: A Practical Guide. 3rd edition. Sudbury, MA: Jones and Bartlett Publishers; 2009.
21. National Commission for Health Education Credentialing, Inc. CHES overview URL: http://www.nchec.org/ches [accessed 2015-08-07] [WebCite Cache ID 6ac0WnrpD]
22. National Commission for Health Education Credentialing, Inc. MCHES overview URL: http://www.nchec.org/mches [accessed 2015-08-07] [WebCite Cache ID 6ac0URul2]
23. Centers for Disease Control and Prevention. 2013. Tips from former smokers URL: http://www.cdc.gov/tobacco/campaign/tips/ [accessed 2015-06-12] [WebCite Cache ID 6ZEpPrGV4]
24. Centers for Disease Control and Prevention. 2011. The health communicator's social media toolkit URL: http://www.cdc.gov/socialmedia/tools/ [accessed 2015-06-12] [WebCite Cache ID 6ZEpYzZ2l]
25. Capurro D, Cole K, Echavarría MI, Joe J, Neogi T, Turner AM. The use of social networking sites for public health practice and research: a systematic review. J Med Internet Res 2014;16(3):e79 [FREE Full text] [doi: 10.2196/jmir.2679] [Medline: 24642014]
26. George DR, Rovniak LS, Kraschnewski JL. Dangers and opportunities for social media in medicine. Clin Obstet Gynecol 2013 Sep;56(3):453-462 [FREE Full text] [doi: 10.1097/GRF.0b013e318297dc38] [Medline: 23903375]
27. Hanson C, West J, Neiger B, Thackeray R, Barnes M, McIntyre E. Use and acceptance of social media among health educators. Am J Health Educ 2011 Jul;42(4):197-204. [doi: 10.1080/19325037.2011.10599188]
28. Harris JK, Mueller NL, Snider D. Social media adoption in local health departments nationwide. Am J Public Health 2013 Sep;103(9):1700-1707. [doi: 10.2105/AJPH.2012.301166] [Medline: 23865660]
29. Centers for Disease Control and Prevention. 2014 Oct 13. In the know: social media for public health URL: https://npin.cdc.gov/training/know-social-media-public-health [accessed 2015-06-12] [WebCite Cache ID 6ZEpuSE4b]
30. AIDS.gov. Washington, DC: US Department of Health & Human Services; 2012. Getting started with new media URL: https://www.aids.gov/using-new-media/getting-started/index.html [accessed 2015-06-12] [WebCite Cache ID 6ZEpyVxPT]
31. DigitalGov. Washington, DC: US General Services Administration DAP: digital metrics guidance and best practices URL: http://www.digitalgov.gov/services/dap/dap-digital-metrics-guidance-%20and-best-practices/ [accessed 2015-06-12] [WebCite Cache ID 6ZEq2vlWb]
32. Crocker L, Algina J. Introduction to Classical and Modern Test Theory. Orlando, FL: Wadsworth Pub Co; 1986.
33. Alreck PL, Settle RB. The Survey Research Handbook. 3rd edition. Boston, MA: McGraw-Hill/Irwin; 2004.
34. Dillman DA, Smyth JD, Christian LM. Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method. 4th edition. Hoboken, NJ: Wiley; Aug 2014.
35. O'Rourke T, O'Rourke D. The ordering and wording of questionnaire items: part I. Am J Health Stud 2001;17(3):156.
36. O'Rourke T, O'Rourke D. The ordering and wording of questionnaire items: part II. Am J Health Stud 2002;17(4):208.
37. Pett MA, Lackey NR, Sullivan JJ. Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research. Thousand Oaks, CA: Sage Publications; 2003.
38. Weiler RM, Sliepcevich EM, Sarvela PD. Development of the Adolescent Health Concerns Inventory. Health Educ Q 1993;20(4):569-583. [Medline: 8307773]

39. Collins D. Pretesting survey instruments: an overview of cognitive methods. Qual Life Res 2003 May;12(3):229-238. [Medline: 12769135]
40. LinkedIn. URL: https://www.linkedin.com/ [accessed 2015-09-10] [WebCite Cache ID 6bRf4nlhC]
41. Atwal A, Money A, Harvey M. Occupational therapists' views on using a virtual reality interior design application within the pre-discharge home visit process. J Med Internet Res 2014;16(12):e283 [FREE Full text] [doi: 10.2196/jmir.3723] [Medline: 25526615]
42. Yuan P, Bare MG, Johnson MO, Saberi P. Using online social media for recruitment of human immunodeficiency virus-positive participants: a cross-sectional survey. J Med Internet Res 2014;16(5):e117 [FREE Full text] [doi: 10.2196/jmir.3229] [Medline: 24784982]
43. Middleton A, Bragin E, Parker M. Finding people who will tell you their thoughts on genomics-recruitment strategies for social sciences research. J Community Genet 2014 Oct;5(4):291-302 [FREE Full text] [doi: 10.1007/s12687-014-0184-2] [Medline: 24535681]
44. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
45. Crosby RA, DiClemente RJ, Salazar LF. Research Methods in Health Promotion. San Francisco, CA: Jossey-Bass; 2006.
46. Clark LA, Watson D. Constructing validity: basic issues in objective scale development. Psychol Assess 1995;7(3):309-319.
47. Raykov T, Marcoulides GA. Introduction to Psychometric Theory. New York, NY: Taylor and Francis Group; 2011.
48. Behavioral Risk Factor Surveillance System 2014 Questionnaire. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention; 2013 Dec 17. URL: http://www.cdc.gov/brfss/questionnaires/pdf-ques/2014_brfss.pdf [accessed 2015-09-18] [WebCite Cache ID 6bdVEYr7b]
49. Pallant J. SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS for Windows. 3rd edition. Maidenhead, UK: Open University Press; 2007.
50. Brown TA. Confirmatory Factor Analysis for Applied Research. New York, NY: The Guilford Press; 2006.
51. Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. Psychol Bull 1980;88(3):588-606. [doi: 10.1037/0033-2909.88.3.588]
52. Teo T. An introduction to the use of structural equation modeling (SEM) in educational technology research. Int J Instr Media 2010;37(3):263-272.
53. Meyer JP. Applied Measurement With jMetrik. New York, NY: Routledge; 2014.
54. Andrich D. A rating formulation for ordered response categories. Psychometrika 1978 Dec;43(4):561-573. [doi: 10.1007/BF02293814]
55. Bond T, Fox C. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. 2nd edition. Mahwah, NJ: Lawrence Erlbaum Associates Publishers; 2007.
56. Yen WM. Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. Appl Psychol Meas 1984 Apr 01;8(2):125-145. [doi: 10.1177/014662168400800201]
57. Christensen KB, Kreiner S, Mesbah M, editors. Rasch Models in Health. London, UK: Wiley; Dec 2012.
58. Linacre JM. Optimizing rating scale category effectiveness. J Appl Meas 2002;3(1):85-106. [Medline: 11997586]
59. Fox C, Jones J. Uses of Rasch modeling in counseling psychology research. J Couns Psychol 1998;45(1):30-45. [doi: 10.1037/0022-0167.45.1.30]
60. Osterlind S. Modern Measurement: Theory, Principles, and Applications of Mental Appraisal. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall; 2006.
61. Marcolin BL, Compeau DR, Munro MC, Huff SL. Assessing user competence: conceptualization and measurement. Information Systems Research 2000 Mar;11(1):37-60. [doi: 10.1287/isre.11.1.37.11782]
62. Montaño DE, Kasprzyk D. Theory of reasoned action, theory of planned behavior, and the integrated behavioral model. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior and Health Education: Theory, Research, and Practice. 4th edition. San Francisco, CA: Jossey-Bass; 2008:67-96.
63. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425-478. [doi: 10.2307/30036540]
64. Norman CD. Social media and health promotion. Glob Health Promot 2012 Dec;19(4):3-6. [doi: 10.1177/1757975912464593] [Medline: 24803437]
65. Krathwohl DR. A revision of Bloom's taxonomy: an overview. Theory Pract 2002 Nov;41(4):212-218. [doi: 10.1207/s15430421tip4104_2]
66. Bloom BS, Krathwohl DR, Masia BB. Taxonomy of Educational Objectives: The Classification of Educational Goals. New York, NY: Longman; 1984.
67. Sehee H, Kim B, Wolfe M. A psychometric revision of the European American values scale for Asian Americans using the Rasch model. Meas Eval Couns Dev 2005 Jan;37(4):194-207.
68. McKenzie J, Seabert D. A retrospective review of those who hold the certified health education specialist (CHES) credential. Am J Health Stud 2009;24(2):276-286.

69. Wright K. Researching Internet-based populations: advantages and disadvantages of online survey research, online questionnaire authoring software packages, and Web survey services. J Comput Mediat Commun 2005 Apr;10(3):00. [doi: 10.1111/j.1083-6101.2005.tb00259.x]
70. Tennant A, Pallant J. Unidimensionality matters. Rasch Measurement Transactions 2006;20:1048-1051.
71. Embretson SE, Reise SP. Item Response Theory for Psychologists. Mahwah, NJ: Lawrence Erlbaum Associates; 2000.

Abbreviations

APHA: American Public Health Association
BRFSS: Behavioral Risk Factor Surveillance System
CCFA: categorical confirmatory factor analysis
CDC: Centers for Disease Control and Prevention
CFI: comparative fit index
CHES: Certified Health Education Specialist(s)
CINAHL: Cumulative Index to Nursing and Allied Health Literature
ERIC: Education Resources Information Center
HHS: US Department of Health & Human Services
IBM: integrated behavioral model
ICC: item characteristic curve
IRT: item response theory
MCHES: Master Certified Health Education Specialist(s)
NCHEC: National Commission for Health Education Credentialing, Inc
NIH: National Institutes of Health
RMSEA: root mean square error of approximation
RSM: rating scale model
SMCI: Social Media Competency Inventory
SOPHE: Society for Public Health Education
TLI: Tucker-Lewis index
UTAUT: unified theory of acceptance and use of technology
WLSMV: weighted least-squares means and variance adjusted

Edited by G Eysenbach; submitted 16.07.15; peer-reviewed by B Chaney, D Reinwand; comments to author 06.08.15; revised version received 10.08.15; accepted 15.08.15; published 23.09.15.

Please cite as:
Alber JM, Bernhardt JM, Stellefson M, Weiler RM, Anderson-Lewis C, Miller MD, MacInnes J
Designing and Testing an Inventory for Measuring Social Media Competency of Certified Health Education Specialists
J Med Internet Res 2015;17(9):e221
URL: http://www.jmir.org/2015/9/e221/
doi:10.2196/jmir.4943
PMID:26399428

Original Paper

Predictors of Response to Web-Based Cognitive Behavioral Therapy With High-Intensity Face-to-Face Therapist Guidance for Depression: A Bayesian Analysis

Ragnhild Sørensen Høifødt[1], PsyD; Matthias Mittner[1], PhD; Kjersti Lillevoll[1], PsyD; Susanne Kvam Katla[2], PsyD; Nils Kolstrup[3], MD, PhD; Martin Eisemann[1], PhD; Oddgeir Friborg[1], PhD; Knut Waterloo[1], PhD

[1]Department of Psychology, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

[2]Division of Substance Abuse and Addiction, Vestre Viken Hospital Trust, Drammen, Norway

[3]Department of Community Medicine, General Practice Research Unit, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

Corresponding Author:
Ragnhild Sørensen Høifødt, PsyD
Department of Psychology
Faculty of Health Sciences
UiT The Arctic University of Norway
9037 Tromsø
Tromsø, 9037
Norway
Phone: 47 776 49230
Fax: 47 776 45291
Email: ragnhild.s.hoifodt@uit.no

Abstract

Background: Several studies have demonstrated the effect of guided Internet-based cognitive behavioral therapy (ICBT) for depression. However, ICBT is not suitable for all depressed patients and there is a considerable level of nonresponse. Research on predictors and moderators of outcome in ICBT is inconclusive.

Objective: This paper explored predictors of response to an intervention combining the Web-based program MoodGYM and face-to-face therapist guidance in a sample of primary care patients with mild to moderate depressive symptoms.

Methods: Participants (N=106) aged between 18 and 65 years were recruited from primary care and randomly allocated to a treatment condition or to a delayed treatment condition. The intervention included the Norwegian version of the MoodGYM program, face-to-face guidance from a psychologist, and reminder emails. In this paper, data from the treatment phase of the 2 groups was merged to increase the sample size (n=82). Outcome was improvement in depressive symptoms during treatment as assessed with the Beck Depression Inventory-II (BDI-II). Predictors included demographic variables, severity variables (eg, number of depressive episodes and pretreatment depression and anxiety severity), cognitive variables (eg, dysfunctional thinking), module completion, and treatment expectancy and motivation. Using Bayesian analysis, predictors of response were explored with a latent-class approach and by analyzing whether predictors affected the slope of response.

Results: A 2-class model distinguished well between responders (74%, 61/82) and nonresponders (26%, 21/82). Our results indicate that having had more depressive episodes, being married or cohabiting, and scoring higher on a measure of life satisfaction had high odds for positively affecting the probability of response. Higher levels of dysfunctional thinking had high odds for a negative effect on the probability of responding. Prediction of the slope of response yielded largely similar results. Bayes factors indicated substantial evidence that being married or cohabiting predicted a more positive treatment response. The effects of life satisfaction and number of depressive episodes were more uncertain. There was substantial evidence that several variables were unrelated to treatment response, including gender, age, and pretreatment symptoms of depression and anxiety.

Conclusions: Treatment response to ICBT with face-to-face guidance may be comparable across varying levels of depressive severity and irrespective of the presence and severity of comorbid anxiety. Being married or cohabiting, reporting higher life satisfaction, and having had more depressive episodes may predict a more favorable response, whereas higher levels of dysfunctional thinking may be a predictor of poorer response. More studies exploring predictors and moderators of Internet-based treatments are needed to inform for whom this treatment is most effective.

Trial Registration: Australian New Zealand Clinical Trials Registry number: ACTRN12610000257066; https://www.anzctr.org.au/trial_view.aspx?id=335255 (Archived by WebCite at http://www.webcitation.org/6GR48iZH4).

(J Med Internet Res 2015;17(9):e197) doi:10.2196/jmir.4351

KEYWORDS

treatment outcome; computer-assisted therapy; cognitive behavior therapy; depression; primary health care; Bayesian analysis

Introduction

Background

Several efficacious psychological and pharmacological treatments for depression exist [1]. A well-documented treatment is cognitive behavioral therapy (CBT), which has shown comparable effects as pharmacotherapy in treating mild to moderate depression with the additional benefit of reducing relapse [2,3].

The therapy model, structure, and short-term format of CBT make it highly suitable for delivery through self-help material. Delivery through Internet services is one example and several studies have demonstrated the efficacy of Internet-based CBT (ICBT) for depression, especially when guided by a therapist (eg, [4,5-8]). In fact, the treatment effects from guided ICBT and standard face-to-face treatment seem to be comparable [9-11]. Despite the positive results, ICBT is not suitable for all depressed patients because the problem of nonresponse is notable ranging from 50% to 65% (eg, [6,7,12]). Therefore, the question of which patients this treatment is effective for is important to address. The aim of this study is to examine pretreatment variables that can predict response to an ICBT protocol that was published previously [13].

General Prognostic Factors

A number of studies have investigated factors predicting the course of depression in primary care and community settings. Factors associated with a poorer course of the depressive disorder include individual characteristics (eg, high levels of neuroticism [14,15]), socioeconomic factors (eg, low educational level [16,17], unemployment [16,17]), relational factors (eg, lack of social support [17-19], loneliness [17]), health-related variables (eg, somatic illness [17,18], severity of somatic symptoms [19,20], poor self-rated health [21], lower levels of mental [20] and global [15] functioning), and factors related to the depressive disorder (eg, baseline depressive severity [17-20,22], history of depression [23], duration of depressive episodes [16,18], dysthymia or double depression [15 24], and comorbidity with anxiety [17,19,24], substance abuse [22], or personality disorders [24]).

Predictors of Response to Cognitive Behavioral Therapy

In the literature on treatment response, the concepts of prognostic and prescriptive factors are discussed [25]. The former represent nonspecific predictors of response and the latter represent moderators which refer to variables predicting differential treatment response between treatments [25,26]. The latter is most useful for informing which treatments seem most suitable for which patient characteristics or subpopulations [27].

Several patient characteristics have been suggested to influence response to CBT for depression. Patient expectancy, perceived treatment credibility, and improvement in the early phases of treatment seem to be powerful predictors of outcome in cognitive therapy and psychotherapy in general [28-32]. Demographic variables such as gender, age, education, and employment status are less consistently related to treatment outcomes [33-36]. However, in a recent study of treatment-resistant depression in primary care, age was found to moderate the effect of CBT with older patients gaining most benefit from this treatment [37]. In addition, married patients seem to respond consistently better to CBT compared to unmarried patients [38-40]. Many studies suggest poorer outcomes in terms of posttreatment symptoms for patients with high baseline depressive severity (eg, [41-43]). This relationship may depend on the definition of outcome; Van et al [34] propose that high initial severity may be associated with more difficulty achieving remission, whereas symptom change may be achieved more readily because higher severity leaves more room for improvement. In addition, regression to the mean effects can be expected to be stronger for those with a higher symptom load. Other features of the depressive disorder, such as high chronicity and younger age of onset, have been found to predict poorer response to CBT [29,38], but the predictive role of number of depressive episodes [29,30,39] and comorbid anxiety remains unclear [44-47]. With its relation to the proposed mechanism of change in CBT, the role of dysfunctional attitudes has received considerable attention and several studies conclude that high baseline levels of dysfunctional attitudes predict a poorer treatment response [29,33,39,48,49].

Predictors of Response to Internet-Based Cognitive Behavioral Therapy

For ICBT, results concerning depressive severity are consistent with previous research on face-to-face CBT [4,50-56]. In contrast to previous research, studies of ICBT have found either no association between marital status and treatment response [6,52,57] or a positive association between being separated, widowed, or divorced and symptom reduction [53]. Two studies of younger and older adults, respectively, found more favorable outcomes for females [52,58]. Donker et al [59] found similar results in a sample with a broader age range, whereas others have not replicated this finding [4,6,12,56,57]. Age itself did not significantly predict outcome in these studies, with the exception of Donker et al [59] in which age was found to be a moderator because older individuals responded more favorably to CBT and younger individuals improved more with interpersonal therapy (IPT). Results have been mixed for educational level, employment status, dysfunctional attitudes, and for clinical variables such as number of depressive episodes and the presence of comorbidity (eg, [6,12,50-53,56,57,59]).

Treatment credibility refers to the extent to which patients endorse a treatment model as logical and meaningful, and 2 studies found this to be unrelated to outcomes of ICBT [60,61]. Results were mixed with respect to treatment expectancy [60,61]. One study indicated that although higher motivation was associated with greater adherence, low and moderate levels were related to better outcomes, perhaps due to unrealistic expectations and proneness to disappointment for highly motivated participants [57]. One may presume that greater adherence leads to better outcomes, but even on this point there are inconsistencies with some studies finding an association [5,60,62-65] and others not [12,54,57,66-68]. A review suggested that the impact of adherence may depend on how it is measured and that module completion may be more consistently related to outcomes for depression than measures such as number of log-ins [69].

Aim of the Study

The aim of this study is to identify prognostic predictors of response to an intervention combining the Web-based program MoodGYM and high-intensity face-to-face therapist guidance in a sample of mildly to moderately depressed primary care patients. Data from a randomized controlled trial (RCT) comparing this intervention to a delayed treatment condition was used. Data from the treatment phase of the 2 groups were collapsed. This increased sample size in the treatment group, but also precludes a clear distinction between general prognostic factors and predictors specific to CBT. This limitation must be borne in mind when interpreting the results. Predictor variables were predominately chosen on the basis of previous research on CBT delivered face-to-face and over the Internet, but some measures were included for exploratory purposes.

Most patients with mild to moderate depression receive all their treatment in primary care where the availability of psychological treatments is often limited [70-73]. If implemented in primary care, this intervention could constitute an alternative to treatment as usual. This paper may indicate which patients in a depressed primary care population may benefit more or less from treatment with MoodGYM and therapist guidance.

Based on previous literature, we hypothesized that (1) more positive expectations would predict a more favorable response to treatment, (2) participants with higher baseline depression severity would improve more, and (3) a higher score on a measure of dysfunctional thinking would predict a poorer treatment response. Because the remaining predictor variables have yielded mixed results in previous studies, no specific hypotheses were formulated for these.

Methods

Study Design

The study was a RCT with 2 conditions: (1) a treatment condition comprising 6 weeks of Web-based CBT with face-to-face therapist guidance and (2) a 6-week waitlist for the same treatment during which time participants could also access treatment as usual. The research protocol was approved by the Regional Committee for Research Ethics in Northern Norway and the Human Ethics Committee of the Australian National University (ANU). The trial was registered in the Australian New Zealand Clinical Registry (ACTRN12610000257066). A more detailed account of the study methods is given in Høifødt et al [13].

Participants and Procedure

Participants (N=106) were recruited from general practitioners (GPs), primary care nurses, and from waitlists of primary care referrals at 2 psychiatric outpatient clinics. Local GPs and primary care nurses were informed about the study and provided their patients with information about the project. Patients on waitlists at the psychiatric outpatient clinics at the Psychiatric Centre for Tromsø at the University Hospital of North Norway and at the Department of Psychology at UiT The Arctic University of Norway were invited by postal mail. Patients consented by signing an informed consent form. Consenting participants were screened for inclusion and randomly allocated to the 2 groups. The study inclusion criteria were (1) aged 18 to 65 years, (2) access to the Internet, and (3) a score between 10 and 40 on the Beck Depression Inventory-II (BDI-II), which indicates mild to moderately severe symptoms of depression. Individuals already attending CBT were excluded. Participants with suicidal intentions, concurrent psychosis, or alcohol or drug abuse disorders were excluded. Participants who used antidepressant medication were stabilized for 1 month prior to entering the trial.

Assessments and treatment took place at the Department of Psychology at UiT The Arctic University of Norway. Because the patients allocated to the 2 study arms showed comparable courses during treatment, data from the treatment phase of the 2 groups was combined to increase statistical power. Seven participants in the control group dropped out during the waiting period and did not complete the pretreatment assessments. Another 7 participants did not meet the inclusion criteria according to the BDI-II at the pretreatment assessment and were excluded, as were 7 participants who provided data only on 1 measurement occasion. In addition, 3 outliers with treatment duration exceeding far beyond that of the rest of the sample (>28 weeks) were excluded. Because slopes of BDI-II were modeled as a function of time, treatment duration is a critical variable; therefore, we chose to base our criterion for outliers on this scale.

Intervention

The guided self-help intervention included (1) The Norwegian version of the ICBT program MoodGYM version 3 [74], (2) face-to-face therapist guidance of high-intensity, and (3) reminder emails between sessions.

The MoodGYM was originally developed at ANU as a free-of-charge automated Web intervention delivered to the public [75]. MoodGYM consists of 5 self-help modules and 29 exercises. The program is based on CBT and was developed to prevent and reduce symptoms of depression and anxiety among adolescents [76], but is efficacious for adult populations also [8,77-79]. MoodGYM focuses on identification and restructuring of dysfunctional thinking, activation of behavioral strategies to increase engagement in positive activities, as well as learning of stress reduction and problem-solving techniques.

Participants were introduced to the program and instructed to complete one module per week. After each module, participants received face-to-face support (15-30 minutes) from a psychologist (RSH or KL). The main elements of the sessions were reinforcement of progress, discussion of key messages from the modules, and helping participants to relate to the material and employ techniques from the program in their daily life. The full intervention included 8 sessions. The mean number of sessions attended was 7.0 (SD 2.2). Due to delays, some participants attended more sessions (9 sessions: n=8; 10 sessions: n=3; 11 sessions: n=1). Mean session length in minutes (excluding screening) was 28.1 (SD 6.9, range 15.8-48.6). Therapists aimed to meet participants weekly. However, the interval between sessions and the number of sessions were allowed to vary to meet individual needs. Thus, treatment duration varied between participants (mean 9.6, SD 4.8, range 1-22) and there was no fixed posttreatment time-point.

Outcome

Several outcome measures were analyzed in the trial focusing on the effect of the intervention [13]. In this paper, analyses are restricted to predicting response on the BDI-II. The BDI-II was administered to all participants at baseline (before randomization) and before every consultation during the intervention phase. The control group also completed an assessment before entering online treatment (pretreatment).

The BDI-II is a 21-item self-report measure of severity of depressive symptoms during the past 2 weeks [80]. Studies consistently support the BDI-II as a reliable, internally consistent, and valid scale for assessing depression [80-82]. In this study, internal consistency (Cronbach alpha) ranged from .79 to .97 and was generally greater than .90 for the measurement occasions T1 to T11 (baseline to session 11).

Predictors of Outcome

Demographic Variables

The variables gender, age, marital status, and employment status were collected during the screening interview before randomization. Marital status and employment status were dichotomized as married/cohabiting versus not married/cohabiting and being employed versus not being employed, respectively.

Severity Variables

This group of variables included pretreatment measures of severity of depressive and anxious symptoms and quality of life, as well as depression and anxiety diagnosis, number of depressive episodes, and alcohol use measured at baseline. In addition, previous treatment was included as a dichotomous variable (1=yes, 0=no) indicating whether participants had previously received pharmacotherapy or psychological treatment for depression.

Severity of anxiety and depression symptoms pretreatment was assessed with The Hospital Anxiety and Depression Scale (HADS). This inventory has 2 subscales of 7 items each, measuring depression and anxiety, respectively, and is reliable and valid [83,84]. In this study, Cronbach alpha was .67 and .81 for the depression and anxiety subscales, respectively.

Another measure of anxiety severity was The Beck Anxiety Inventory (BAI) [85]. The inventory possesses robust internal consistency, reliability, and validity [86-88]. Cronbach alpha in the present study was .92.

The Mini-International Neuropsychiatric Interview (MINI) [89] was used to identify participants who fulfilled the criteria for a major depressive episode (MDE) or any anxiety disorder, and to determine the number of previous depressive episodes (0=no lifetime MDE; 1=single lifetime MDE; 2=2-4 lifetime MDEs; 3=≥5 lifetime MDEs).

Alcohol use was assessed with the Alcohol Use Disorders Identification Test (AUDIT) [90]. The instrument has favorable internal consistency, reliability, and criterion validity [90,91]. Participants with scores greater than 20 were excluded from the study. Cronbach alpha in this study was .81.

Health-related quality of life was assessed with the EuroQol 5-Dimension Self-Report Questionnaire (EQ-5D) [92]. Respondents mark their level of functioning for each of 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression).

The Satisfaction With Life Scale (SWLS) measures global life satisfaction according to the individual's own criteria [93]. The scale has sound psychometric properties [94,95]. Cronbach alpha in this study was .78.

Cognitive Variables

Dysfunctional thinking and self-efficacy were explored as potential predictors of response. Dysfunctional thinking patterns were measured with the Warpy Thoughts Quiz, which is part of the first module of MoodGYM [96]. The 42-item quiz covers 7 areas of dysfunctional thinking: the need for approval, love, to succeed, and to be perfect; expectations of rights; influence on others; and the view that happiness depends on external things. Items are rated from 1 (strongly agree) to 5 (strongly disagree). Higher scores indicate more dysfunctional thinking. Norms were based on a sample aged 20 to 32 years (N=153) [97], and the scale demonstrates good internal consistency (Cronbach alpha=.77-.84) [96]. A 20-item short form of the scale correlates strongly with the Automatic Thoughts Questionnaire (r=.51) and moderately (r=.39) with measures of depression and anxiety [98].

The General Self-Efficacy Scale (GSE) assesses broad and stable beliefs about one's ability to deal with various demands and challenges [99]. The GSE has satisfactory reliability and construct and criterion cross-cultural validity [100-103]. Cronbach alpha in this study was .89.

Expectancy, Motivation, and Use

Expectancy, attitudes toward using an Internet-based program, and motivation were measured after introducing CBT and MoodGYM using questions developed for the purpose of this study:

1. To which degree do you expect that an Internet-based self-help program can be helpful for your depressive symptoms?

2. How is your attitude toward using an Internet-based self-help program?
3. How likely is it that you will use this Internet-based self-help program?

For the first 2 questions, 5-point Likert scales (1=very high expectations, 5=very low expectations; 1=very negative attitude, 5=very positive attitude) were used. Responses to the item on motivation (question 3) were given on an 11-point scale from 0% to 100%. User data on module completion was registered online and was denoted by a number between zero and 4, with zero indicating no use and 4 indicating completion of the module.

Statistical Analyses Using Bayesian Statistics

Motivation for Using Bayesian Methods

Bayesian methods were used for data analyses instead of the more commonly used null-hypothesis significance testing (NHST) approach (for a general introduction to Bayesian methods, see [104,105]). In a Bayesian framework, we directly estimated the posterior probability distribution of the parameters taking data and model structure into account. Bayesian methods are suitable in the current setting for several reasons. First, the use of Bayesian hierarchical modeling allows the design of custom models that are appropriate for the data without relying on approximations as is necessary in NHST methods. Furthermore, Bayesian modeling is highly flexible because the posterior distribution can be readily transformed into easily interpretable quantities and the uncertainty inherent to the analysis is propagated and available at each level of analysis. As such, Bayesian analysis relies much less on point estimates and an arbitrary choice of significance levels. Indeed, the strong critique on *P* values (regarding, for example, their biasing impact on which results are trusted/reported and the problems with their interpretation [106,107]) emerging in many relevant scientific fields such as medicine [108] and psychology [106] has triggered the development of Bayesian methods in these fields (eg, [109,110]). Instead of reporting *P* values and relying on the problematic concept of statistical significance using an arbitrary significance level, Bayesian methods report the results of an analysis in terms of probabilities, odds ratios, and Bayes factors that give a more graded and readily interpretable summary of the conclusions supported by the data.

Odds ratios are ratios of probabilities or densities indicating the probability of one event occurring relative to another. Similarly, the Bayes factor quantifies how much more likely one hypothesis is with respect to another by dividing the posterior model odds by the prior model odds. Note that the Bayes factor integrates the probability over the complete parameter space and, therefore, automatically punishes overly complex models. Jeffreys [111] discussed how Bayes factors could be interpreted in terms of strength of evidence for and against a hypothesis (see Table 1) and it has been shown that Bayes factors are less prone to overestimating effects from psychological experiments compared to *P* values [112].

Using Bayes factors, Bayesian modeling may quantify the support for the null hypothesis and to what extent the null hypothesis (H_0) is more likely than the alternative (H_1). This is advantageous compared to traditional NHST-based tests which can only “not reject” the null hypothesis. This is a desirable feature when investigating the potential impact of predictor variables on treatment efficiency.

Table 1. Evidence categories for Bayes factors (BF_{10}).[a]

Bayes factor	Interpretation
>100	Decisive evidence for H_1
30-100	Very strong evidence for H_1
10-30	Strong evidence for H_1
3-10	Substantial evidence for H_1
1-3	Anecdotal evidence for H_1
1	No evidence
1/3-1	Anecdotal evidence for H_0
1/10-1/3	Substantial evidence for H_0
1/30-1/10	Strong evidence for H_0
1/100-1/30	Very strong evidence for H_0
<1/100	Decisive evidence for H_0

[a] Adapted from Wetzels et al [112]. BF_{10} is the odds for the alternative hypothesis (H_1) divided by the odds for the null hypothesis (H_0).

Statistical Models

Depression scores from BDI-II were acquired for each individual over several weeks of treatment. Because the intervention allowed a flexible session schedule there was resulting variation in measurement occasions; therefore, the effects of time from treatment could not be disentangled. Because participants could use the self-help program between sessions, we hypothesized that participants would continuously benefit from the treatment between sessions. Therefore, time (in weeks) was chosen as the

repeating variable because this was considered to be the most correct representation of the data. We conducted a model selection procedure (for details see Multimedia Appendix 1) to find the most faithful representation of our data from among a linear, a quadratic, and an exponential model. Based on the results from this procedure, we modeled the BDI-II scores on the individual level as an exponential function of time and constrained the individual regression coefficients by a group-level distribution (hierarchical model). In Bayesian analysis, the specification of prior belief is essential. We specified a weakly informative prior such that the estimates were allowed to vary across a large number of parameter values while constraining them to be in a plausible range [105,113].

We implemented 2 complementary models, one for predicting probability of responding to treatment and another one for quantifying the strength of the response. The models were fit using Markov chain Monte Carlo (MCMC) algorithms implemented in the Just Another Gibbs Sampler (JAGS) software [114] and convergence was ensured by visual inspection and the Gelman-Rubin diagnostic [115]. We also conducted posterior predictive checks to ensure that the model fit the data well [105] (see Figures S1-S3 in Multimedia Appendix 1).

Predicting Probability of Response

Response to depression treatment varies substantially across individuals [27]. Latent-class approaches allow for the modeling of different growth trajectories across subgroups and captures this unobserved heterogeneity in trajectories by employing a categorical latent variable [116,117]. Class membership is initially unknown, but is inferred based on observed data resulting in identified classes of individuals with more similar response patterns within each group than between groups [116]. Thus, different classes of individuals may vary around different mean growth curves with potentially unique forms and parameter values. This can be advantageous compared to conventional growth modeling which assumes that all individuals are drawn from the same population and estimates the average growth curve for this population [118]. Furthermore, covariates can be included in the model to predict class membership and, in this way, individual characteristics predicting differential trajectories may be identified. Previous investigations have successfully employed latent-class methods to identify different distributions for groups of responders and nonresponders to treatment [32,119,120]. Therefore, we chose to fit a model that assumed 2 different distributions from which subject-level parameters could be drawn. Predictor variables were used as regressors on probability of class membership using a logit link function (for details see Multimedia Appendix 1) resulting in estimates β_i for each predictor. The resulting model effectively distinguished between responders and nonresponders (see Figure S4 in Multimedia Appendix 1).

Predicting the Strength of Response

In a next step, we aimed to explain variation in responsiveness by identifying variables that correlated with the slope of the response. This is an alternative way to look at prediction of response and it has the advantage of being more directly comparable to previous studies because latent-class approaches have not been widely used in the field. We modeled this situation by adding the subject-level covariates as linear predictors on the estimate of the first-level regression slope. Because changes of the slope parameter in the exponential model are not reflected linearly (a unit change on a low slope parameter has strong impact whereas the same change on a higher slope parameter has less impact), we relied on the quadratic model for this approach. This resulted in estimates α_i for the regression coefficient for each predictor.

Results

Sample Characteristics

A total of 106 participants were included in the study and randomized to an intervention condition (n=52) or a delayed treatment control condition (n=54). Figure 1 describes the flow of participants through the trial. Of the 54 participants in the control group, 47 (87%) showed up for pretreatment assessment after being on a waitlist. For the control and intervention groups, 21 of 47 (45%) and 15 of 52 (29%) participants, respectively, dropped out between pre- and posttreatment assessments.

Treatment adherence was moderate with 31 of 52 participants (60%) in the intervention group and 20 of 54 participants (37%) in the control group adhering to treatment (completing MoodGYM and attending at least 7 sessions). The average number of completed modules and pretreatment characteristics of the sample are presented in Table 2. Distributions for the predictors are shown in Figure S5 in Multimedia Appendix 1.

Table 2. Participant characteristics (N=82).

Variables	Participants
Demographic variables	
Gender (female), n (%)	60 (73)
Age (years)	
Mean (SD)	36.0 (11.7)
Range	18-63
Marital status (married/cohabiting), n (%)	44 (54)
Educational level (higher education),[a] n (%)	41 (50)
Employment status (employed),[b] n (%)	56 (68)
Severity variables	
Symptom measures,[c] mean (SD)	
Beck Depression Inventory-II	21.3 (6.6)
Beck Anxiety Inventory	13.0 (10.2)
HADS Depression	8.3 (2.9)
HADS Anxiety	9.7 (4.1)
Satisfaction With Life Scale	16.7 (5.1)
EQ-5D	0.7 (0.2)
AUDIT	5.0 (4.1)
Depression diagnosis, n (%)	44 (54)
Number of major depressive episodes,[d] n (%)	
0	5 (6)
1	27 (33)
2-4	25 (31)
≥5	19 (23)
Comorbid anxiety,[e] n (%)	27 (33)
Earlier treatment,[f] n (%)	49 (60)
Present treatment (antidepressants or other[g]), n (%)	23 (28)
Cognitive variables, mean (SD)	
Warpy Thoughts Quiz[f]	82.8 (25.1)
General self-efficacy[f]	26.6 (4.9)
Expectancy, motivation, and use	
Expectancy (1=very high expectations), mean (SD)	2.6 (0.7)
Attitude (5=very positive), mean (SD)	4.1 (0.8)
Motivation, mean (SD)	94.0 (12.2)
Number of modules, mean (SD)	3.8 (1.7)
Treatment duration (weeks)	
Mean (SD)	9.6 (4.8)
Range	1-22
Treatment sessions, mean (SD)	7.0 (2.2)

[a] Data for 1% (1/82) missing.

[b] Employed: full-time or part-time employment. Not employed: unemployed, student, homemaker, long-term sick.

[c] Hospital Anxiety and Depression Scale (HADS): 4% (3/82) missing, Satisfaction With Life Scale: 10% (8/82) missing, EuroQol 5-Dimension (EQ-5D) Self-Report Questionnaire: 11% (9/82) missing, Alcohol Use Disorders Identification Test (AUDIT): 1% (1/82) missing.

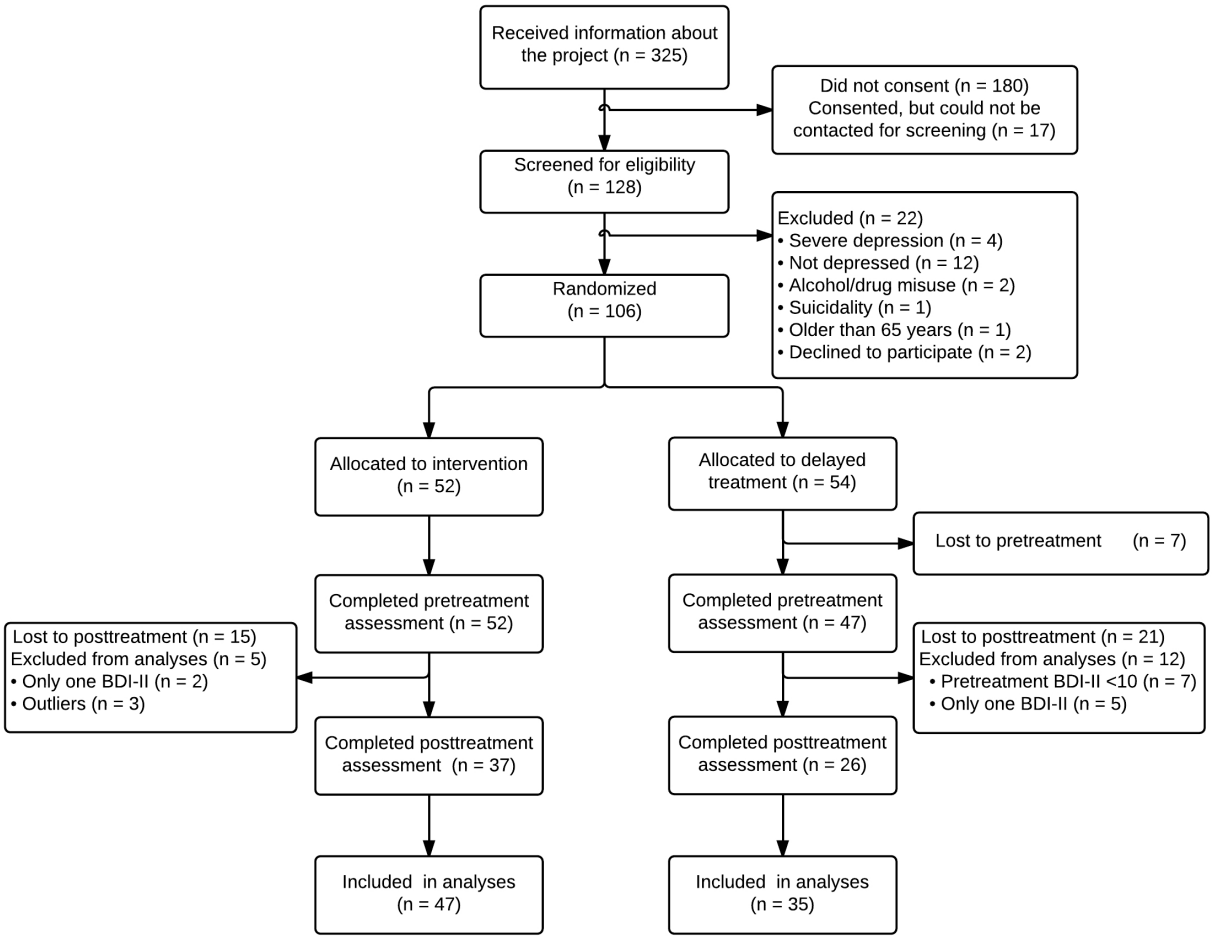
[d] Data for 7% (6/82) missing.

[e] Includes panic disorder, agoraphobia, social phobia, and generalized anxiety.

[f] Data from 2% (2/82) missing.

[g] Psychological therapy other than CBT.

Figure 1. Flow of participants through the trial.



Predicting Probability of Response

The restricted 2-class model distinguished well between responders and nonresponders (ie, most participants either have a very low or a very high probability of belonging to the responder group, P_{resp}) (see Figure S4 in Multimedia Appendix 1). Using P_{resp}=.05 as split criterion, we found that 21 of 82 (26%) participants did not respond to treatment, whereas 61 of 82 (74%) did. These results were based on the conditional latent-class exponential model encompassing all predictor variables. A corresponding analysis using the quadratic model found qualitatively similar results. The results of the regression of the covariates on the probability to respond to the treatment are reported in Table 3. The odds ratios indicate the degree of evidence that each covariate has a positive/negative impact relative to the probability of the opposite (eg, as indicated in Table 3, it is almost 15 times more likely that a subject's score on the Warpy Thoughts Quiz affects the probability of him or her responding to treatment negatively rather than positively). Thus, the odds ratios give an indication of the likely direction of the effect of a covariate on the probability of response, but do not delineate the strength of this effect. To give an indication of the strength of the effect, the probability of being in the responder group as a function of each of the covariates is plotted in Figure 2. This relatively complex reporting of the strength of effects was necessary for this analysis since the estimation of Bayes factors in latent-class models is computationally complicated and still a topic of ongoing research.

Table 3. Posterior mode, highest density interval (HDI), and odds ratios for the beta coefficients predicting probability of being a responder. The odds ratios indicate the probability that each covariate has a positive/negative impact relative to the probability of the opposite (+: positive effect; -: negative effect), but do not indicate the strength of this effect.

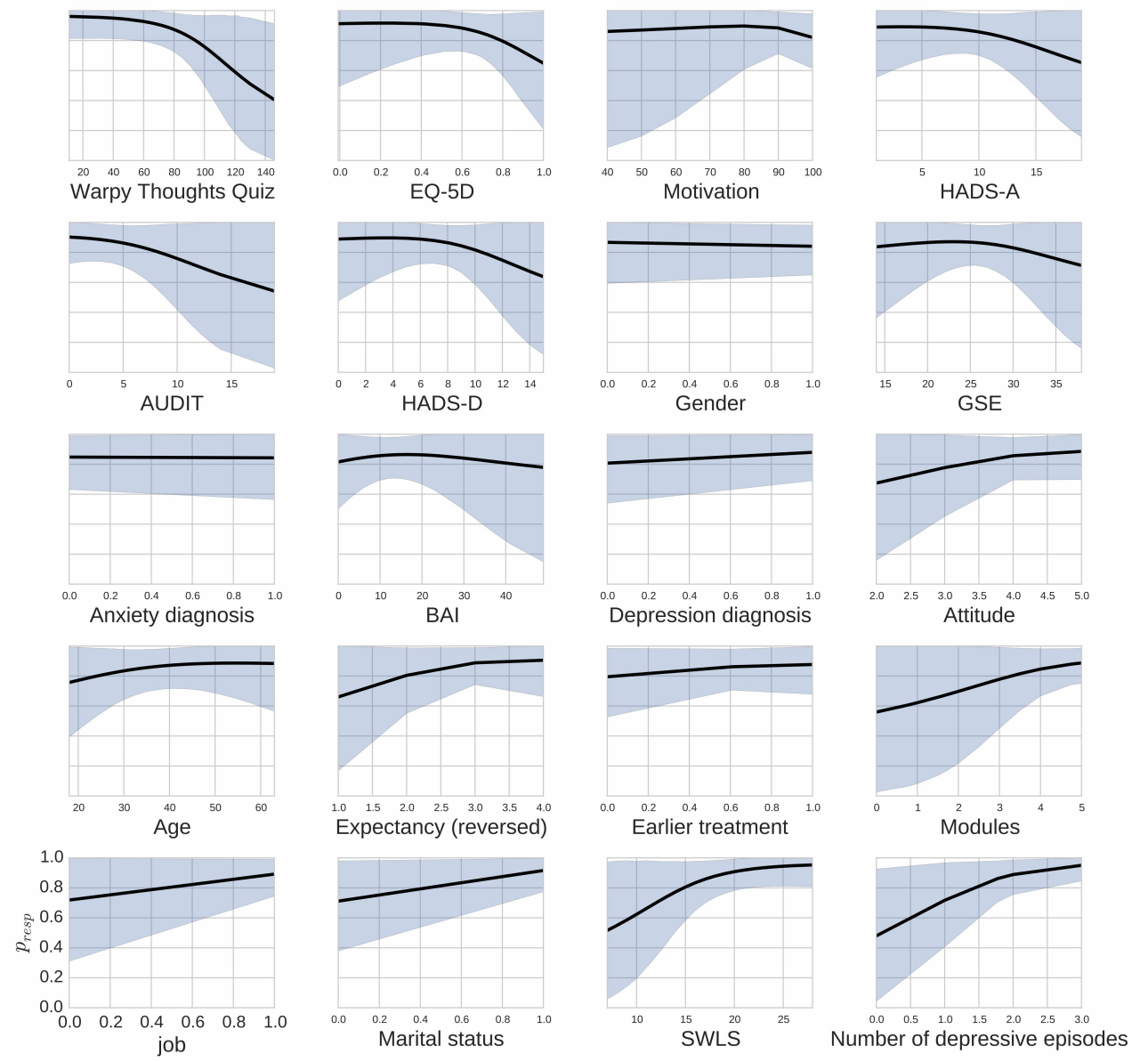
Variable[a]	Posterior mode (HDI)	OR $\|\beta_i\| > 0$
Warpy Thoughts Quiz	−0.93 (–2.27, 0.30)	14.55−
EQ-5D	−0.71 (–2.21, 0.76)	4.84−
Motivation	−0.70 (–2.13, 1.05)	3.42−
AUDIT	−0.49 (–1.82, 0.73)	3.67−
HADS-A	−0.48 (–1.87, 0.88)	3.11−
HADS-D	−0.44 (–1.72, 0.90)	2.99−
GSE	−0.21 (–1.71, 1.19)	1.70−
Gender	−0.15 (–1.46, 0.97)	1.74−
Anxiety diagnosis	0.04 (–1.29, 1.39)	1.10+
BAI	0.08 (–1.27, 1.63)	1.31+
Earlier treatment	0.33 (–0.88, 1.53)	2.44+
Depression diagnosis	0.33 (–0.96, 1.51)	2.19+
Age	0.37 (–0.86, 1.60)	2.56+
Expectancy (reversed)	0.39 (–0.77, 1.70)	3.23+
Attitude	0.42 (–0.87, 1.60)	2.54+
Modules	0.45 (–0.82, 1.82)	3.22+
Employment status	0.51 (–0.76, 1.81)	3.77+
Marital	0.83 (–0.48, 2.09)	8.17+
SWLS	0.88 (–0.44, 2.20)	10.92+
Number of depressive episodes	1.02 (–0.14, 2.28)	23.91+

[a] AUDIT: Alcohol Use Disorders Identification Test; BAI: Beck Anxiety Inventory; EQ-5D: EuroQol 5-Dimension Self-Report Questionnaire; GSE: General Self-Efficacy Scale; HADS-A: Hospital Anxiety and Depression Scale-anxiety subscale; HADS-D: Hospital Anxiety and Depression Scale-depression subscale; SWLS: Satisfaction With Life Scale.

In summary, having had more depressive episodes, being married or cohabiting, and scoring higher on life satisfaction (SWLS) had high odds for positively affecting the probability of response. Tentative positive effects were found for the number of completed modules and having a paid job. Figure 2 shows that the effects are strongest for number of depressive episodes and scores on the SWLS with the probability for response approaching 1 for those with 5 or more depressive episodes and those with highest levels of life satisfaction, whereas those never having had a major depressive episode (only symptoms) and those with the lowest level of life satisfaction had only approximately .50 probability of response.

In the opposite direction, higher scores on the Warpy Thoughts Quiz were likely to have a negative effect on the probability of responding to treatment. Tentative negative effects were found for health-related quality of life (EQ-5D), motivation, expectancy, scores on both subscales of the HADS, and for alcohol use (AUDIT). Figure 2 shows that high scores on the Warpy Thoughts Quiz appear to be associated with a substantially reduced probability of response (P_{resp}~.40). The impact of the other covariates are more limited (P_{resp}~.60-.80 for participants with scores in the highest range; see Figure 2).

Figure 2. Probability of being in the responder group as a function of the predictor variables (assuming all other predictors remained at their baseline level). Black line is the mean posterior probability and shaded area is the 95% highest density interval.



Predicting the Strength of Response

The analysis of variation in responsiveness indicated that the predictors having the highest impact on response were largely consistent with the results from the latent-class model with the most important variables being the Warpy Thoughts Quiz, number of depressive episodes, life satisfaction (SWLS), module completion, and marital status. Results are summarized in terms of odds ratios in Table 4. Results from a separate analysis exploring the variation in responsiveness in the subgroup of responders (n=61) extracted by the latent-class model described in the previous section are presented in Multimedia Appendix 1.

Bayes factors quantify the strength of evidence for the null hypothesis (the covariate does not affect treatment response) and for the alternative hypothesis (the covariate affects response to treatment).

The results were largely consistent with the results from the odds ratio analyses with regard to which variables were most influential (see Table 5). However, the evidence was substantial only for the effect of marital status. There was substantial evidence for the null hypothesis for several variables, indicating that these variables are likely to be unrelated to treatment response in the present trial. This included the variables gender and age, and several severity variables including pretreatment symptoms of depression and anxiety, as well as treatment expectancy, attitude, and motivation. Inconsistent with the results from the odds ratio analyses, there was substantial evidence that the Warpy Thoughts Quiz was unrelated to treatment response.

Table 4. Posterior mode, highest density interval (HDI), and odds ratios for the α coefficients predicting the strength of the response. The α coefficients are the group-level regression coefficients on the slope of the treatment effect in the quadratic model (see Equation 5 in Multimedia Appendix 1). The odds ratios indicate the probability that each covariate has a positive/negative impact relative to the probability of the opposite (+: positive effect, –: negative effect), but do not indicate the strength of this effect.

Variable[a]	Posterior mode (HDI)	OR $\|\alpha_i\| > 0$
Warpy Thoughts Quiz	−0.23 (–0.50, 0.05)	18.28−
Motivation	−0.20 (–0.50, 0.07)	13.00−
GSE	−0.17 (–0.45, 0.11)	7.35−
EQ-5D	−0.09 (–0.34, 0.16)	2.95−
Earlier treatment	−0.08 (–0.35, 0.19)	2.46−
AUDIT	−0.05 (–0.29, 0.18)	2.07−
HADS-D	−0.04 (–0.36, 0.25)	1.62−
Age	−0.04 (–0.31, 0.23)	1.53−
Attitude	−0.01 (–0.25, 0.23)	1.10−
BAI	−0.01 (–0.33, 0.30)	1.16−
Gender	0.01 (–0.21, 0.24)	1.29+
HADS-A	0.02 (–0.28, 0.33)	1.22+
Depression diagnosis	0.02 (–0.27, 0.30)	1.18+
Anxiety diagnosis	0.03 (–0.27, 0.33)	1.34+
Expectancy (reversed)	0.07 (–0.18, 0.32)	2.28+
Employment status	0.09 (–0.20, 0.34)	2.38+
Marital status	0.13 (–0.14, 0.41)	4.72+
Modules	0.18 (–0.12, 0.45)	7.43+
Number of depressive episodes	0.23 (–0.02, 0.49)	29.24+
SWLS	0.24 (–0.04, 0.52)	22.83+

[a] AUDIT: Alcohol Use Disorders Identification Test; BAI: Beck Anxiety Inventory; EQ-5D: EuroQol 5-Dimension Self-Report Questionnaire; GSE: General Self-efficacy Scale; HADS-A: Hospital Anxiety and Depression Scale-anxiety subscale; HADS-D: Hospital Anxiety and Depression Scale-depression subscale; SWLS: Satisfaction With Life Scale.

Table 5. Bayes factors (BF_{10}) quantifying the evidence for alternative hypotheses (H_1) over the null hypothesis (H_0). Variables are sorted with respect to its Bayes factor in ascending order. The null hypothesis is that the predictor does not have an impact on treatment response (H_0: $\alpha_1=0$) and the alternative is that it does have an effect (H_1: $\alpha_1\neq0$). BF_{10} is the odds for H_1 divided by the odds for H_0.

Variable[a]	BF_{10}	Evidence for
Earlier treatment	0.15	H_0: substantial
Gender	0.15	H_0: substantial
GSE	0.16	H_0: substantial
BAI	0.16	H_0: substantial
Expectancy	0.17	H_0: substantial
Depression diagnosis	0.17	H_0: substantial
EQ-5D	0.18	H_0: substantial
Anxiety diagnosis	0.18	H_0: substantial
HADS-A	0.20	H_0: substantial
HADS-D	0.20	H_0: substantial
Attitude	0.22	H_0: substantial
Motivation	0.23	H_0: substantial
Age	0.26	H_0: substantial
Warpy Thoughts Quiz	0.29	H_0: substantial
AUDIT	0.37	H_0: anecdotal
Employment status	0.42	H_0: anecdotal
Modules	0.47	H_0: anecdotal
Number of depressive episodes	0.82	H_0: anecdotal
SWLS	1.82	H_1: anecdotal
Marital status	3.24	H_1: substantial

[a] AUDIT: Alcohol Use Disorders Identification Test; BAI: Beck Anxiety Inventory; EQ-5D: EuroQol 5-Dimension Self-Report Questionnaire; GSE: General Self-efficacy Scale; HADS-A: Hospital Anxiety and Depression Scale-anxiety subscale; HADS-D=Hospital Anxiety and Depression Scale-depression subscale; SWLS: Satisfaction With Life Scale.

Discussion

Principal Findings

This paper explored predictors to a treatment combining the MoodGYM program and high-intensity face-to-face guidance. Using Bayesian methods and a latent-class approach, a 2-class model classifying 74% of participants as responders and 26% as nonresponders was identified. The variation in responsiveness was also explored by analyzing whether predictors affected the slope of response. The results suggest that treatment effects were unrelated to baseline depressive severity, gender, and age. In addition, the presence and severity of comorbid anxiety did not predict differential response to treatment. Having a partner and reporting higher life satisfaction at baseline were associated with a more favorable treatment response. Results also indicated that having experienced more depressive episodes may predict more positive treatment effects, whereas higher scores on the Warpy Thoughts Quiz, which is a measure of dysfunctional thinking, may predict poorer response to treatment.

Limitations

The results of this study must be interpreted in light of some methodological limitations. Despite merging data from both treatment groups, the size of the sample is limited and may be too small to allow reliable testing of effects. Small sample sizes are a common problem in research on prediction of treatment outcome [121]. Collapsing the data from the 2 groups increased the sample size in the treatment group, but precluded the identification of predictors or moderators of differential treatment response [27]. This means that this study cannot accurately distinguish between nonspecific predictors of good prognosis, nonspecific predictors of response to any treatment, and moderators (predicts differential response to treatments). In an effort to ameliorate this limitation, we separately investigated whether the individual predictors could explain the change in the BDI-II score during the waiting time for the waitlist control group. Due to the limited sample size, Bayes factors indicated no evidence for the alternative hypothesis for any of the predictors and could neither establish confidence in the null hypothesis for most of the variables.

Multiple comparisons in small samples also introduces a risk of chance findings. Studies with low power have a high chance of overestimating effect sizes or even making sign errors (eg, [122]). Bayesian methods allow us to model all data in a joint context and reduce the multiple comparison problem by constraining individual model coefficients by an overarching distribution (for details see [104]). In addition, formulating the results in terms of probabilities and odds ratios rather than making dichotomized decisions about whether or not a variable serves as a predictor or not can prevent overinterpretation of results.

Another limitation is that the intervention allowed a flexible session schedule and hence a variation in the spacing between measurement occasions. This means that the effects of time from treatment cannot be disentangled. Because participants could use the self-help program between sessions, we hypothesized that participants would continuously benefit from the treatment also between sessions. Therefore, time was chosen as the repeating variable because this was considered to be the most correct representation of the data.

The choice of outcome and predictor variables may also be criticized. Although demographic variables and baseline axis-I diagnoses were well covered, several variables that may have important contributions, such as personality variables, were not investigated. Therefore, these results can only give a partial description of factors influencing treatment response. The sole reliance on self-report is another limitation. Furthermore, treatment expectancy, attitudes, and motivation were measured using invalidated single items developed for this study. In addition, the convergent and discriminant validity of the Warpy Thoughts Quiz have not been established. This leaves uncertainty regarding how well these constructs were captured and calls for caution in interpreting the results.

A limitation of the 2-class model was that we were unable to estimate Bayes factors due to statistical complexity. This would have provided additional information about the strength of effects. Bayesian methods are a field of active research and development, and improved methods will surely be available in the future.

Finally, although a strength of the study is the recruitment of a relatively heterogeneous sample of primary care patients with regard to the range of depression and anxiety symptoms, the generalizability of the results is uncertain because the sample was a self-selected group. Nevertheless, an estimated uptake of 39% indicates that the trial sample may be representative of a considerable proportion of the targeted patient group [13]. Because some participants were excluded from analyses (eg, participants present at only one measurement occasion), results are based on a subsample of trial participants which further limits generalizability.

Variables Unrelated to Treatment Response

Bayesian methods may be used to indicate the likelihood of the null hypotheses. The present analyses provide substantial evidence for the absence of any effect for several variables, such as pretreatment symptoms of depression and anxiety, depression or anxiety diagnosis, earlier treatment, and the demographic variables gender and age. This implies that MoodGYM combined with face-to-face guidance of relatively high intensity may be expected to work equally well for adult patients of varying ages, for women and men, and for various mild to moderate depressive symptom profiles, as well as for patients with comorbid anxiety of varying severity. Previous results regarding the predictive role of anxiety have been mixed [44-47,51,56,59,123]. With regard to depressive severity, several studies of CBT have found a larger response in terms of symptoms change for patients with higher severity (eg, [34,43,50,52,55]). However, these patients also tend to have more difficulties with achieving remission [39,43,124]. This trial did not find evidence for more improvement among participants having higher initial depressive severity; nevertheless, the results suggest that patients with higher depressive severity appear to benefit from treatment. Whether remission was achieved at comparable rates for participants with more or less severe depression cannot be answered by the present analyses. In addition, because the range of symptom severities was restricted because patients with severe depression were excluded and the proportion of patients having severe anxiety was small, no conclusions can be drawn with regard to more severe cases.

Predictors of Improved Response

Being married or cohabiting was the most robust predictor of favorable response to treatment. This effect was evident both in the latent-class model and the analysis exploring the strength of response, and the Bayes factor indicated substantial evidence for a predictive effect. These results are in accordance with previous research on CBT delivered face-to-face [38-40]. In fact, some studies have suggested that marital status may be a prescriptive predictor for better outcomes in CBT compared to medications or IPT [38,40]. Although, this study cannot identify moderators, these past results indicate that having a partner is likely to be a predictor of treatment response and not merely of good prognosis. Supportive relationships were emphasized in interviews with participants from the current trial [125]. Participants described how important others encouraged them and facilitated their engagement in treatment (eg, by helping them make time to use MoodGYM or attend sessions). This strengthened their hope for recovery and motivation. Although important others also include friends and other family, one may hypothesize that living with a partner may facilitate such reinforcing processes. Also, being married or cohabiting may reflect a better ability to establish and maintain close relationships and this may in itself be an important factor for success in treatments that include interaction with a therapist [39]. This study included high-intensity face-to-face support. This may explain why this effect was evident in the current trial, whereas most studies of ICBT have failed to find any relation between marital status and response [6,52,57]. Replications within other contexts may decide whether this effect is unique to interventions including face-to-face contact or if similar processes operate also in Internet-based interventions including less support.

Life satisfaction also emerged as a possible predictor of better response to treatment, although the Bayes factor analysis indicated only anecdotal evidence for this effect. Life

satisfaction may be regarded as an indirect measure of illness severity. The SWLS does not directly tap into constructs such as affect, but it is significantly negatively correlated with measures of depression and anxiety [94,95]. This result is consistent with an early study of ICBT in which higher quality of life, although assessed with a different scale, was associated with better outcomes [51]. However, this has not been replicated in other studies [56,59]. Why health-related quality of life (EQ-5D) showed tendencies toward predicting more inferior response in this study is more of a riddle. However, the 2 scales assess quite different constructs with the SWLS focusing on how satisfied individuals are with life according to their own criteria and not based on the presence or absence of specific ailments or impairments. The EQ-5D, on the other hand, focuses on the latter. These 2 constructs need not be highly correlated as is supported in this study's data (r=.23). Whether life satisfaction is a more potent predictor of better treatment response remains to be replicated.

The results indicate that more depressive episodes have high odds for predicting a more favorable response. However, the result of the Bayes factor analysis was more ambiguous. This result is puzzling given that high rates of recurrence have been related to poor treatment outcomes [30,51] and treatment resistance [126,127] in previous studies. However, the findings are inconsistent and other studies have found no negative effect of high rates of recurrence on treatment outcomes [4,38,39,53,59]. There are some possible explanations for this finding. Compared with participants with a single or no depressive episodes, more participants with recurrent depression received antidepressant medication or additional psychological therapy. Although most did not receive additional treatment (~65%) and medications were stabilized for 1 month before entering the trial, one cannot rule out the possibility of this influencing the treatment effect. This would be consistent with a meta-analysis finding significantly better effects when adding psychotherapy to pharmacotherapy [128]. Another explanation may be related to the nature of recurrent depression in the general population because studies have suggested that subsequent episodes are shorter in duration than first episodes and have a mean duration of only 3 months [129,130]. This sample was recruited from GPs and is likely to be more similar to a general population sample than to a clinical population recruited from specialist mental health services. In accordance with these epidemiological studies, recurrent depression may be a predictor of shorter episode duration in general population samples. Finally, given that this finding was not fully robust across analyses and was in the opposite direction of most previous results, it may represent a chance finding as a result of random fluctuations in small samples.

The effect of module completion was ambiguous with the odds ratios indicating a tentative positive effect, but the Bayes factor indicating anecdotal support for no effect. Previous results have been mixed on the association between adherence to treatment and response [56,57,63-65,67-69]. The addition of supportive sessions in this trial may have confounded the effect of module completion and although there was high correlation between completing modules and attending sessions (r=.86), a measure reflecting adherence to both treatment components could have been a more potent predictor.

Predictors of Poorer Response

The negative predictive effect of high scores on the Warpy Thoughts Quiz was evident in both models, but was not supported by the Bayes factor, which challenges the robustness of the finding. The Warpy Thoughts Quiz has not been used previously in studies of prediction. It is not entirely equivalent to the much-used Dysfunctional Attitude Scale [131], but taps into many of the same constructs including perfectionism and the need for success, love, and approval [96,132]. Worse treatment response has been associated with higher levels of dysfunctional attitudes in previous studies of face-to-face CBT [29,39,48,49] and some studies of ICBT [59], but not others [50,56]. Dysfunctional attitudes moderated treatment response in one study in which those with severe dysfunctional attitudes responded better to IPT and those with lower levels experienced better effects with CBT [29]. Again, this can indicate that this variable may be a predictor of response to treatment rather than a predictor of general prognosis. A proposed explanation is that patients having less severe dysfunctional attitudes may have greater cognitive flexibility [39] making them more able to profit from utilizing cognitive techniques [29].

The Role of Expectations and Motivation

The results were somewhat mixed for treatment expectancy, attitude, and motivation with some analyses indicating a possible negative effect of motivation and expectancy, whereas the Bayes factors indicated substantial support for no effect for all 3 variables. The lack of effects in this study is inconsistent with our hypothesis and with previous studies of face-to-face therapy in which expectancy is considered an important predictor of outcome [29,31]. However, results have been inconclusive with respect to ICBT [60,61]. These results may be due to the fact that most individuals entering a research trial have fairly positive attitudes, expectations, and high motivation, which restricts the range of these variables as is reflected by the distributions displayed in Figure S5 in Multimedia Appendix 1. These variables may be more valuable predictors in a regular practice setting. In addition, these constructs were assessed using single items, which call the validity of these measures into question.

Conclusion

The findings of the present study indicate that within a population of primary care patients with mild to moderate depression, treatment response to Web-based CBT with face-to-face guidance of high intensity was comparable across varying levels of initial depressive severity and irrespective of the presence and severity of comorbid anxiety. Whether the treatment is suitable for more severe depression is still uncertain. Treatment effects were also comparable for men and women and for patients of various ages. Being married or cohabiting and reporting higher life satisfaction predicted more favorable response to treatment. More positive response was also indicated for individuals with more previous depressive episodes, whereas having a higher level of dysfunctional thinking may predict poorer treatment response.

The purpose of this paper was primarily exploratory. Therefore, the results must be interpreted as hypotheses to inform further research rather than firm conclusions. Nevertheless, the results add to the knowledge base concerning differential treatment response, knowledge that is crucial for further implementation of Internet-based treatments in regular practice. Future studies should continue to explore predictors and, preferably, moderators of different Internet-based treatments compared to face-to-face treatments. In addition, studies exploring different patterns of response may also give important information about the differential response of various subgroups of patients.

Acknowledgments

This study was financially funded by the Research Council of Norway (196423/V50). MoodGYM was provided by the Australian National University and the authors would like to acknowledge the contribution to the project by Professor Kathleen M Griffiths, director at the National Institute for Mental Health Research at ANU. We would also like to thank Anthony Bennett, IT manager of the e-hub, the e-mental health research and development group at the National Institute for Mental Health Research at ANU and Kylie Bennett, Collaborative Research and Development Manager at e-hub, who provided technical expertise in administration of the trial; and Ada Tam, Web Developer at ANU, who provided technical support for users in the trial. Last, but not least, we would like to give a special thanks to those participating in the trial.

Authors' Contributions

RSH and MM wrote the paper. MM conducted the statistical analyses and OF contributed to the statistical analyses. All authors contributed to project design, data collection, and/or preparation of the manuscript.

Conflicts of Interest

KW and ME contributed in the process of translating MoodGYM into Norwegian. RSH, MM, KL, SKK, NK, and OF have no financial or nonfinancial interests to declare in relation to this study.

Multimedia Appendix 1

Supplemental material: Details on statistical methods and additional analyses.

[PDF File (Adobe PDF File), 1MB - jmir_v17i9e197_app1.pdf]

Multimedia Appendix 2

CONSORT-EHEALTH checklist V1.6.2 [133].

[PDF File (Adobe PDF File), 151KB - jmir_v17i9e197_app2.pdf]

References

1. Cuijpers P, van Straten A, Andersson G, van Oppen P. Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. J Consult Clin Psychol 2008 Dec;76(6):909-922. [doi: 10.1037/a0013075] [Medline: 19045960]
2. Churchill R, Hunot V, Corney R, Knapp M, McGuire H, Tylee A, et al. A systematic review of controlled trials of the effectiveness and cost-effectiveness of brief psychological treatments for depression. Health Technol Assess 2001;5(35):1-173 [FREE Full text] [Medline: 12387733]
3. Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. Clin Psychol Rev 2006 Jan;26(1):17-31. [doi: 10.1016/j.cpr.2005.07.003] [Medline: 16199119]
4. Ruwaard J, Schrieken B, Schrijver M, Broeksteeg J, Dekker J, Vermeulen H, et al. Standardized web-based cognitive behavioural therapy of mild to moderate depression: a randomized controlled trial with a long-term follow-up. Cogn Behav Ther 2009 Dec;38(4):206-221. [doi: 10.1080/16506070802408086] [Medline: 19221919]
5. Andersson G, Bergström J, Holländare F, Carlbring P, Kaldo V, Ekselius L. Internet-based self-help for depression: randomised controlled trial. Br J Psychiatry 2005 Nov;187:456-461 [FREE Full text] [doi: 10.1192/bjp.187.5.456] [Medline: 16260822]
6. Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F. Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. Cogn Behav Ther 2011 Dec;40(4):251-266. [doi: 10.1080/16506073.2011.616531] [Medline: 22060248]
7. Johansson R, Sjöberg E, Sjögren M, Johnsson E, Carlbring P, Andersson T, et al. Tailored vs. standardized internet-based cognitive behavior therapy for depression and comorbid symptoms: a randomized controlled trial. PLoS One 2012;7(5):e36905 [FREE Full text] [doi: 10.1371/journal.pone.0036905] [Medline: 22615841]

8. Lintvedt OK, Griffiths KM, Sørensen K, Østvik AR, Wang CEA, Eisemann M, et al. Evaluating the effectiveness and efficacy of unguided internet-based self-help intervention for the prevention of depression: a randomized controlled trial. Clin Psychol Psychother 2013;20(1):10-27. [doi: 10.1002/cpp.770] [Medline: 21887811]
9. Cuijpers P, Donker T, van Straten A, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. Psychol Med 2010 Dec;40(12):1943-1957. [doi: 10.1017/S0033291710000772] [Medline: 20406528]
10. 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]
11. Wagner B, Horn AB, Maercker A. Internet-based versus face-to-face cognitive-behavioral intervention for depression: a randomized controlled non-inferiority trial. J Affect Disord 2014 Jan;152-154:113-121. [doi: 10.1016/j.jad.2013.06.032] [Medline: 23886401]
12. Vernmark K, Lenndin J, Bjärehed J, Carlsson M, Karlsson J, Öberg J, et al. Internet administered guided self-help versus individualized e-mail therapy: A randomized trial of two versions of CBT for major depression. Behav Res Ther 2010 May;48(5):368-376. [doi: 10.1016/j.brat.2010.01.005] [Medline: 20152960]
13. Høifødt RS, Lillevoll KR, Griffiths KM, Wilsgaard T, Eisemann M, Waterloo K, et al. The clinical effectiveness of web-based cognitive behavioral therapy with face-to-face therapist support for depressed primary care patients: randomized controlled trial. J Med Internet Res 2013;15(8):e153 [FREE Full text] [doi: 10.2196/jmir.2714] [Medline: 23916965]
14. Morris BH, Bylsma LM, Rottenberg J. Does emotion predict the course of major depressive disorder? A review of prospective studies. Br J Clin Psychol 2009 Sep;48(Pt 3):255-273. [doi: 10.1348/014466508X396549] [Medline: 19187578]
15. Rhebergen D, Beekman AT, Graaf RD, Nolen WA, Spijker J, Hoogendijk WJ, et al. The three-year naturalistic course of major depressive disorder, dysthymic disorder and double depression. J Affect Disord 2009 Jun;115(3):450-459. [doi: 10.1016/j.jad.2008.10.018] [Medline: 19042028]
16. Gilchrist G, Gunn J. Observational studies of depression in primary care: what do we know? BMC Fam Pract 2007;8:28 [FREE Full text] [doi: 10.1186/1471-2296-8-28] [Medline: 17493280]
17. van Beljouw IM, Verhaak PF, Cuijpers P, van Marwijk HW, Penninx BW. The course of untreated anxiety and depression, and determinants of poor one-year outcome: a one-year cohort study. BMC Psychiatry 2010;10:86 [FREE Full text] [doi: 10.1186/1471-244X-10-86] [Medline: 20961414]
18. Spijker J, de Graaf R, Bijl RV, Beekman AT, Ormel J, Nolen WA. Determinants of persistence of major depressive episodes in the general population. Results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). J Affect Disord 2004 Sep;81(3):231-240. [doi: 10.1016/j.jad.2003.08.005] [Medline: 15337327]
19. Rubenstein LV, Rayburn NR, Keeler EB, Ford DE, Rost KM, Sherbourne CD. Predicting outcomes of primary care patients with major depression: development of a depression prognosis index. Psychiatr Serv 2007 Aug;58(8):1049-1056. [doi: 10.1176/appi.ps.58.8.1049] [Medline: 17664515]
20. Stegenga BT, Kamphuis MH, King M, Nazareth I, Geerlings MI. The natural course and outcome of major depressive disorder in primary care: the PREDICT-NL study. Soc Psychiatry Psychiatr Epidemiol 2012 Jan;47(1):87-95 [FREE Full text] [doi: 10.1007/s00127-010-0317-9] [Medline: 21057769]
21. Ambresin G, Chondros P, Dowrick C, Herrman H, Gunn JM. Self-rated health and long-term prognosis of depression. Ann Fam Med 2014;12(1):57-65 [FREE Full text] [doi: 10.1370/afm.1562] [Medline: 24445104]
22. Riihimäki KA, Vuorilehto MS, Melartin TK, Isometsä ET. Five-year outcome of major depressive disorder in primary health care. Psychol Med 2014 May;44(7):1369-1379. [doi: 10.1017/S0033291711002303] [Medline: 22085687]
23. Dowrick C, Shiels C, Page H, Ayuso-Mateos JL, Casey P, Dalgard OS, et al. Predicting long-term recovery from depression in community settings in Western Europe: evidence from ODIN. Soc Psychiatry Psychiatr Epidemiol 2011 Feb;46(2):119-126. [doi: 10.1007/s00127-009-0179-1] [Medline: 20035318]
24. Skodol AE, Grilo CM, Keyes KM, Geier T, Grant BF, Hasin DS. Relationship of personality disorders to the course of major depressive disorder in a nationally representative sample. Am J Psychiatry 2011 Mar;168(3):257-264 [FREE Full text] [doi: 10.1176/appi.ajp.2010.10050695] [Medline: 21245088]
25. Driessen E, Hollon SD. Cognitive behavioral therapy for mood disorders: efficacy, moderators and mediators. Psychiatr Clin North Am 2010 Sep;33(3):537-555 [FREE Full text] [doi: 10.1016/j.psc.2010.04.005] [Medline: 20599132]
26. Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. Arch Gen Psychiatry 2002 Oct;59(10):877-883. [Medline: 12365874]
27. Simon GE, Perlis RH. Personalized medicine for depression: can we match patients with treatments? Am J Psychiatry 2010 Dec;167(12):1445-1455 [FREE Full text] [doi: 10.1176/appi.ajp.2010.09111680] [Medline: 20843873]
28. Tadić A, Helmreich I, Mergl R, Hautzinger M, Kohnen R, Henkel V, et al. Early improvement is a predictor of treatment outcome in patients with mild major, minor or subsyndromal depression. J Affect Disord 2010 Jan;120(1-3):86-93. [doi: 10.1016/j.jad.2009.04.014] [Medline: 19428118]
29. Sotsky SM, Glass DR, Shea MT, Pilkonis PA, Collins JF, Elkin I, et al. Patient predictors of response to psychotherapy and pharmacotherapy: findings in the NIMH Treatment of Depression Collaborative Research Program. Am J Psychiatry 1991 Aug;148(8):997-1008. [doi: 10.1176/ajp.148.8.997] [Medline: 1853989]

30. Carter JD, Luty SE, McKenzie JM, Mulder RT, Frampton CM, Joyce PR. Patient predictors of response to cognitive behaviour therapy and interpersonal psychotherapy in a randomised clinical trial for depression. J Affect Disord 2011 Feb;128(3):252-261. [doi: 10.1016/j.jad.2010.07.002] [Medline: 20674982]
31. Meyer B, Pilkonis PA, Krupnick JL, Egan MK, Simmens SJ, Sotsky SM. Treatment expectancies, patient alliance, and outcome: further analyses from the National Institute of Mental Health Treatment of Depression Collaborative Research Program. J Consult Clin Psychol 2002 Aug;70(4):1051-1055. [Medline: 12182269]
32. Lutz W, Stulz N, Köck K. Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. J Affect Disord 2009 Nov;118(1-3):60-68. [doi: 10.1016/j.jad.2009.01.019] [Medline: 19217669]
33. Hamilton KE, Dobson KS. Cognitive therapy of depression: pretreatment patient predictors of outcome. Clin Psychol Rev 2002 Jul;22(6):875-893. [Medline: 12214329]
34. Van HL, Schoevers RA, Dekker J. Predicting the outcome of antidepressants and psychotherapy for depression: a qualitative, systematic review. Harv Rev Psychiatry 2008;16(4):225-234. [doi: 10.1080/10673220802277938] [Medline: 18661365]
35. Thase ME, Reynolds CF, Frank E, Simons AD, McGeary J, Fasiczka AL, et al. Do depressed men and women respond similarly to cognitive behavior therapy? Am J Psychiatry 1994 Apr;151(4):500-505. [Medline: 8147447]
36. Cuijpers P, Weitz E, Twisk J, Kuehner C, Cristea I, David D, et al. Gender as predictor and moderator of outcome in cognitive behavior therapy and pharmacotherapy for adult depression: an "individual patient data" meta-analysis. Depress Anxiety 2014 Nov;31(11):941-951. [doi: 10.1002/da.22328] [Medline: 25407584]
37. Button KS, Turner N, Campbell J, Kessler D, Kuyken W, Lewis G, et al. Moderators of response to cognitive behavioural therapy as an adjunct to pharmacotherapy for treatment-resistant depression in primary care. J Affect Disord 2015 Mar 15;174:272-280. [doi: 10.1016/j.jad.2014.11.057] [Medline: 25527998]
38. Fournier JC, DeRubeis RJ, Shelton RC, Hollon SD, Amsterdam JD, Gallop R. Prediction of response to medication and cognitive therapy in the treatment of moderate to severe depression. J Consult Clin Psychol 2009 Aug;77(4):775-787 [FREE Full text] [doi: 10.1037/a0015401] [Medline: 19634969]
39. Jarrett RB, Eaves GG, Grannemann BD, Rush AJ. Clinical, cognitive, and demographic predictors of response to cognitive therapy for depression: a preliminary report. Psychiatry Res 1991 Jun;37(3):245-260. [Medline: 1891508]
40. Barber JP, Muenz LR. The role of avoidance and obsessiveness in matching patients to cognitive and interpersonal psychotherapy: empirical findings from the treatment for depression collaborative research program. J Consult Clin Psychol 1996 Oct;64(5):951-958. [Medline: 8916624]
41. Elkin I, Gibbons RD, Shea MT, Sotsky SM, Watkins JT, Pilkonis PA, et al. Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. J Consult Clin Psychol 1995 Oct;63(5):841-847. [Medline: 7593878]
42. Shapiro DA, Barkham M, Rees A, Hardy GE, Reynolds S, Startup M. Effects of treatment duration and severity of depression on the effectiveness of cognitive-behavioral and psychodynamic-interpersonal psychotherapy. J Consult Clin Psychol 1994 Jun;62(3):522-534. [Medline: 8063978]
43. Thase ME, Simons AD, Cahalane J, McGeary J, Harden T. Severity of depression and response to cognitive behavior therapy. Am J Psychiatry 1991 Jun;148(6):784-789. [Medline: 2035722]
44. Gelhart RP, King HL. The influence of comorbid risk factors on the effectiveness of cognitive-behavioral treatment of depression. Cognitive and Behavioral Practice 2001 Dec;8(1):18-28. [doi: 10.1016/S1077-7229(01)80039-0]
45. Scheibe G, Albus M. Prospective follow-up study lasting 2 years in patients with panic disorder with and without depressive disorders. Eur Arch Psychiatry Clin Neurosci 1994;244(1):39-44. [Medline: 7918700]
46. Forand NR, Derubeis RJ. Pretreatment anxiety predicts patterns of change in cognitive behavioral therapy and medications for depression. J Consult Clin Psychol 2013 Oct;81(5):774-782 [FREE Full text] [doi: 10.1037/a0032985] [Medline: 23647285]
47. Smits JA, Minhajuddin A, Thase ME, Jarrett RB. Outcomes of acute phase cognitive therapy in outpatients with anxious versus nonanxious depression. Psychother Psychosom 2012;81(3):153-160 [FREE Full text] [doi: 10.1159/000334909] [Medline: 22398963]
48. Rude SS, Rehm LP. Response to treatments for depression: The role of initial status on targeted cognitive and behavioral skills. Clinical Psychology Review 1991 Jan;11(5):493-514. [doi: 10.1016/0272-7358(91)90001-B]
49. Blatt SJ, Zuroff DC, Hawley LL, Auerbach JS. Predictors of sustained therapeutic change. Psychother Res 2010 Jan;20(1):37-54. [doi: 10.1080/10503300903121080] [Medline: 19757328]
50. de Graaf LE, Hollon SD, Huibers MJ. Predicting outcome in computerized cognitive behavioral therapy for depression in primary care: A randomized trial. J Consult Clin Psychol 2010 Apr;78(2):184-189. [doi: 10.1037/a0018324] [Medline: 20350029]
51. Andersson G, Bergström J, Holländare F, Ekselius L, Carlbring P. Delivering cognitive behavioural therapy for mild to moderate depression via the Internet: predicting outcome at 6-month follow-up. Verhaltenstherapie 2004;14(3):185-189. [doi: 10.1159/000080914]
52. Spek V, Nyklícek I, Cuijpers P, Pop V. Predictors of outcome of group and internet-based cognitive behavior therapy. J Affect Disord 2008 Jan;105(1-3):137-145. [doi: 10.1016/j.jad.2007.05.001] [Medline: 17543392]

53. Button KS, Wiles NJ, Lewis G, Peters TJ, Kessler D. Factors associated with differential response to online cognitive behavioural therapy. Soc Psychiatry Psychiatr Epidemiol 2012 May;47(5):827-833. [doi: 10.1007/s00127-011-0389-1] [Medline: 21541696]
54. Moritz S, Schilling L, Hauschildt M, Schröder J, Treszl A. A randomized controlled trial of internet-based therapy in depression. Behav Res Ther 2012 Aug;50(7-8):513-521. [doi: 10.1016/j.brat.2012.04.006] [Medline: 22677231]
55. 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]
56. Warmerdam L, Van Straten A, 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]
57. Farrer LM, Griffiths KM, Christensen H, Mackinnon AJ, Batterham PJ. Predictors of adherence and outcome in Internet-based cognitive behavior therapy delivered in a telephone counseling setting. Cogn Ther Res 2013 Oct 27;38(3):358-367. [doi: 10.1007/s10608-013-9589-1]
58. Clarke G, Kelleher C, Hornbrook M, Debar L, Dickerson J, Gullion C. Randomized effectiveness trial of an Internet, pure self-help, cognitive behavioral intervention for depressive symptoms in young adults. Cogn Behav Ther 2009;38(4):222-234 [FREE Full text] [doi: 10.1080/16506070802675353] [Medline: 19440896]
59. Donker T, Batterham PJ, Warmerdam L, Bennett K, Bennett A, Cuijpers P, et al. Predictors and moderators of response to internet-delivered Interpersonal psychotherapy and cognitive behavior therapy for depression. J Affect Disord 2013 Oct;151(1):343-351. [doi: 10.1016/j.jad.2013.06.020] [Medline: 23953024]
60. de Graaf LE, Huibers MJ, Riper H, Gerhards SA, Arntz A. Use and acceptability of unsupported online computerized cognitive behavioral therapy for depression and associations with clinical outcome. J Affect Disord 2009 Aug;116(3):227-231. [doi: 10.1016/j.jad.2008.12.009] [Medline: 19167094]
61. Cavanagh K, Shapiro DA, Van Den Berg S, Swain S, Barkham M, Proudfoot J. The acceptability of computer-aided cognitive behavioural therapy: a pragmatic study. Cogn Behav Ther 2009;38(4):235-246. [doi: 10.1080/16506070802561256] [Medline: 19306147]
62. Meyer B, Berger T, Caspar F, Beevers CG, Andersson G, Weiss M. Effectiveness of a novel integrative online treatment for depression (Deprexis): randomized controlled trial. J Med Internet Res 2009;11(2):e15 [FREE Full text] [doi: 10.2196/jmir.1151] [Medline: 19632969]
63. Hilvert-Bruce Z, Rossouw PJ, Wong N, Sunderland M, Andrews G. Adherence as a determinant of effectiveness of internet cognitive behavioural therapy for anxiety and depressive disorders. Behav Res Ther 2012 Aug;50(7-8):463-468. [doi: 10.1016/j.brat.2012.04.001] [Medline: 22659155]
64. Christensen H, Griffiths K, Groves C, Korten A. Free range users and one hit wonders: community users of an Internet-based cognitive behaviour therapy program. Aust N Z J Psychiatry 2006 Jan;40(1):59-62. [doi: 10.1111/j.1440-1614.2006.01743.x] [Medline: 16403040]
65. Christensen H, Griffiths KM, Korten A. Web-based cognitive behavior therapy: analysis of site usage and changes in depression and anxiety scores. J Med Internet Res 2002;4(1):e3 [FREE Full text] [doi: 10.2196/jmir.4.1.e3] [Medline: 11956035]
66. Warmerdam L, van Straten A, Twisk J, Riper H, Cuijpers P. Internet-based treatment for adults with depressive symptoms: randomized controlled trial. J Med Internet Res 2008;10(4):e44 [FREE Full text] [doi: 10.2196/jmir.1094] [Medline: 19033149]
67. Clarke G, Eubanks D, Reid E, Kelleher C, O'Connor E, DeBar LL, et al. Overcoming Depression on the Internet (ODIN) (2): a randomized trial of a self-help depression skills program with reminders. J Med Internet Res 2005;7(2):e16 [FREE Full text] [doi: 10.2196/jmir.7.2.e16] [Medline: 15998607]
68. Clarke G, Reid E, Eubanks D, O'Connor E, DeBar LL, Kelleher C, et al. Overcoming Depression on the Internet (ODIN): a randomized controlled trial of an Internet depression skills intervention program. J Med Internet Res 2002 Dec;4(3):E14 [FREE Full text] [doi: 10.2196/jmir.4.3.e14] [Medline: 12554545]
69. 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;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
70. Davidsen A. Experiences of carrying out talking therapy in general practice: a qualitative interview study. Patient Educ Couns 2008 Aug;72(2):268-275. [doi: 10.1016/j.pec.2008.03.020] [Medline: 18472244]
71. Backenstrass M, Joest K, Rosemann T, Szecsenyi J. The care of patients with subthreshold depression in primary care: is it all that bad? A qualitative study on the views of general practitioners and patients. BMC Health Serv Res 2007;7:190 [FREE Full text] [doi: 10.1186/1472-6963-7-190] [Medline: 18031573]
72. Mykletun A, Knudsen AK, Tangen T, Øverland S. General practitioners' opinions on how to improve treatment of mental disorders in primary health care. Interviews with one hundred Norwegian general practitioners. BMC Health Serv Res 2010;10:35 [FREE Full text] [doi: 10.1186/1472-6963-10-35] [Medline: 20144205]

73. Wittchen HU, Jacobi F, Rehm J, Gustavsson A, Svensson M, Jönsson B, et al. The size and burden of mental disorders and other disorders of the brain in Europe 2010. Eur Neuropsychopharmacol 2011 Sep;21(9):655-679. [doi: 10.1016/j.euroneuro.2011.07.018] [Medline: 21896369]
74. Australian National University, Centre for Mental Health Research. The MoodGYM training program: mark III: learn cognitive behaviour therapy skills for preventing and coping with depression. 2008. URL: https://moodgym.anu.edu.au/welcome [accessed 2013-05-10] [WebCite Cache ID 6GVUeWUTc]
75. Bennett K, Reynolds J, Christensen H, Griffiths KM. e-hub: an online self-help mental health service in the community. Med J Aust 2010 Jun 7;192(11 Suppl):S48-S52. [Medline: 20528710]
76. Calear AL, Christensen H, Mackinnon A, Griffiths KM, O'Kearney R. The YouthMood Project: a cluster randomized controlled trial of an online cognitive behavioral program with adolescents. J Consult Clin Psychol 2009 Dec;77(6):1021-1032. [doi: 10.1037/a0017391] [Medline: 19968379]
77. Powell J, Hamborg T, Stallard N, Burls A, McSorley J, Bennett K, et al. Effectiveness of a web-based cognitive-behavioral tool to improve mental well-being in the general population: randomized controlled trial. J Med Internet Res 2013;15(1):e2 [FREE Full text] [doi: 10.2196/jmir.2240] [Medline: 23302475]
78. Christensen H, Griffiths KM, Jorm AF. Delivering interventions for depression by using the internet: randomised controlled trial. BMJ 2004 Jan 31;328(7434):265 [FREE Full text] [doi: 10.1136/bmj.37945.566632.EE] [Medline: 14742346]
79. Farrer L, Christensen H, Griffiths KM, Mackinnon A. Internet-based CBT for depression with and without telephone tracking in a national helpline: randomised controlled trial. PLoS One 2011;6(11):e28099 [FREE Full text] [doi: 10.1371/journal.pone.0028099] [Medline: 22140514]
80. Beck A, Steer R, Brown G. BDI-II, Beck Depression Inventory: Manual. San Antonio, TX: The Psychological Corporation; 1996.
81. Arnau RC, Meagher MW, Norris MP, Bramson R. Psychometric evaluation of the Beck Depression Inventory-II with primary care medical patients. Health Psychol 2001 Mar;20(2):112-119. [Medline: 11315728]
82. Dozois DJA, Dobson KS, Ahnberg JL. A psychometric evaluation of the Beck Depression Inventory-II. Psychological Assessment 1998;10(2):83-89. [doi: 10.1037/1040-3590.10.2.83]
83. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res 2002 Feb;52(2):69-77. [Medline: 11832252]
84. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [Medline: 6880820]
85. Beck A, Steer R. Beck Anxiety Inventory: Manual. San Antonio, TX: The Psychological Corporation; 1993.
86. Fydrich T, Dowdall D, Chambless DL. Reliability and validity of the Beck Anxiety Inventory. Journal of Anxiety Disorders 1992 Jan;6(1):55-61. [doi: 10.1016/0887-6185%2892%2990026-4]
87. Steer RA, Ranieri WF, Beck AT, Clark DA. Further evidence for the validity of the beck anxiety inventory with psychiatric outpatients. Journal of Anxiety Disorders 1993 Jul;7(3):195-205. [doi: 10.1016/0887-6185%2893%2990002-3]
88. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. J Consult Clin Psychol 1988 Dec;56(6):893-897. [Medline: 3204199]
89. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33;quiz 34. [Medline: 9881538]
90. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. Addiction 1993 Jun;88(6):791-804. [Medline: 8329970]
91. Reinert DF, Allen JP. The alcohol use disorders identification test: an update of research findings. Alcohol Clin Exp Res 2007 Feb;31(2):185-199. [doi: 10.1111/j.1530-0277.2006.00295.x] [Medline: 17250609]
92. EuroQol Group. EQ-5D. 2013. About EQ-5D URL: http://www.euroqol.org/about-eq-5d.html [accessed 2013-05-08] [WebCite Cache ID 6GSukzvVz]
93. Diener E, Emmons RA, Larsen RJ, Griffin S. The Satisfaction With Life Scale. J Pers Assess 1985 Feb;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
94. Pavot W, Diener E. Review of the Satisfaction With Life Scale. Psychological Assessment 1993;5(2):164-172. [doi: 10.1037/1040-3590.5.2.164]
95. Glaesmer H, Grande G, Braehler E, Roth M. The German version of the Satisfaction With Life Scale (SWLS). European Journal of Psychological Assessment 2011 May 3;27(2):127-132. [doi: 10.1027/1015-5759/a000058]
96. Christensen H, Griffiths K, Groves C. MoodGYM Training Program: Clinician's Manual: A Manual for the Use of the MoodGYM Training Program in Clinical Practice. Canberra: Centre for Mental Health Research; 2004.
97. Jorm AF, Medway J, Christensen H, Korten AE, Jacomb PA, Rodgers B. Attitudes towards people with depression: effects on the public's help-seeking and outcome when experiencing common psychiatric symptoms. Aust N Z J Psychiatry 2000 Aug;34(4):612-618. [Medline: 10954392]
98. Parslow RA, Christensen H, Griffiths KM, Groves C. The Warpy Thoughts Scale: a new 20-item instrument to measure dysfunctional attitudes. Cogn Behav Ther 2006;35(2):106-116. [doi: 10.1080/16506070500372279] [Medline: 16754265]

99. Schwarzer R, Jerusalem M. Generalized Self-Efficacy Scale. In: Weinman J, Wright S, Johnston M, editors. Measures in Health Psychology: A User's Portfolio Causal and Control Beliefs. Windsor, UK: NFER-NELSON; 1995:35-37.
100. Leganger A, Kraft P, R ysamb E. Perceived self-efficacy in health behaviour research: conceptualisation, measurement and correlates. Psychology & Health 2000 Feb;15(1):51-69. [doi: 10.1080/08870440008400288]
101. Luszczynska A, Gutiérrez - Doña B, Schwarzer R. General self - efficacy in various domains of human functioning: evidence from five countries. International Journal of Psychology 2005 Apr;40(2):80-89. [doi: 10.1080/00207590444000041]
102. Luszczynska A, Scholz U, Schwarzer R. The general self-efficacy scale: multicultural validation studies. J Psychol 2005 Sep;139(5):439-457. [doi: 10.3200/JRLP.139.5.439-457] [Medline: 16285214]
103. Scholz U, Gutiérrez Doña B, Sud S, Schwarzer R. Is general self-efficacy a universal construct? European Journal of Psychological Assessment 2002 Sep 1;18(3):242-251. [doi: 10.1027//1015-5759.18.3.242]
104. Kruschke J. Doing Bayesian Data Analysis: A Tutorial with R and BUGS. Oxford: Academic Press; 2011.
105. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. 3rd edition. Boca Raton, FL: Chapman & Hall/CRC Press; 2013.
106. Wagenmakers E. A practical solution to the pervasive problems of p values. Psychon Bull Rev 2007 Oct;14(5):779-804. [Medline: 18087943]
107. Wagenmakers E, Wetzels R, Borsboom D, van der Maas HL. Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). J Pers Soc Psychol 2011 Mar;100(3):426-432. [doi: 10.1037/a0022790] [Medline: 21280965]
108. Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med 1999 Jun 15;130(12):995-1004. [Medline: 10383371]
109. Kruschke JK. Bayesian estimation supersedes the t test. J Exp Psychol Gen 2013 May;142(2):573-603. [doi: 10.1037/a0029146] [Medline: 22774788]
110. Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G. Bayesian t tests for accepting and rejecting the null hypothesis. Psychon Bull Rev 2009 Apr;16(2):225-237. [doi: 10.3758/PBR.16.2.225] [Medline: 19293088]
111. Jeffreys H. Theory of probability. Oxford: Oxford University Press; 1961.
112. Wetzels R, Matzke D, Lee MD, Rouder JN, Iverson GJ, Wagenmakers E. Statistical evidence in experimental psychology: an empirical comparison using 855 t tests. Perspectives on Psychological Science 2011 May 18;6(3):291-298. [doi: 10.1177/1745691611406923]
113. Gelman A, Shalizi CR. Philosophy and the practice of Bayesian statistics. Br J Math Stat Psychol 2013 Feb;66(1):8-38. [doi: 10.1111/j.2044-8317.2011.02037.x] [Medline: 22364575]
114. Plummer M. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. 2003 Presented at: Proceedings of the 3rd International Workshop on Distributed Statistical Computing; March 20–22, 2003; Vienna, Austria.
115. Gelman A, Rubin D. Inference from iterative simulation using multiple sequences. Stat Sci 1992;7(4):457-472.
116. Jung T, Wickrama KA. An introduction to latent class growth analysis and growth mixture modeling. Social Pers Psych Compass 2008 Jan;2(1):302-317. [doi: 10.1111/j.1751-9004.2007.00054.x]
117. Tofighi D, Enders CK. Identifying the correct number of classes in growth mixture models. In: Hancock GR, Samuelsen KM, editors. Advances in Latent Variable Mixture Models. Greenwich, CT: Information Age Publishing; 2007.
118. Muthén B. Latent variable analysis: growth mixture modeling and related techniques for longitudinal data. In: Kaplan D, editor. The Sage Handbook of Quantitative Methodology for the Social Sciences. Thousand Oaks, CA: SAGE Publications; 2004:345-368.
119. Stulz N, Lutz W, Leach C, Lucock M, Barkham M. Shapes of early change in psychotherapy under routine outpatient conditions. J Consult Clin Psychol 2007 Dec;75(6):864-874. [doi: 10.1037/0022-006X.75.6.864] [Medline: 18085904]
120. Muthén B, Brown HC, Hunter AM, Cook IA, Leuchter AF. General approaches to analysis of course: applying growth mixture modeling to randomized trials of depression medication. In: Shrout PE, Keyes K, Ornstein K, editors. Causality and Psychopathology: Finding the Determinants of Disorders and their Cures. New York: Oxford University Press; 2010:159-177.
121. Cuijpers P, Reynolds CF, Donker T, Li J, Andersson G, Beekman A. Personalized treatment of adult depression: medication, psychotherapy, or both? A systematic review. Depress Anxiety 2012 Oct;29(10):855-864. [doi: 10.1002/da.21985] [Medline: 22815247]
122. Gelman A, Carlin J. Beyond power calculations: assessing type S (sign) and type M (magnitude) errors. Perspectives on Psychological Science 2014 Nov 17;9(6):641-651. [doi: 10.1177/1745691614551642]
123. Köhler S, Unger T, Hoffmann S, Steinacher B, Fydrich T. Acute and long-term treatment outcome in depressed inpatients with vs. without anxious features: results of a one-year follow-up study. J Affect Disord 2013 Sep 25;150(3):1055-1061. [doi: 10.1016/j.jad.2013.05.043] [Medline: 23764380]
124. Persons J, Bostrom A, Bertagnolli A. Results of randomized controlled trials of cognitive therapy for depression generalize to private practice. Cognit Ther Res 1999;23(5):535-548. [doi: 10.1023/A:1018724505659]
125. Wilhelmsen M, Lillevoll K, Risør MB, Høifødt R, Johansen M, Waterloo K, et al. Motivation to persist with internet-based cognitive behavioural treatment using blended care: a qualitative study. BMC Psychiatry 2013;13:296 [FREE Full text] [doi: 10.1186/1471-244X-13-296] [Medline: 24199672]

126. Dudek D, Rybakowski JK, Siwek M, Pawłowski T, Lojko D, Roczeń R, et al. Risk factors of treatment resistance in major depression: association with bipolarity. J Affect Disord 2010 Oct;126(1-2):268-271. [doi: 10.1016/j.jad.2010.03.001] [Medline: 20381154]
127. Bennabi D, Aouizerate B, El-Hage W, Doumy O, Moliere F, Courtet P, et al. Risk factors for treatment resistance in unipolar depression: a systematic review. J Affect Disord 2015 Jan 15;171:137-141. [doi: 10.1016/j.jad.2014.09.020] [Medline: 25305428]
128. Cuijpers P, Dekker J, Hollon SD, Andersson G. Adding psychotherapy to pharmacotherapy in the treatment of depressive disorders in adults: a meta-analysis. J Clin Psychiatry 2009 Sep;70(9):1219-1229. [doi: 10.4088/JCP.09r05021] [Medline: 19818243]
129. 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 2002 Sep;181:208-213 [FREE Full text] [Medline: 12204924]
130. 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]
131. Weissman A, Beck A. Development and validation of the Dysfunctional Attitude Scale: a preliminary investigation. 1978 Presented at: Annual Meeting of the American Educational Research Association (62nd); March 27-31, 1978; Toronto, ON.
132. Chioqueta AP, Stiles TC. Factor structure of the Dysfunctional Attitude Scale (Form A) and the Automatic Thoughts Questionnaire: an exploratory study. Psychol Rep 2006 Aug;99(1):239-247. [doi: 10.2466/pr0.99.1.239-247] [Medline: 17037474]
133. Eysenbach G, Consort- E. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

ANU: Australian National University
AUDIT: Alcohol Use Disorders Identification Test
BAI: Beck Anxiety Inventory
BDI-II: Beck Depression Inventory-II
BF: Bayes factor
CBT: cognitive behavioral therapy
EQ-5D: EuroQol 5-Dimension Self-Report Questionnaire
GP: general practitioner
GSE: General Self-efficacy Scale
HADS-A: Hospital Anxiety and Depression Scale-Anxiety subscale
HADS-D: Hospital Anxiety and Depression Scale-Depression subscale
HDI: highest density interval
ICBT: Internet-based cognitive behavioral therapy
IPT: interpersonal therapy
MCMC: Markov chain Monte Carlo
MDE: major depressive episode
NHST: null-hypothesis significance testing
Presp: probability of response
RCT: randomized controlled trial
SWLS: Satisfaction With Life Scale

Edited by G Eysenbach; submitted 17.02.15; peer-reviewed by K Button, B Meyer; comments to author 12.03.15; revised version received 08.05.15; accepted 08.07.15; published 02.09.15.

Please cite as:
Høifødt RS, Mittner M, Lillevoll K, Katla SK, Kolstrup N, Eisemann M, Friborg O, Waterloo K
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;17(9):e197
URL: http://www.jmir.org/2015/9/e197/
doi:10.2196/jmir.4351
PMID:26333818

Original Paper

What Online Communities Can Tell Us About Electronic Cigarettes and Hookah Use: A Study Using Text Mining and Visualization Techniques

Annie T Chen[1], MSIS, PhD; Shu-Hong Zhu[2], PhD; Mike Conway[2], PhD

[1]School of Information and Library Science, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[2]Department of Family Medicine and Public Health, University of California, San Diego, La Jolla, CA, United States

Corresponding Author:
Annie T Chen, MSIS, PhD
School of Information and Library Science
University of North Carolina at Chapel Hill
216 Lenoir Drive/CB #3360
100 Manning Hall
Chapel Hill, NC, 27514
United States
Phone: 1 919 962 8366
Fax: 1 919 962 8071
Email: atchen@email.unc.edu

Abstract

Background: The rise in popularity of electronic cigarettes (e-cigarettes) and hookah over recent years has been accompanied by some confusion and uncertainty regarding the development of an appropriate regulatory response towards these emerging products. Mining online discussion content can lead to insights into people's experiences, which can in turn further our knowledge of how to address potential health implications. In this work, we take a novel approach to understanding the use and appeal of these emerging products by applying text mining techniques to compare consumer experiences across discussion forums.

Objective: This study examined content from the websites Vapor Talk, Hookah Forum, and Reddit to understand people's experiences with different tobacco products. Our investigation involves three parts. First, we identified contextual factors that inform our understanding of tobacco use behaviors, such as setting, time, social relationships, and sensory experience, and compared the forums to identify the ones where content on these factors is most common. Second, we compared how the tobacco use experience differs with combustible cigarettes and e-cigarettes. Third, we investigated differences between e-cigarette and hookah use.

Methods: In the first part of our study, we employed a lexicon-based extraction approach to estimate prevalence of contextual factors, and then we generated a heat map based on these estimates to compare the forums. In the second and third parts of the study, we employed a text mining technique called topic modeling to identify important topics and then developed a visualization, Topic Bars, to compare topic coverage across forums.

Results: In the first part of the study, we identified two forums, Vapor Talk Health & Safety and the Stopsmoking subreddit, where discussion concerning contextual factors was particularly common. The second part showed that the discussion in Vapor Talk Health & Safety focused on symptoms and comparisons of combustible cigarettes and e-cigarettes, and the Stopsmoking subreddit focused on psychological aspects of quitting. Last, we examined the discussion content on Vapor Talk and Hookah Forum. Prominent topics included equipment, technique, experiential elements of use, and the buying and selling of equipment.

Conclusions: This study has three main contributions. Discussion forums differ in the extent to which their content may help us understand behaviors with potential health implications. Identifying dimensions of interest and using a heat map visualization to compare across forums can be helpful for identifying forums with the greatest density of health information. Additionally, our work has shown that the quitting experience can potentially be very different depending on whether or not e-cigarettes are used. Finally, e-cigarette and hookah forums are similar in that members represent a "hobbyist culture" that actively engages in information exchange. These differences have important implications for both tobacco regulation and smoking cessation intervention design.

(J Med Internet Res 2015;17(9):e220) doi:10.2196/jmir.4517

KEYWORDS

electronic cigarettes; hookah smoking; cigarettes; tobacco products; social media; text mining

Introduction

In recent years, researchers have begun to realize the value of social media (including online discussion forums) as a data source for understanding health-related phenomena. The pervasiveness, ubiquity, and real-time nature of social media makes it useful for biosurveillance applications such as mining for influenza mentions, as well as studies of information dissemination and public sentiment towards topics such as vaccination [1-3]. Various terms have been used to describe this new and growing field, including infodemiology [4], digital disease detection [5], and digital epidemiology [6]. Moreover, social media mining has also been employed to understand the public's impression of products that have health implications [7]. The content of health discussion forums can provide rich details concerning the context in which patients experience various health issues, including temporal and emotional factors, which may help us tailor information to fit their needs [8]. In recent years, there has been increased interest in leveraging the use of online social networks for interventions to promote population-level smoking cessation [9].

This study is focused on leveraging the rich detail that is often provided in discussion forums to understand more about the experiences of users of three tobacco products—combustible cigarettes, electronic cigarettes (e-cigarettes), and water pipes (also known as "hookah")—and their potential health implications. E-cigarettes have increasingly gained popularity, particularly in those markets with well-developed tobacco control policies like the United States and (parts of) the European Union [10-12]. Current smokers and tobacco users are more likely to try e-cigarettes than those who have never smoked or used tobacco [12]. Dual use of e-cigarettes and combustible cigarettes is common among smokers who are considering quitting in the next 6 months [13].

Previous literature has enumerated various motivations for e-cigarette use, including quitting smoking for health reasons, the belief that e-cigarettes are safer than regular cigarettes, e-cigarettes are cheaper than regular cigarettes, e-cigarettes are allowed in locations where regular cigarettes are not, avoiding disturbing others with secondhand smoke, the sheer pleasure of smoking, and "just because" [12,14]. Reasons for stopping use included users thinking they did not need them anymore or that they would not relapse to smoking if they stopped, poor product quality, no reduction in cravings, relapse to smoking, and the lack of efficacy in helping users to quit smoking [15].

Aside from e-cigarettes, there has been increasing concern about the growing use of hookah (also known by other names such as waterpipe, shisha, and hubble bubble) worldwide [16]. Hookah is a centuries old practice that experienced a resurgence in the Middle East in the 1990s [17]. A hookah consists of a bowl where the burning tobacco is placed, an ashtray, stem, air valve, water base, and one or more hoses and mouthpieces. During use, smoke from the burning tobacco descends to the bowl of water that it bubbles through and is then inhaled by the smoker through a mouthpiece.

Hookah use is often a social behavior, and hookah bars or lounges appear to play an important role in the increased popularity of hookah smoking [18]. Aspects of group use such as group size and the number of waterpipes available per table may affect toxicant exposure; thus, it is important to consider the social and contextual factors associated with use [19]. Factors that have contributed to the rise in hookah use include availability in cafes and restaurants, social aspects, affordability, appeal of hookah designs, sensory aspects of the hookah smoking experience, and media influence [20]. Predictors of hookah use include current and past cigarette use, and alcohol and marijuana use [21-23].

Use of hookah may have various negative health effects, for example, developing chronic obstructive pulmonary disease and chronic bronchitis, increased risk of lung cancer and esophageal cancer, and adverse effects on cardiovascular health [24]. However, previous research suggests that hookah users believe that hookah is less harmful than traditional cigarettes, and thus the argument has been made that there is a greater need for education concerning the potential health dangers of hookah use [22,25].

Though there is a considerable research currently being undertaken on the health effects of e-cigarettes and hookah, there is less work focused on how people are using these tobacco products in naturalistic settings. However, in recent years, there have been a number of studies that have investigated e-cigarette and hookah mentions in social media, including symptoms that were reported by participants in three discussion forums [26], sentiment towards e-cigarette and hookah use on Twitter [4], marketing of electronic cigarettes on Twitter [27], hookah references on Facebook profiles of American college students [28], and e-cigarette and hookah videos on YouTube [29].

There are many different kinds of social media, and it can be problematic to employ social media data from a single source, or even multiple sources, to make population-level inferences [30]. This is not what we endeavor to do in this study. Rather, we demonstrate methods that can be used to compare across data sources and mitigate the effects of source differences, to make inferences about the sample that is being studied. We also try to provide enough contextual detail to enable readers to understand the extent to which the results may be applicable to other populations and to generate hypotheses for future research. In this study, we used several data sources to understand how different online communities might address the same topic.

As far as we know, there has not been a text mining study that has taken a comparative approach to examining online communities and tobacco products, and more specifically, examining what the discussion content may suggest about the appeal and motivation for use. With this study, we have

endeavored to fill that gap. We selected multiple online communities, in order to develop a better sense of the diversity of online content with these products. We focused on six different discussion forums on three different websites: Vapor Talk, Hookah Forum, and Reddit. We expected that these samples might differ on a variety of characteristics and thus serve as an appropriate set of samples for comparison.

This study is structured into three distinct parts. In the first part, we employed a heat map visualization to compare different aspects of e-cigarette and hookah use behavior across multiple forums to identify the forums with the highest concentration of reports concerning social and contextual factors of e-cigarette and hookah use, including the settings where use behaviors occur (eg, restaurant, lounge, and party), time, social relationships, and sensory experience. The heat map facilitated a quick visual scan enabling us to determine which discussion forums might contain the richest discussions of behavior relevant to e-cigarette and hookah use, and thus, enabled selection of data subsets for further analyses.

In the second part, we integrated text mining and visualization techniques to render a visualization, Topic Bars, to compare discussion content in two forums: the Health & Safety forum on Vapor Talk, which is focused on e-cigarettes, and the Stopsmoking subreddit, which is primarily concerned with quitting traditional, combustible cigarettes (analogs).

In the third and last part, we compare experiences with e-cigarette and hookah use. How does the nature of content on these two products differ? We examined this question through a Topic Bars visualization depicting the general discussion forums for Vapor Talk (focused on e-cigarettes) and Hookah Forum (focused on hookah).

Methods

Harvesting of the Document Collection

We downloaded publicly available content from three websites: (1) Vapor Talk [31], a forum dedicated to e-cigarettes, (2) Hookah Forum [32], and (3) Reddit [33], a platform that hosts subforums or “subreddits” on a wide variety of topics.

Vapor Talk and Hookah Forum are online communities that are dedicated to e-cigarettes and hookah, respectively. At the time the data were collected, Vapor Talk and Hookah Forum appeared among the top results on the Google search engine when searching using keywords such as “e-cigarette”, “vaping”, “hookah”, “health”, and “forum”. Vapor Talk has also been examined in previous research [26]. Vapor Talk features a number of different forums; we selected “General E-Cig Discussion” and “Health & Safety.” These two forums were selected to acquire a general sense of what the nature of discussion concerning e-cigarettes is like, as well as the community’s specific health concerns.

Reddit is a generic platform featuring “subreddits” on a broad range of topics. The platform is more popular among younger people [34,35]. On Reddit, we examined the “stopsmoking”, “electronic_cigarette”, and “hookah” subreddits.

Publicly available content for each discussion forum was downloaded using a Web crawler, Wget, between April and June 2014. Crawls of each site focused on the discussion content, and no explicit attempt was made to crawl user profiles. The pages of discussion content from Vapor Talk and Hookah Forum include some basic user metadata such as username, gender, and member level. The post content and metadata were extracted using Python code and inserted into a MySQL database.

Comparing Contextual Factors of Tobacco Use Across Datasets

In this study, we were interested in using social media to understand more about differences in people’s experiences and motivations for using e-cigarettes and hookah, as an understanding of how consumers use different tobacco products is vital for both advancing tobacco regulatory science and smoking cessation intervention design. We identified a set of factors to use to compare across datasets. Understanding the factors that influence people’s behavior can be invaluable for developing strategies to encourage more healthful behaviors. Previous literature has argued that an individual’s behavior is affected by a variety of individual and social factors, including an individual’s beliefs, social interactions, and organizational and policy factors [36]. In addition, factors such as space and time are often critical aspects of health context [37].

These factors include health perceptions about the safety of e-cigarettes versus smoking, cost, sensory pleasure, effect on social relations (eg, not inconveniencing others), and popularity in social settings. We classified these by three main categories of interest: (1) subject matter (e-cigarette and hookah), (2) health (symptoms, quitting, health perceptions, and health care practitioners), and (3) context (social relationships, setting, time, cost and sensory experience). We employed lexicons containing words that represented these categories. By using these words to match against the online discussion content, we could come to understand to what degree the discussion content contained information about these categories of interest. The higher the proportion of this content, the more we might be interested in examining the content in that forum. Table 1 depicts the categories, their definitions, and example terms. The terms in the lexicon are provided in Multimedia Appendix 1.

The process of lexicon development was a hybrid one consisting of both a literature review and iterative testing involving examination of the discussion content. The Symptoms and Quitting terms primarily came from the empirical literature but were augmented using online consumer-generated content, such as guides written for novice users, discussion forums, and websites advertising e-cigarette and hookah products. The other dimensions were primarily drawn from user-generated content and supplemented using empirical research. Lexicon development was an iterative process of adding keywords until the addition of new keywords did not result in substantive differentiation across the datasets being compared.

Table 1. Contextual factors of tobacco use: Lexicon definitions and examples.

Contextual factors	Definition
Subject matter	
E-cigarette	The types and parts of e-cigarettes, eg, ecig, vape, “atty” (atomizer), “carto” (cartomizer).
Hookah	The types and parts of hookahs, eg, hookah, waterpipe, shisha, mouthpiece.
Health	
Symptoms	This set of concepts was constructed from existing literature on the health effects of e-cigarette and hookah use, particularly [26], and also through examination of the discussion content harvested in this study, eg, throat, cough, migraine, craving.
Quitting	Pertaining to experience of quitting, including motivations (eg, “stigma” and “stink”), difficulties in quitting (eg, “stress”), and tobacco cessation aids; also includes psychological factors such as “depression” and “anxiety”.
Perceptions	Perceptions of the safety of and potential health implications of e-cigarettes and hookah use, eg, toxic, dangerous, safe.
Health care practitioners	Various types of health care practitioners, eg, doctors, physicians, therapists, counselors.
Context	
Social relationships	Social relations that are often mentioned in discussion forums, eg, family, friends, children.
Setting	Settings where vaping and hookah use may occur, eg, home, bar, party.
Time	Timing of e-cigarette and hookah use, eg, morning, afternoon, evening.
Cost	Cost aspects of tobacco use, eg, cheap, expensive, price, saving.
Sensory experience	Sensory aspects of tobacco use, eg, hit, cloud, buzz.

We used these lexicons to estimate the prevalence of each category of interest, and then we rendered a heat map visualization to compare across forums. Heat maps are often used in genetics to display gene expression patterns [38,39] or to show the results of hierarchical clustering. In a classic cluster heat map, one axis of the heat map might represent samples, and the other, genes [40]. Each cell is colored based on the level of expression of the gene in the corresponding sample.

Topic Modeling and Visualization

In the second and third parts of our study, we used topic models to compare the content of online discussion forums. To model topics, we used a generative probabilistic modeling algorithm, Latent Dirichlet Allocation (LDA). LDA is a technique that models documents as random mixtures over topics, where a topic is characterized as a distribution of words [41].

We employed the LDA implementation that is available with the MALLET toolkit [42]. Previous research has observed that results with and without stemming yield comparable results and that stemmed results are more difficult to interpret [43]. In this study, we opted not to stem because viewing the original versions of the words facilitated interpretation of the context in which words were used. We used an augmented stop word list that included the original MALLET stop word list, as well as other common online forms of non-substantive words and word fragments, such as “ill” (“I’ll) and “dont” and forum members’ usernames. The augmented stop words, with the exception of forum members’ usernames, have been provided in Multimedia Appendix 2.

We trained topic models for four forums: Vapor Talk General E-Cig Discussion, Hookah Forum General Discussion, Vapor Talk Health & Safety, and the Stopsmoking subreddit. We experimented with different numbers of topics in order to find a level of granularity that showed the diversity of discussion topics, while at the same time avoiding topics that were thematically similar. We named all of the topics through a combination of examining keywords and manual examination of posts that were representative of those topics. To reduce complexity, we then grouped these topics together into categories if they were thematically similar. A list of all the topics and their respective categories, for each topic model, is available in Multimedia Appendix 3.

The output of topic modeling includes a set of topics and the main words associated with that topic, as well as a list of documents, with estimates of the proportion of each topic present in each document. Thus, from these outputs, one could say, for example, that if 60% of document A consists of topic X, then document A primarily consists of topic X, with trace amounts of all other topics. Similarly, a document B that is predicted to be 30% topic Y and 30% topic Z might be said to primarily consist of topics Y and Z, with trace amounts of all other topics. One final example would be that a document contains small amounts of all the topics but is not that representative of any topic in particular.

In order to summarize the prevalence of the topics generated, we used an estimate of main “document-topics”. By document topic, we refer to the instances where a topic is a major constituent of a given document. A topic was considered a major constituent of a document if it was predicted to constitute a given minimum proportion of that document. The thresholds were determined by iteratively testing different candidate values until the number of “document-topics” was close to the number of total posts in the discussion forum. The selection of this criterion was to maximize the proportion of content that was represented.

We calculated the number of document-topic elements for each topic and then divided by the number of total document-topic elements, to determine the proportion of a forum that was constituted by each topic. We then used these proportions to render a horizontal stacked bar chart, which supports a visual comparison of topic prevalence within and across discussion forums.

Research Ethics Statement

Publicly available social media content can be an invaluable complement to data provided by study participants in more explicit research contexts because it is a rich source of information on how behaviors with health impacts may naturally occur in the real world. In order to protect the identities of forum users, we have not provided explicit quotations, but instead described the content in as much detail as possible, both quantitatively and qualitatively, in line with ethical guidelines [44,45]. The work reported in this paper has been certified as exempt from review under 45 CFR 46.101(b), category 4 by the University of California San Diego Institutional Review board (Project #140844X).

Results

Harvesting of the Document Collection

We examined content from three different websites: (1) Vapor Talk, a website devoted to e-cigarettes, (2) Hookah Forum, a forum devoted to hookah use, and (3) Reddit, a site featuring discussion forums on a wide variety of topics. On Reddit, we chose to focus on three different discussion forums: “electronic_cigarette”, “hookah”, and “stopsmoking”. On Vapor Talk, we focused on two subforums: “General E-cig Discussion” and “Health & Safety.” On Hookah Forum, we focused on the general discussion forum only, as this website does not have a forum dedicated specifically to health topics. The forums differed considerably in terms of the number of total posts, the mean number of users, and mean post length (Table 2).

Table 2. Corpus statistics.

	Vapor Talk		Hookah Forum	Reddit		
	General	Health	General	Stop-smoking	Electronic cigarette	Hookah
Posts, n	11,438	2376	17,761	2092	89,119	43,501
Threads, n	690	172	413	177	2093	2994
Users, n	773	423	1659	760	14,277	4374
Post length, mean (SD)	356.35 (447.33)	487.39 (653.45)	323.16 (520.97)	267.49 (441.77)	189.29 (378.29)	155.88 (263.75)

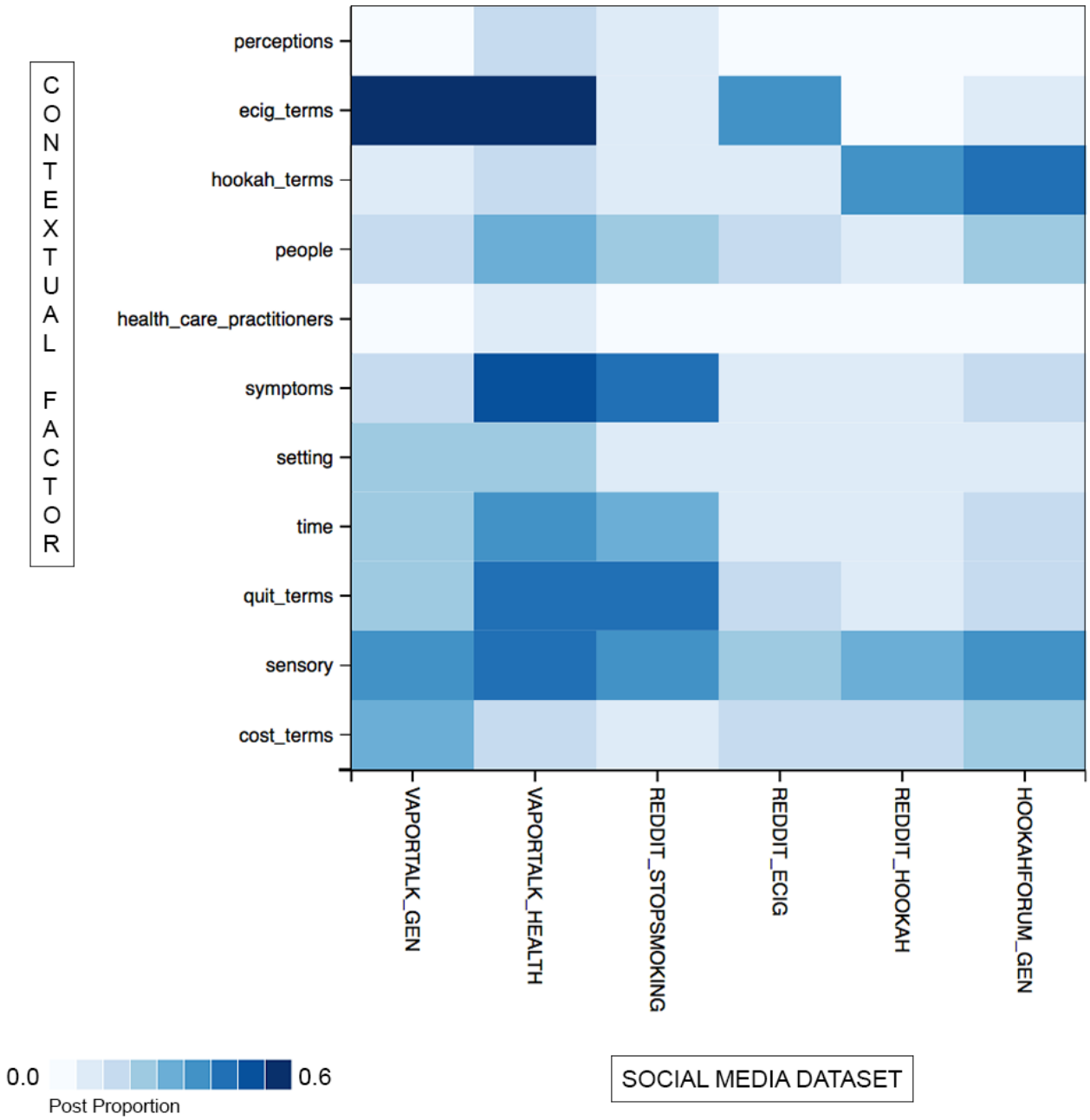
Comparing Contextual Factors of Tobacco Use Across Datasets

In our first research question, we asked what differences there were in the prevalence of contextual factors of e-cigarette and hookah use across different online communities. The prevalence of contextual factors was calculated as the proportion of posts containing a term from the relevant contextual factor lexicon, and a heat map was rendered based on these prevalence estimates (Figure 1). The darker the hue, the higher the proportion of that type of content in the forums, with the darkest cells representing approximately 60% of the forum content.

As we might expect, e-cigarette–related content was most popular in the Vapor Talk forums and on the Electronic_cigarette subreddit, and hookah content was most popular in the hookah forums. The two general forums on Vapor Talk and Hookah Forum contained more content on the cost and purchasing of equipment. Examination of the content showed an active discussion of the “ins and outs” of these products (ie, the detailed description of the intricacies of product use) and cost implications of product use. Descriptions of sensory experience appear common in most of the forums, which suggests that the sensory aspects of use are important across multiple types of tobacco products.

The purpose of the heat map visualization was to identify forums that contained the richest information about contextual factors in e-cigarette and hookah use. We saw that the mentions of people, symptoms, time, quitting, and sensory experience were highest in density in the Vapor Talk Health & Safety forum and in the Stopsmoking subreddit. Examining the discussion content, we saw that a substantial part of this discussion addressed people’s health situations as pertaining to e-cigarette use (in Vapor Talk Health & Safety) and to quitting without e-cigarettes (in the Stopsmoking subreddit).

Figure 1. Contextual factors of e-cigarette and hookah use.



Topic Modeling and Visualization

We trained topic models for four forums: Vapor Talk General E-Cig Discussion, Hookah Forum General Discussion, Vapor Talk Health & Safety, and the Stopsmoking subreddit. We experimented with different numbers of topics in order to find a level of granularity that showed the diversity of discussion topics, while at the same time avoiding topics that were thematically similar. Ultimately, we generated 20 topics for each of the subforums, with the exception of Hookah Forum. Hookah Forum had a greater number of posts than the other forums, as well as a shorter mean post length. With fewer numbers of topics, the themes were not as coherent; thus, we generated 40 topics for Hookah Forum.

We labeled all of these topics and set a minimum threshold for document topics as discussed in the Methods section. In the Stopsmoking subreddit, topics were dispersed in more trace amounts throughout the other posts; thus, it was necessary to lower the threshold to preserve a similar number of document-topics. Aggregate statistics for the four topic models are presented in Table 3.

Table 3. Topic modeling results overview.

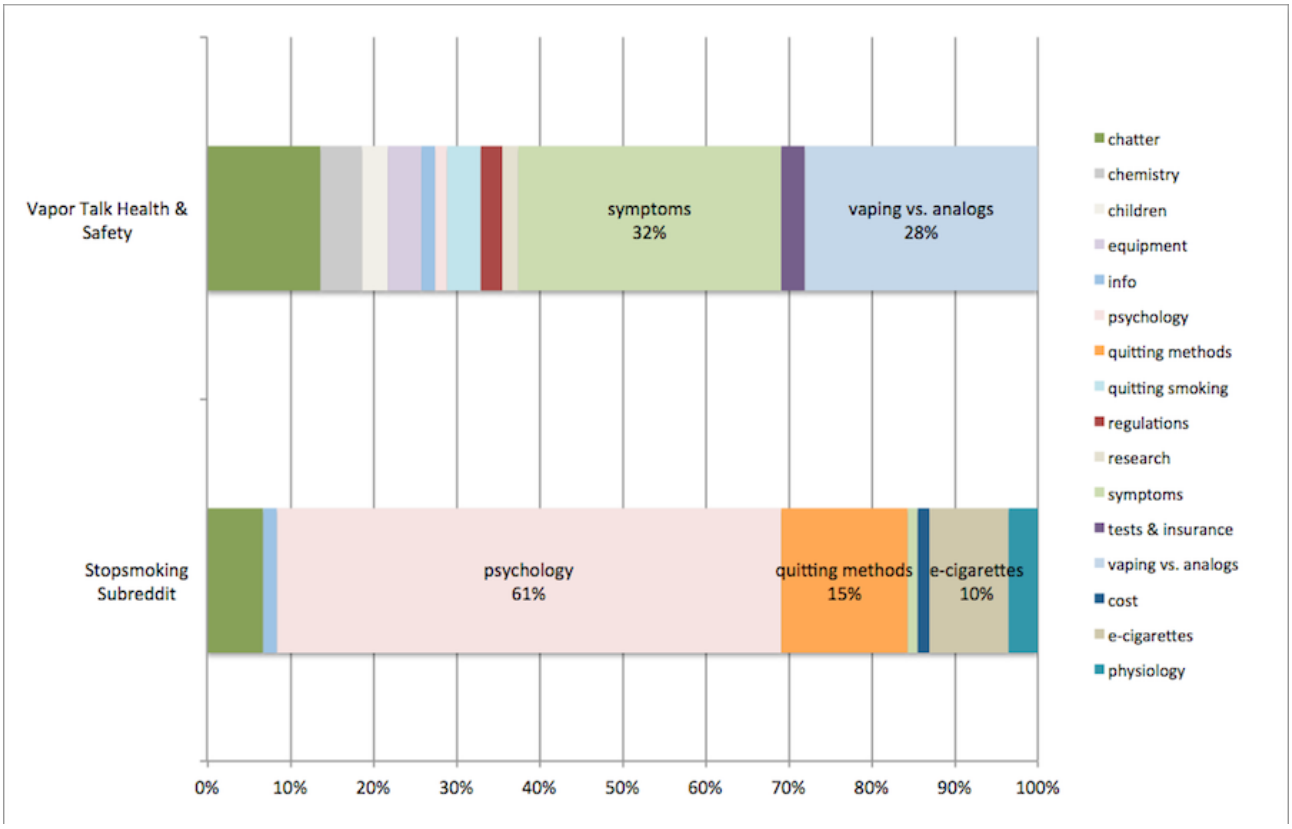
	Vapor Talk		Reddit	Hookah Forum
	Health	General	Stop-smoking	General
Total posts, n	2376	11,438	2092	17,761
Total topics, n	20	20	20	40
Document-topic threshold, n	0.3	0.3	0.2	0.3
Post length, mean (SD)	487.39 (653.45)	356.35 (447.33)	267.49 (441.77)	323.16 (520.97)

E-Cigarette Versus Combustible Cigarette Use

We used the topic modeling results to render a Topic Bars visualization to compare the two forums with the richest discussion of contextual factors: Vapor Talk Health & Safety, and the Stopsmoking subreddit (Figure 2). In Vapor Talk Health, the two most prominent categories were Symptoms and Vaping versus Analogs. With regard to Symptoms, common topics were the health dangers of smoking cigarettes, problems that forum members have encountered in the mouth and throat, the use of propylene glycol ("pg") as opposed to vegetable glycerin ("vg"), and sleep quality.

In the Stopsmoking subreddit, we saw a much different picture. The most salient bars were Psychology (60.60%, 1435/2368 document-topics) and Quitting Methods (15.29%, 362/2368). In Psychology, the topics discussed included overcoming cravings, dealing with friends, and encouragement that cravings would pass. The Quitting Methods category had only one constituent topic, Quitting Mechanisms, which included terms such as "cold turkey", "gum", and "patch". It is useful to observe that in the Stopsmoking subreddit (Figure 2, bottom), only 9.50% of the discussion content is focused on e-cigarettes (225/2368).

Figure 2. Topic Bars: Quitting in Vapor Talk Health & Safety versus the Stopsmoking Subreddit.



E-Cigarettes Versus Hookah

In the last part of our study, we considered the two products: e-cigarettes and hookah. Are these communities different, and if so, how? To consider this question, we compared the Topic Bars visualization for Vapor Talk General E-Cig Discussion and Hookah Forum General Discussion (Figure 3).

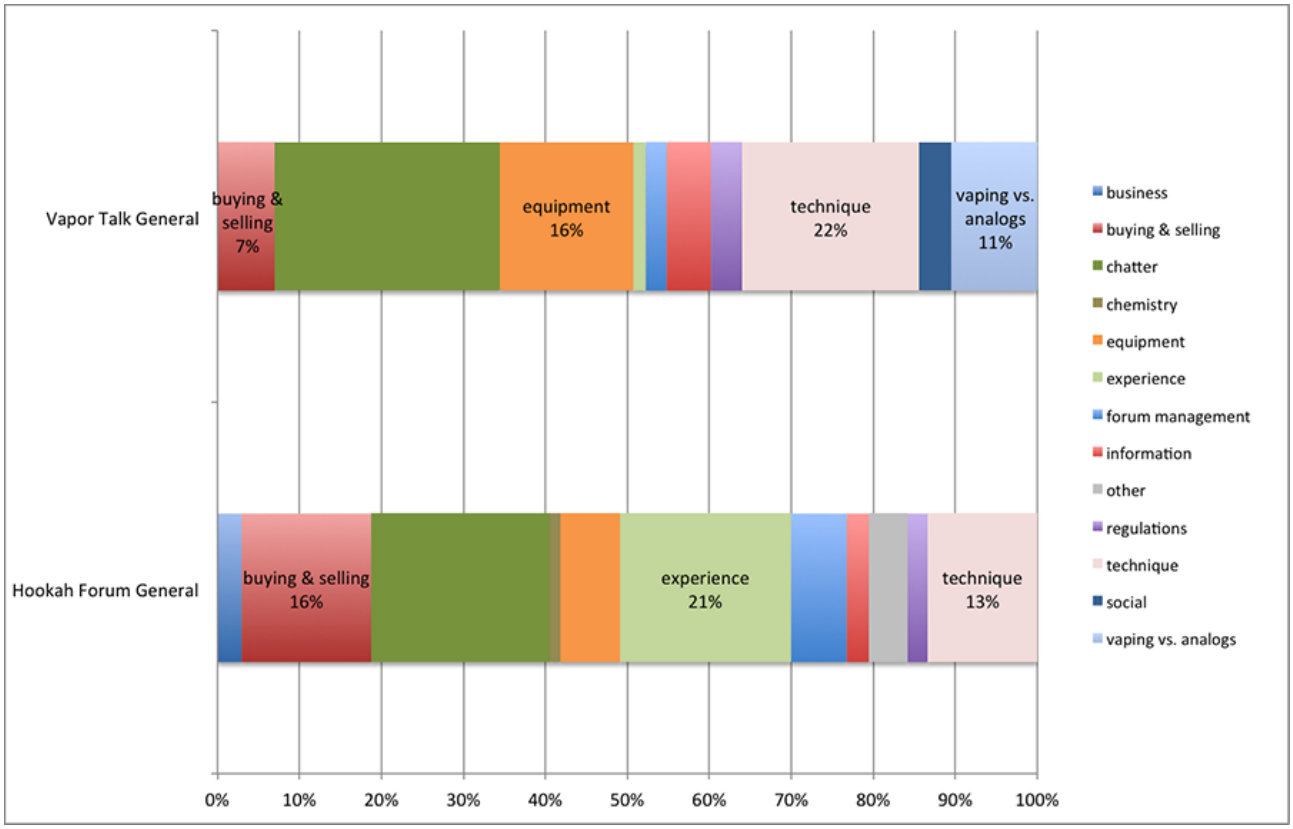
There are similarities between the categories of discussions on Vapor Talk and Hookah Forum. In both forums, there was a substantial amount of general chatter (dark green). In addition, both forums featured discussion on buying and selling equipment for e-cigarettes and hookah (red). From the dialogue content, the consumers in Vapor Talk appeared to primarily be end consumers, whereas the consumers in Hookah Forum consisted both of individuals interested in the purchase of hookah equipment for personal use, as well as proprietors of hookah lounges. There were also individuals in both forums whose member type indicated that they were a vendor.

There were also many topics relating to technique (pink). In Vapor Talk, topics concerning technique included how to get a good taste and how different characteristics of the juices affect the vaping experience. In Hookah Forum, sample topics included how to pack the bowl and whether it is a good idea to put other things (eg, alcohol) in the base. Thus, e-cigarette and hookah forums are similar in that their members are actively engaged in information exchange concerning technical and cost-related aspects of the use of their products of choice.

The most prominent difference between Hookah Forum and Vapor Talk is the greater focus on equipment in Vapor Talk (orange), as opposed to the focus on the use experience in Hookah Forum (light green). In Hookah Forum, there is a great deal of discussion of different flavors, "buzz", and clouds. A large proportion of Vapor Talk is devoted to equipment, that is, discussion of the different types and parts of e-cigarettes, including mods, tanks, coils, atomizers, cartomizers, and batteries.

There is some discussion in these two forums about health—a substantive part of the conversation in Vapor Talk focuses on vaping as opposed to smoking "analogs" (traditional cigarettes), and though not as prominent in the discussion content, a number of health concerns were also expressed in Hookah Forum, relating to headaches, lung issues, and vocal chord problems. There was also discussion on ways to prevent getting sick from smoking hookah, including eating prior to smoking and staying properly hydrated—though one might consider this not a matter of health concern, but rather, a practical consideration in order to enjoy the experience.

Figure 3. Topic Bars: Vapor Talk General E-Cig versus Hookah Forum General Discussion.



Discussion

Principal Findings

In this paper, we used text mining and visualization techniques to examine the use of different tobacco products. At the outset, we identified contextual factors of these behaviors, particularly in terms of health impacts and concerns. Then we generated a heat map that enabled us to compare forum content in terms of these factors of interest. Based on this information, we selected two forums that contained the highest densities of these factors and rendered a topic modeling-based visualization, Topic Bars, to compare these forums. This comparison enabled us to gain insights concerning the experience of tobacco use with e-cigarettes and the experience of tobacco use without e-cigarettes. Last, we constructed another Topic Bars visualization to compare general e-cigarette and hookah discussion, to investigate similarities and differences between the communities.

The main contributions of this paper are as follows. First, we have demonstrated an approach using text mining and visualization techniques to select particular social media datasets out of a larger pool, for a particular health behavior. The crux of this technique is to identify factors of interest for developing strategies to facilitate behavioral change and then employ relevant lexicons to assess and compare the amount of content concerning these factors, across datasets. This technique can be helpful for characterizing discussion forums as a whole, as well as in the selection and differentiation of social media datasets to investigate specific research questions.

Second, this paper shows that e-cigarettes provide a very different experience of tobacco use as compared to combustible cigarettes. When smokers who are trying to quit visit a discussion forum, they report on the difficulties they are having trying to quit, and others in the forum chime in to offer their encouragement. The psychological element is extremely salient, and the focus is on quitting. In the case of e-cigarettes, we saw that much of the discussion focused on symptoms that people were experiencing as they were using e-cigarettes. People using e-cigarettes appear less likely to engage in the psychological battle of quitting. The e-cigarette has diverted their attention to a different activity, dealing with concrete problems to avert particular physiological symptoms associated with e-cigarette use. Moreover, at least for some Vapor Talk users, their goal is to be analog free rather than nicotine free, and hence a psychological struggle is less evident.

The difference in psychological state and engagement of the consumer is an important concern on two levels. In terms of regulation and policy concerning electronic cigarettes, there are no clear answers, but the findings of this study highlight the importance of considering impacts on psychological state and engagement in the regulation of electronic cigarettes as opposed to combustible cigarettes. On an individual level, users of tobacco products interact with electronic cigarettes in very different ways than they do combustible cigarettes, and thus, the pathway that one faces in quitting the use of all tobacco products appears to be fundamentally different. Counselors and those designing educational programs designed for smokers should be aware of the differences so that they can provide different types of support to facilitate changes in health behavior.

Last, this study examined the general content in discussion forums for e-cigarette and hookah. There are strong similarities, and ultimately, both are focused on improving the use experience, which has a strong sensory component. These are “hobbyist cultures” in that their members are enthusiastic users and sharers of information concerning their common activity. Particularly given the rapid rate at which the two products are growing in popularity, online communities, as common sites of information diffusion and as sources of the latest information, are ideal environments to study both.

Limitations

This work has various limitations. First, we harvested data from three websites, and there are certainly many other online communities relating to tobacco products. We deliberately selected different types of communities and subsetted the communities in order to examine similarities and differences within and between communities. As we expected, the selected communities vary in many characteristics, suggesting that they represent a range of tobacco users’ experiences. However, this investigation focused on a subset of online communities that are available to users of tobacco products, and it would be valuable to examine additional communities in the future, for example, by comparing multiple forums for e-cigarettes and/or multiple forums for hookah in order to characterize the variability in topics addressed in online communities for the same product type. Additional research might also consider the content in relation to the demographic characteristics of the users, which was out of the scope of the current study.

Second, the users in an online community are not necessarily representative of users of tobacco products, cigarettes, e-cigarettes, and/or hookah in general. While we agree that this is true, today, if a typical user goes to a search engine and types “e-cigarette sore throat”, among the first entries to come up would be links to specific threads on this subject in discussion forums including Vapor Talk. Thus, the potential for exposure to a much larger number of people, those who do not actively participate in discussion forums, is a reality.

Third, in this study we constructed lexicons to assess contextual factors of interest for a particular type of behavior with health impacts. The lexicon is not necessarily generalizable to other types of health behaviors, nor would it necessarily perform comparably over time. It is likely that as language evolves, the lexicon would need to be augmented. However, there is potential here to extend the lexicon for application to other health contexts and time periods.

Atmosphere of the Forums and Implications

In this paper, we employed two primary techniques, a contextual heatmap, and a Topic Bars visualization, in order to explore differences between data sets. The Topic Bars visualization enabled us to specifically compare different discussion forums. We now consider some of the differences in topics between forums and what this may mean.

The results of the topic models on Vapor Talk Health and Stopsmoking subreddits suggest that those who attempt to quit smoking combustible cigarettes and those who use e-cigarettes have very different experiences. It appears that many who use e-cigarettes encounter problems that may lead them to do research and perhaps find a solution; thus, the forums contain detailed accounts of the technical intricacies of vaping and the health issues that may be encountered. Though a minority of the members of the Stopsmoking subreddit appear to use e-cigarettes, for the most part this group appears to take more traditional approaches to quitting, with emphasis on mutual encouragement and support, and coping with the psychological aspects of this experience. These topic modeling results suggest that, without e-cigarettes, the aspect of quitting that is most salient is the psychological hurdle, though it is important to state that users may be using e-cigarettes but not reporting this activity in their Stopsmoking subreddit discussion.

The information exchanged and atmosphere of support in these two forums appears to be quite different. Whereas Vapor Talk includes detailed reports of symptoms and their temporal context (eg, how long the symptoms have lasted and when they started), the Stopsmoking subreddit appears to be focused on mutual encouragement, reinforcement of the value of quitting, and strategies for overcoming cravings. Time is important here also, but the nature of that time is different. Many forum participants report how long it has been since they quit, and others add words of encouragement and how long it has been since they quit. Thus, there are many shorter posts here.

The interactions in the two forums have both similarities and differences to existing literature on online support groups for

smoking cessation. Previous studies of discussion forums for quitting smoking have found that most participants were women, and that they used the forum mostly as a source of emotional support and encouragement, and less often for the purposes of eliciting practical information and quitting tips [46,47]. Consistent with this work, there appeared to be a substantial amount of support and encouragement. However, in contrast to prior work, there did appear to be information and quitting tips exchanged. In Vapor Talk Health & Safety, the tips often took the form of concrete advice about the types of e-liquids to use, how to inhale, and so on, which could potentially alleviate problems with the mouth and throat. In the Stopsmoking subreddit, the tips were often psychological, concerning how to overcome the desire to smoke.

In our topic modeling results comparing e-cigarette and hookah discussion (Vapor Talk and Hookah Forum), initially there appear to be substantive differences in the content. However, there are similarities in the nature of the communities. In the case of hookah, the use experience is prominent, including discussion of the "buzz", smoke rings, and clouds. In the case of e-cigarettes, the equipment and techniques features more prominently, but much of that discussion is on how to get a "throat hit" or a better taste. Thus, improving the experience is a common theme in both forums.

In summary, the results of these two topic models suggest similarities in the e-cigarette and hookah general discussion. Both are communities composed of enthusiastic users of a product who are actively engaged in the discovery and sharing of new information on how to obtain or enjoy the products that they champion. As such, this content can be invaluable in terms of providing knowledge of the day-to-day use problems that may occur with the two products.

Acknowledgments

This work was supported by a grant from the National Cancer Institute U01 CA154280. The authors would also like to thank Jessica Sun for providing useful reference material for the development of the Factors for Tobacco Use lexicon and the reviewers for their helpful suggestions.

Authors' Contributions

ATC and MC collected the data, and all authors participated in the conceptualization of the study. ATC performed the analyses and MC reviewed the results, including the topic labeling and the topic-category mappings. AC drafted the paper, ATC and MC revised the paper, and SHZ provided feedback.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Factors of tobacco use lexicon.

[XLSX File (Microsoft Excel File), 40KB - jmir_v17i9e220_app1.xlsx]

Multimedia Appendix 2

Additional stop words.

[PDF File (Adobe PDF File), 13KB - jmir_v17i9e220_app2.pdf]

Multimedia Appendix 3

Topic modeling results.

[XLSX File (Microsoft Excel File), 62KB - jmir_v17i9e220_app3.xlsx]

References

1. Corley CD, Cook DJ, Mikler AR, Singh KP. Text and structural data mining of influenza mentions in Web and social media. Int J Environ Res Public Health 2010 Feb;7(2):596-615 [FREE Full text] [doi: 10.3390/ijerph7020596] [Medline: 20616993]
2. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: twitter and antibiotics. Am J Infect Control 2010 Apr;38(3):182-188 [FREE Full text] [doi: 10.1016/j.ajic.2009.11.004] [Medline: 20347636]
3. Salathé M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. PLoS Comput Biol 2011 Oct;7(10):e1002199 [FREE Full text] [doi: 10.1371/journal.pcbi.1002199] [Medline: 22022249]

4. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
5. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection--harnessing the Web for public health surveillance. N Engl J Med 2009 May 21;360(21):2153-5, 2157 [FREE Full text] [doi: 10.1056/NEJMp0900702] [Medline: 19423867]
6. Salathé M, Bengtsson L, Bodnar TJ, Brewer DD, Brownstein JS, Buckee C, et al. Digital epidemiology. PLoS Comput Biol 2012;8(7):e1002616 [FREE Full text] [doi: 10.1371/journal.pcbi.1002616] [Medline: 22844241]
7. 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;15(8):e174 [FREE Full text] [doi: 10.2196/jmir.2534] [Medline: 23989137]
8. Chen AT. Exploring online support spaces: using cluster analysis to examine breast cancer, diabetes and fibromyalgia support groups. Patient Educ Couns 2012 May;87(2):250-257. [doi: 10.1016/j.pec.2011.08.017] [Medline: 21930359]
9. Cobb NK, Graham AL, Byron MJ, Niaura RS, Abrams DB, Workshop P. Online social networks and smoking cessation: a scientific research agenda. J Med Internet Res 2011;13(4):e119 [FREE Full text] [doi: 10.2196/jmir.1911] [Medline: 22182518]
10. Regan AK, Promoff G, Dube SR, Arrazola R. Electronic nicotine delivery systems: adult use and awareness of the 'e-cigarette' in the USA. Tob Control 2013 Jan;22(1):19-23. [doi: 10.1136/tobaccocontrol-2011-050044] [Medline: 22034071]
11. Vardavas CI, Filippidis FT, Agaku IT. Determinants and prevalence of e-cigarette use throughout the European Union: a secondary analysis of 26 566 youth and adults from 27 Countries. Tob Control 2015 Sep;24(5):442-448. [doi: 10.1136/tobaccocontrol-2013-051394] [Medline: 24935441]
12. Zhu S, Gamst A, Lee M, Cummins S, Yin L, Zoref L. The use and perception of electronic cigarettes and snus among the U.S. population. PLoS One 2013;8(10):e79332 [FREE Full text] [doi: 10.1371/journal.pone.0079332] [Medline: 24250756]
13. Pearson JL, Richardson A, Niaura RS, Vallone DM, Abrams DB. e-Cigarette awareness, use, and harm perceptions in US adults. Am J Public Health 2012 Sep;102(9):1758-1766 [FREE Full text] [doi: 10.2105/AJPH.2011.300526] [Medline: 22813087]
14. Etter J. Electronic cigarettes: a survey of users. BMC Public Health 2010;10:231 [FREE Full text] [doi: 10.1186/1471-2458-10-231] [Medline: 20441579]
15. Etter J, Bullen C. Electronic cigarette: users profile, utilization, satisfaction and perceived efficacy. Addiction 2011 Nov;106(11):2017-2028. [doi: 10.1111/j.1360-0443.2011.03505.x] [Medline: 21592253]
16. Maziak W, Ward KD, Afifi SRA, Eissenberg T. Tobacco smoking using a waterpipe: a re-emerging strain in a global epidemic. Tob Control 2004 Dec;13(4):327-333 [FREE Full text] [doi: 10.1136/tc.2004.008169] [Medline: 15564614]
17. Martinasek MP, McDermott RJ, Martini L. Waterpipe (hookah) tobacco smoking among youth. Curr Probl Pediatr Adolesc Health Care 2011 Feb;41(2):34-57. [doi: 10.1016/j.cppeds.2010.10.001] [Medline: 21232693]
18. Soule EK, Barnett TE, Curbow BA. Keeping the night going: the role of hookah bars in evening drinking behaviours. Public Health 2012 Dec;126(12):1078-1081. [doi: 10.1016/j.puhe.2012.06.010] [Medline: 22819668]
19. Blank MD, Brown KW, Goodman RJ, Eissenberg T. An observational study of group waterpipe use in a natural environment. Nicotine Tob Res 2014 Jan;16(1):93-99 [FREE Full text] [doi: 10.1093/ntr/ntt120] [Medline: 23943842]
20. Nakkash RT, Khalil J, Afifi RA. The rise in narghile (shisha, hookah) waterpipe tobacco smoking: a qualitative study of perceptions of smokers and non smokers. BMC Public Health 2011;11:315 [FREE Full text] [doi: 10.1186/1471-2458-11-315] [Medline: 21569577]
21. Sutfin EL, Song EY, Reboussin BA, Wolfson M. What are young adults smoking in their hookahs? A latent class analysis of substances smoked. Addict Behav 2014 Jul;39(7):1191-1196 [FREE Full text] [doi: 10.1016/j.addbeh.2014.03.020] [Medline: 24746345]
22. Heinz AJ, Giedgowd GE, Crane NA, Veilleux JC, Conrad M, Braun AR, et al. A comprehensive examination of hookah smoking in college students: use patterns and contexts, social norms and attitudes, harm perception, psychological correlates and co-occurring substance use. Addict Behav 2013 Nov;38(11):2751-2760. [doi: 10.1016/j.addbeh.2013.07.009] [Medline: 23934006]
23. Chan WC, Leatherdale ST, Burkhalter R, Ahmed R. Bidi and hookah use among Canadian youth: an examination of data from the 2006 Canadian Youth Smoking Survey. J Adolesc Health 2011 Jul;49(1):102-104. [doi: 10.1016/j.jadohealth.2010.11.250] [Medline: 21700168]
24. Blachman-Braun R, Del Mazo-Rodríguez RL, López-Sámano G, Buendía-Roldán I. Hookah, is it really harmless? Respir Med 2014 May;108(5):661-667. [doi: 10.1016/j.rmed.2014.01.013] [Medline: 24582881]
25. Aljarrah K, Ababneh ZQ, Al-Delaimy WK. Perceptions of hookah smoking harmfulness: predictors and characteristics among current hookah users. Tob Induc Dis 2009;5(1):16 [FREE Full text] [doi: 10.1186/1617-9625-5-16] [Medline: 20021672]
26. Hua M, Alfi M, Talbot P. Health-related effects reported by electronic cigarette users in online forums. J Med Internet Res 2013;15(4):e59 [FREE Full text] [doi: 10.2196/jmir.2324] [Medline: 23567935]
27. Huang J, Kornfield R, Szczypka G, Emery SL. A cross-sectional examination of marketing of electronic cigarettes on Twitter. Tob Control 2014 Jul;23 Suppl 3:iii26-iii30 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051551] [Medline: 24935894]

28. Brockman LN, Pumper MA, Christakis DA, Moreno MA. Hookah's new popularity among US college students: a pilot study of the characteristics of hookah smokers and their Facebook displays. BMJ Open 2012;2(6):e001709 [FREE Full text] [doi: 10.1136/bmjopen-2012-001709] [Medline: 23242241]
29. Carroll MV, Shensa A, Primack BA. A comparison of cigarette- and hookah-related videos on YouTube. Tob Control 2013 Sep;22(5):319-323 [FREE Full text] [doi: 10.1136/tobaccocontrol-2011-050253] [Medline: 22363069]
30. Murphy J, Link MW, Childs JH, Tesfaye CL, Dean E, Stern M, et al. Social Media in Public Opinion Research: Report of the AAPOR Task Force on Emerging Technologies in Public Opinion Research. 2014. URL: https://www.aapor.org/AAPORKentico/AAPOR_Main/media/MainSiteFiles/AAPOR_Social_Media_Report_FNL.pdf [accessed 2015-07-02] [WebCite Cache ID 6ZkvWErIN]
31. Vapor Talk. URL: https://www.vaportalk.com/ [accessed 2015-09-15] [WebCite Cache ID 6bZTfYa6r]
32. Hookah Forum. URL: http://www.hookahforum.com/ [accessed 2015-04-05] [WebCite Cache ID 6XZou3N0m]
33. Reddit. URL: http://www.reddit.com [accessed 2015-04-06] [WebCite Cache ID 6XZohdFyL]
34. Bogers T, Wernersen RN. How 'social' are social news sites? Exploring the motivations for using Reddit. 2014 Presented at: iConference 2014; March 4-7, 2014; Berlin, Germany p. 329-344 URL: https://www.ideals.illinois.edu/bitstream/handle/2142/47295/108_ready.pdf?sequence=5 [doi: 10.9776/14108]
35. 6% of Online Adults are Reddit Users. Washington, DC: Pew Research Center; 2014. URL: http://www.pewinternet.org/files/old-media/Files/Reports/2013/PIP_reddit_usage_2013.pdf [accessed 2015-07-02] [WebCite Cache ID 6ZkrUJnFS]
36. McLeroy KR, Bibeau D, Steckler A, Glanz K. An ecological perspective on health promotion programs. Health Educ Q 1988;15(4):351-377. [Medline: 3068205]
37. Rainham D, McDowell I, Krewski D, Sawada M. Conceptualizing the healthscape: contributions of time geography, location technologies and spatial ecology to place and health research. Soc Sci Med 2010 Mar;70(5):668-676. [doi: 10.1016/j.socscimed.2009.10.035] [Medline: 19963310]
38. Barkow S, Bleuler S, Prelic A, Zimmermann P, Zitzler E. BicAT: a biclustering analysis toolbox. Bioinformatics 2006 May 15;22(10):1282-1283 [FREE Full text] [doi: 10.1093/bioinformatics/btl099] [Medline: 16551664]
39. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One 2010;5(11):e13984 [FREE Full text] [doi: 10.1371/journal.pone.0013984] [Medline: 21085593]
40. Wilkinson L, Friendly M. The History of the Cluster Heat Map. The American Statistician 2009 May;63(2):179-184. [doi: 10.1198/tas.2009.0033]
41. Blei D, Ng A, Jordan M. Latent Dirichlet Allocation. J Mach Learn Res 2003;3:993-1022.
42. McCallum AK. Machine Learning for Language Toolkit. 2002. URL: http://mallet.cs.umass.edu/ [accessed 2015-04-05] [WebCite Cache ID 6XZgiQKil]
43. Yang TI, Torget AJ, Mihalcea R. Topic Modeling on Historical Newspapers. USA: Association for Computational Linguistics; 2011 Presented at: 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities; June 24, 2011; Portland, Oregon p. 96-104 URL: http://www.aclweb.org/anthology/W11-1513
44. Eysenbach G, Till JE. Ethical issues in qualitative research on internet communities. BMJ 2001 Nov 10;323(7321):1103-1105 [FREE Full text] [Medline: 11701577]
45. McKee R. Ethical issues in using social media for health and health care research. Health Policy 2013 May;110(2-3):298-301. [doi: 10.1016/j.healthpol.2013.02.006] [Medline: 23477806]
46. Burri M, Baujard V, Etter J. A qualitative analysis of an internet discussion forum for recent ex-smokers. Nicotine Tob Res 2006 Dec;8 Suppl 1:S13-S19. [Medline: 17491166]
47. Cobb NK, Graham AL, Abrams DB. Social network structure of a large online community for smoking cessation. Am J Public Health 2010 Jul;100(7):1282-1289 [FREE Full text] [doi: 10.2105/AJPH.2009.165449] [Medline: 20466971]

Abbreviations

LDA: Latent Dirichlet Allocation

Edited by G Eysenbach; submitted 13.04.15; peer-reviewed by J Hawkins, J Pugatch; comments to author 24.06.15; accepted 25.07.15; published 29.09.15.

Please cite as:
Chen AT, Zhu SH, Conway M
What Online Communities Can Tell Us About Electronic Cigarettes and Hookah Use: A Study Using Text Mining and Visualization Techniques
J Med Internet Res 2015;17(9):e220
URL: http://www.jmir.org/2015/9/e220/
doi:10.2196/jmir.4517
PMID:26420469

Original Paper

How Consumers and Physicians View New Medical Technology: Comparative Survey

Debra L Boeldt[1], PhD; Nathan E Wineinger[1], PhD; Jill Waalen[1], PhD; Shreya Gollamudi[1]; Adam Grossberg[2]; Steven R Steinhubl[1], MD; Anna McCollister-Slipp[1]; Marc A Rogers[2], PhD; Carey Silvers[2]; Eric J Topol[1], MD

[1]Scripps Translational Science Institute, Scripps Health, The Scripps Research Institute, La Jolla, CA, United States

[2]WebMD, New York, NY, United States

Corresponding Author:
Eric J Topol, MD
Scripps Translational Science Institute
Scripps Health
The Scripps Research Institute
3344 North Torrey Pines Ct
Suite 300
La Jolla, CA, 92037
United States
Phone: 1 858 554 5708
Fax: 1 858 546 9272
Email: etopol@scripps.edu

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2015/12/e284

Abstract

Background: As a result of the digital revolution coming to medicine, a number of new tools are becoming available and are starting to be introduced in clinical practice.

Objective: We aim to assess health care professional and consumer attitudes toward new medical technology including smartphones, genetic testing, privacy, and patient-accessible electronic health records.

Methods: We performed a survey with 1406 health care providers and 1102 consumer responders.

Results: Consumers who completed the survey were more likely to prefer new technologies for a medical diagnosis (437/1102, 39.66%) compared with providers (194/1406, 13.80%; *P*<.001), with more providers (393/1406, 27.95%) than consumers (175/1102, 15.88%) reporting feeling uneasy about using technology for a diagnosis. Both providers and consumers supported genetic testing for various purposes, with providers (1234/1406, 87.77%) being significantly more likely than consumers (806/1102, 73.14%) to support genetic testing when planning to have a baby (*P*<.001). Similarly, 91.68% (1289/1406) of providers and 81.22% (895/1102) of consumers supported diagnosing problems in a fetus (*P*<.001). Among providers, 90.33% (1270/1406) were concerned that patients would experience anxiety after accessing health records, and 81.95% (1149/1406) felt it would lead to requests for unnecessary medical evaluations, but only 34.30% (378/1102; *P*<.001) and 24.59% (271/1102; *P*<.001) of consumers expressed the same concerns, respectively. Physicians (137/827, 16.6%) reported less concern about the use of technology for diagnosis compared to medical students (21/235, 8.9%; *P*=.03) and also more frequently felt that patients owned their medical record (323/827, 39.1%; and 30/235, 12.8%, respectively; *P*<.001).

Conclusions: Consumers and health professionals differ significantly and broadly in their views of emerging medical technology, with more enthusiasm and support expressed by consumers.

(J Med Internet Res 2015;17(9):e215) doi:10.2196/jmir.4456

KEYWORDS

digital revolution; healthcare; medical technology; physician and consumer attitudes; electronic health record; mobile health

Introduction

As a result of the digital revolution coming to medicine, new tools are becoming available and are starting to be introduced in clinical practice, including genome sequencing and commercially available medical technologies, such as mobile phone-enabled self-monitoring of physiologic metrics or replacements of traditional laboratory tests. Many of these new digital tools raise questions about their impact on the patient-physician relationship, ethical standards, privacy, and security [1,2]. Yet there is limited widespread knowledge about the perceptions and support by consumers and health care professionals of these technologies. Assessment until now has been limited with respect to both scope and inclusion of views for both health care professionals and consumers. Understanding both patient and provider attitudes is essential if such technology is to be implemented in the future. Accordingly, we conducted a large-scale survey of the perceptions and comfort level towards new technologies by patients and providers by directing the same survey to both groups, adapted for each audience.

Methods

Study Participants and Data Collection

The technology survey assessed perspectives in two separate population samples, classified as either providers or consumers. A total of 21,812 health care professional members of Medscape who were active in the past year were invited via email to complete an online 15-item survey. Respondents completed the survey between August 22 and September 8, 2014. Lay WebMD website visitors from August 18-27, 2014, were invited via an interstitial invitation to complete a nearly identical online survey. This invitation was extended to total 456,243 consumers.

The Scripps Health Institutional Review Board reviewed and deemed this study exempt.

Survey

Health care providers and consumers completed very similar 15-item surveys assessing attitudes toward new technology in medicine (see Multimedia Appendix 1). Participants also provided age and gender information. Consumers answered additional demographic questions, while providers reported their area of expertise and current work setting. To participate in the study, providers were categorized to one of the following occupations: doctor, nurse practitioner, physician assistant, nurse, or medical student. If the provider did not meet these criteria, their participation in the survey was terminated (n=144).

New Technology

Health care providers and consumers responded to questions about the use of new medical technology for self-diagnosis of non–life-threatening conditions (see Multimedia Appendix 2, Q4). Respondents rated their willingness to use technology on a 3-point scale and rated whether they would use genetic testing for eight different medical scenarios. The use of mobile phones for conducting blood tests or submitting health information to a health care provider for four different conditions (eg, suspicious skin problem) was also evaluated.

Privacy

Health care providers and consumers responded to one question assessing privacy and security concerns. The participants rated their level of reluctance to use digital technology due to concerns about privacy. More specifically, the question assessed levels of concern about storage, access, and sharing health records online, in addition to communicating electronically with health care providers.

Medical Health Records

Four questions addressed attitudes towards electronic medical records. Providers and consumers rated whether patients should have access to lab results and doctor notes/procedures, or if doctors should share only information they deem appropriate. Moreover, consumers and health care providers identified ownership of a medical record, and whether access to medical records would cause patient anxiety, management of health, or unnecessary medical evaluations. Attitudes towards the immediacy of accessing lab test results were also assessed.

Cost and Transparency

Three questions gauged likelihood to ask about medical costs prior to a procedure, patient rights to receive medical cost information prior to treatment, and access to prices charged by other providers. Providers were also asked whether they were willing to compete on the basis of price.

Physical Exams and Imaging

Attitudes towards annual physical exams were evaluated with one question. Health care providers and consumers reported whether they felt an annual exam is necessary or whether there is interest in alternative forms of monitoring health. Additional concerns about exposure to radiation (eg, x-rays, mammograms, angiograms) were rated on a 7-point Likert scale in one question.

Data Analysis

Age and gender differences between groups were assessed using chi-square statistics. Probit regression was conducted on survey items with categorical outcomes. Multinomial probit regression was used on 5 items to assess differences among multiple categorical (polytomous) outcomes. Linear regression was used for one item with a continuous outcome. For each survey item, all statistical analyses involving between-group comparisons were conducted accounting for age (continuous) and gender as covariates. Other covariates, though likely different between groups (eg, education, income), were unavailable. Significance results are presented without correction for multiple testing. All data analysis was conducted in R.

Results

A total of 2508 surveys were completed, representing 1406 health care providers and 1102 consumers. Of the total number of Medscape members emailed (21,812), 6.4% of providers responded. A total of 456,243 consumers visiting the webpages owned and operated by WebMD were invited to participate in the survey. Health care provider respondents were younger than consumer respondents (mean age 45 versus 60 years respectively, $P<.001$) and included fewer females (providers:

704/1406, 50.07%; consumers: 776/1102, 70.42%; $P<.001$) (Table 1). Thus, all between-group comparisons are also presented accounting for age and gender covariates.

Table 1. Characteristics of providers and consumers.

Characteristics	Provider (N=1406), n (%)	Consumer (N=1102), n (%)	*P* value
Sex (% female)	704 (50.01)	776 (70.42)	<.001[a]
Age in years			<.001[a]
20-29	228 (16.21)	40 (3.63)	
30-39	323 (22.97)	41 (3.72)	
40-49	311 (22.12)	119 (10.80)	
50-59	332 (23.61)	296 (26.86)	
60-69	175 (12.45)	373 (33.85)	
70+	37 (3.63)	233 (21.14)	
Politics			
Fiscally conservative, but socially liberal		159 (14.43)	
Fiscally conservative, socially conservative		253 (22.96)	
Fiscally liberal, socially conservative		24 (2.18)	
Fiscally liberal, socially liberal		118 (10.71)	
Middle of the road fiscally and socially		341 (30.94)	
None of the above		207 (18.78)	
Education			
Some high school		24 (2.18)	
High school graduate		172 (15.61)	
Some college		326 (29.58)	
College (2 year)		115 (10.44)	
College (4 year)		205 (18.60)	
Postgraduate work		260 (23.59)	
Marital status			
Married		616 (55.90)	
Domestic partner		51 (4.62)	
Never married		134 (12.16)	
Divorced/separated		203 (18.42)	
Widow		98 (8.89)	
Income (USD)			
Under $16,000		106 (9.62)	
$16,000-29,999		129 (11.71)	
$30,000-44,999		173 (15.70)	
$45,000-64,999		160 (14.52)	
$65,000-79,999		119 (10.80)	
$80,000-99,999		82 (7.44)	
>$100,000		159 (14.43)	
Declined to answer		174 (15.79)	
Ethnicity			
African American/black		82 (7.44)	
Caucasian/white		835 (75.77)	
Hispanic (any)		52 (4.72)	

Characteristics	Provider (N=1406), n (%)	Consumer (N=1102), n (%)	*P* value
Other		49 (4.45)	
Declined to answer		84 (7.62)	

[a]Chi-square test.

Consumers were primarily college-educated with nearly one-quarter (260/1102) having some post-graduate training, 60.53% (667/1102) were married or had a domestic partner, and 75.77% (835/1102) were Caucasian.

The majority of health care providers were doctors (827/1406, 58.82%) with nurses representing the smallest group (85/1406, 6.05%) (Table 2). The most common physician specialties were family medicine (280/1406, 19.91%), internal medicine (224/1406, 15.93%), and pediatrics (168/1406, 11.95%).

Table 2. List of provider occupations, settings, and specialty (N=1406).

Provider characteristics	n (%)
Occupation	
Doctor	827 (58.82)
Medical student	235 (16.71)
Nurse practitioner	152 (10.81)
Physician assistant	107 (7.61)
Nurse	85 (6.05)
Primary practice setting	
Hospital	326 (23.19)
Solo/group practice	311 (22.12)
Outpatient clinic	190 (13.51)
Academic, research, military, government	157 (11.17)
Group practice owned by hospital	149 (10.60)
Health care organization	131 (9.32)
Specialty	
Family medicine	280 (19.91)
Internal medicine	224 (15.93)
Other specialty	231 (16.43)
Pediatrics	168 (11.95)
Other	153 (10.88)
Psychiatry	67 (4.77)
OB/GYN & women's health	61 (4.34)
General surgery	57 (4.05)
Cardiology	47 (3.34)
Neurology	47 (3.34)
Emergency medicine	43 (3.06)
Hematology/Oncology	28 (1.99)

Technology

New Technology

Consumers were more likely to prefer using technology for self-diagnosis of non–life-threatening medical conditions (437/1102, 39.66%) compared with providers (194/1406, 13.80%), with more providers (393/1406, 27.95%) than consumers (175/1102, 15.88%) reporting feeling uneasy about consumers using technology for self-diagnosis. The majority of providers (819/1406, 58.25%) preferred a diagnosis be made by a professional compared with 44.46% (490/1102) among consumers (Table 3, Q1).

Table 3. Comparison of survey results between providers and consumers (relative risks [RR] in reference to providers:consumers).

Survey items	Provider, n (%)	Consumer, n (%)	RR	RR 95% CI	*P* value[d]
Q1. Technology [a] (choose one)					<.001
Like technology, prefer professional diagnosis	819 (58.25)	490 (44.46)	1.3	1.2-1.4	
Like technology for diagnosis	194 (13.80)	437 (39.66)	0.70	0.66-0.74	
Uneasy using technology	393 (27.95)	175 (15.88)	1.2	1.1-1.2	
Q2. Support genetic testing [b] (% No)					
Having a baby	172 (12.23)	296 (26.86)	0.83	0.80-0.87	<.001
Diagnose problems in fetus	117 (8.32)	207 (18.78)	0.89	0.86-0.92	<.001
Treat disease	39 (2.77)	70 (6.35)	0.96	0.95-0.98	<.001
Disease prevention	80 (5.69)	73 (6.62)	0.99	0.97-1.0	.66
Treat infections	228 (16.07)	122 (10.16)	1.1	1.0-1.1	<.001
Identify drug side effects	172 (12.23)	145 (13.16)	0.99	0.96-1.0	.81
Prolong lifespan	394 (28.02)	277 (25.14)	1.0	0.99-1.1	.05
Identify cause of death	209 (14.86)	172 (15.61)	0.99	0.96-1.0	.56
Q3. Blood tests using smartphone[b] (% No)	530 (37.70)	399(36.21)	1.0	0.96-1.1	.029
Q4. Send/accept information via smartphone [b] (% No)					
Skin problem	737 (52.42)	412 (37.39)	1.3	1.2-1.4	<.001
Heart rate/rhythm	554 (39.40)	379 (34.39)	1.1	1.0-1.1	<.001
Eye exam	983 (69.91)	565 (51.27)	1.6	1.5-1.8	<.001
Ear exam	962 (68.42)	509 (46.19)	1.7	1.6-1.9	<.001
Q5. Hesitant due to privacy concerns[b] (% true)	492 (34.99)	466 (42.29)	.89	.83-.95	.033
Q6. Ownership of medical record [a] (choose one)					<.001
Provider owns records	613 (43.60)	258 (23.41)	1.4	1.3-1.4	
Patient owns records	431 (30.65)	594 (54.90)	0.66	0.62-0.71	
Don’t know who owns records	362 (25.75)	250 (22.69)	1.0	0.99-1.1	
Q7. Access to med records [b] (% I/Patient have/has a right to see)					
Patient has right to see all test results	1339 (95.23)	1060 (96.19)	0.80	0.55-1.2	.54
Patient has right to see all doctors’ notes	884 (62.87)	984 (89.29)	0.29	0.24-0.35	<.001
Q8. Access to EHR information [b] (% No)					
Could lead to feeling anxious about results	136 (9.67)	724 (65.70)	0.38	0.35-0.41	<.001
Could lead to better management of my health	375 (26.67)	80 (7.26)	1.3	1.2-1.3	<.001
Could lead to requesting unnecessary medical evaluations	257 (18.28)	831 (75.41)	0.30	0.28-0.33	<.001
Q9. Access to lab tests [a] (choose one)					<.001
Provider should review	1096 (77.95)	641 (58.17)	1.9	1.7-2.1	
Patients should have access	182 (12.94)	377 (34.21)	0.76	0.72-0.79	
Doctors review results that may cause concern	128 (9.10)	84 (7.62)	1.0	0.99-1.0	
Q10. Cost medical procedure[b] (% No)	718 (51.07)	549 (49.82)	1.0	0.94-1.1	.34
Q11. Right to know full cost of procedure[b] (% agree)	1364 (97.01)	1057 (95.92)	1.4	0.90-2.1	.13
Q12. Prices charged by different providers[b] (% No)	133 (9.46)	71 (6.44)	1.0	1.0-1.1	.40
Q13. Annual physicals [a] (choose one)					.039

Survey items	Provider, n (%)	Consumer, n (%)	RR	RR 95% CI	*P* value[d]
Annual exam is necessary	837 (59.53)	683 (61.98)	0.94	0.85-1.0	
Alternatives for monitoring health	456 (32.43)	340 (30.85)	1.0	0.97-1.1	
Annual physical unnecessary	113 (9.46)	79 (7.17)	1.0	0.99-1.0	
Q14. Concern about radiation exposure[c], mean (SD)	4.28 (1.7)	3.53 (2.0)			<.001
Q15. Feelings about new technology [b] (choose one)					<.001
Must be mastered	806 (57.33)	405 (36.75)	1.5	1.4-1.6	
It is exciting	548 (38.98)	487 (44.19)	0.91	0.86-0.98	
It is beyond me	37 (2.63)	166 (15.06)	0.87	0.85-0.90	
It scares me	15 (1.07)	44 (3.99)	0.97	0.96-0.98	

[a]Multinomial probit regression.

[b]Probit regression.

[c]Linear regression.

[d]Age and gender modeled as covariates.

Genetic Testing

The majority of both providers and consumers supported genetic testing in medical situations, ranging from identifying and treating diseases such as cancer (providers: 1367/1406, 97.23%; consumers: 1032/1102, 93.65%) to prolonging lifespan (providers: 1012/1406, 71.98%; consumers: 825/1102, 74.86%). Providers and consumers similarly reported high support for using genetic testing for disease prevention, identifying drug side effects, and cause of death. Providers were significantly more likely than consumers to support the use of genetic testing when planning to have a baby (providers: 1234/1406, 87.77%; consumers: 806/1102, 73.14%) and diagnosing problems with a fetus (providers: 1289/1406, 91.68%; consumers: 895/1102, 81.22%). Consumers were more likely to support using genetic testing in treating infections (980/1102, 88.93%) than providers (1178/1406, 83.78%), although the absolute difference was not large (Table 3, Q2).

Smartphone Utilization

Health care providers were less likely to support the use of smartphones (*P*=.029) to perform blood tests. In contrast, consumers showed significantly greater support than providers for using smartphones for diagnosis of most of the surveyed conditions in place of an office visit, with 50-60% of consumers supporting smartphone delivery of information about skin problems, eye examinations, and ear examinations compared with 32-48% of providers (Table 3, Q4). Both providers (852/1406, 60.60%) and consumers (723/1102, 65.61%) endorsed using a smartphone to collect or provide heart rate information.

Privacy

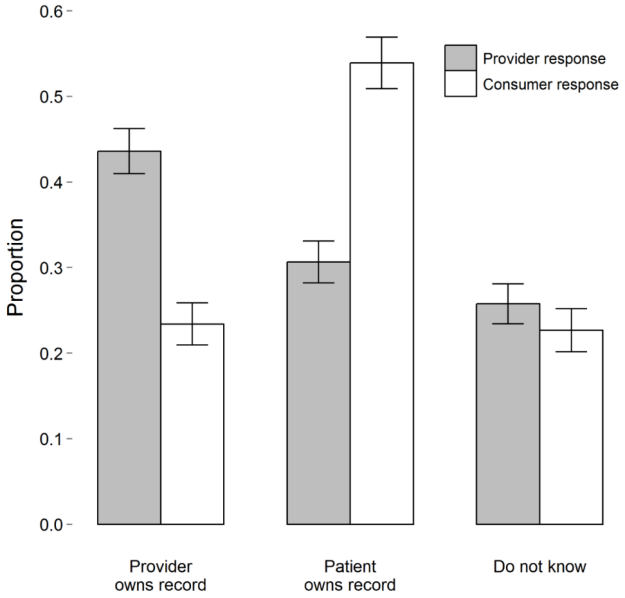
A substantial minority of both providers and consumers expressed hesitancy about privacy and security concerns when using digital health technology, although concern levels were significantly higher among consumers (466/1102, 42.29%) compared with providers (492/1406, 34.99%) (*P*=.033).

Medical Health Records

Ownership

When asked about ownership of the medical record, consumers and providers expressed significant differences in opinion: 43.60% (613/1406) of providers reported that they own their patients' medical records, whereas 53.90% (594/1102) of consumers believed that patients own their medical record (see Figure 1 and Table 3, Q6). Approximately 20% of both groups responded that they did not know who owned the records.

Figure 1. Proportion of responders who believed the patient or provider owned a patient's medical record (error bars represent 95% confidence intervals).



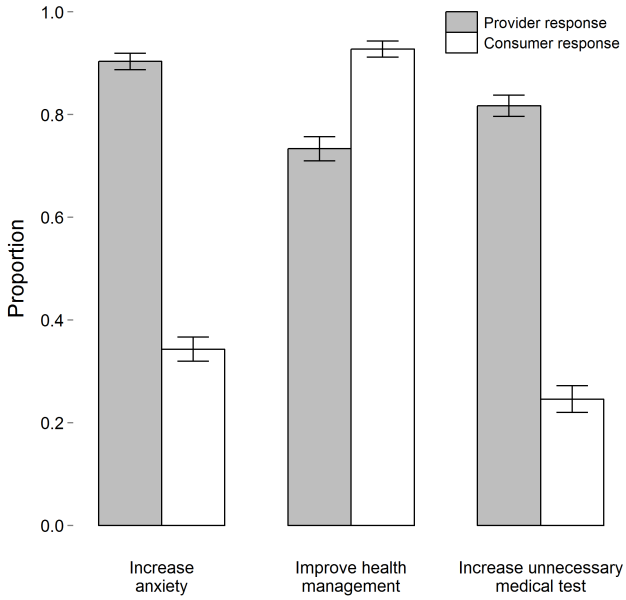
Access to Medical Record Information

A high percentage of both providers (1339/1406, 95.23%) and consumers (1060/1102, 96.19%) agreed that patients should have a right to see all laboratory and diagnostic test results. In contrast, there was a significant difference in opinion regarding access to notes written by the doctor after visits or procedures, with 62.87% (884/1406) of providers agreeing that patients should have access to all notes compared with 89.29% (984/1102) of consumers (Table 3, Q7).

Providers and consumers also differed in their beliefs regarding the consequences of patient access to detailed electronic health records (Figure 2). Most providers (1270/1406, 90.33%) reported concern that patients would experience anxiety after accessing health records, and 81.72% (1149/1406) felt it would lead to requests for unnecessary medical evaluations. Only 34.30% (378/1102) and 24.59% (271/1102) of consumers expressed the same concerns, respectively. While the majority of both groups agreed that access to records could lead to better management of the patient's health, significantly fewer providers (1031/1406, 73.33%) than consumers (1022/1102, 92.74%) shared this belief (Figure 2).

When asked about access to lab tests, one-third of consumers (377/1102) agreed that patients should have access to all of their test results immediately compared to only 12.94% (182/1406) of providers, with a higher percentage of providers than consumers believing that doctors should review all lab results prior to sharing the information with patients (providers: 1096/1406, 77.95%; consumers: 641/1102, 58.17%). Fewer than 10% of individuals in both groups reported that doctors should review lab tests and determine which results may cause the patient worry before sharing the information.

Figure 2. Proportion of responders who believed access to electronic health records would increase anxiety in patients, improve health management, or increase the number of unnecessary medical tests (error bars represent 95% confidence intervals).



Cost and Transparency

Half of the respondents in both groups reported they (consumers) or their patients (providers) ask about the cost of medical services prior to pursuing treatment (Table 3, Q10). More than 90% of both groups agreed that patients should have the right to know the full cost prior to deciding to have a medical procedure and that patients should have access to prices charged by other providers (Table 3, Q11 and Q12).

Physical Exams and Imaging

Physical Exams

Providers and consumers responded similarly to questions about physical exams. A majority of respondents reported that annual exams are necessary (providers: 837/1406, 59.53%; consumers: 683/1102, 61.98%). Less than 10% of both providers and consumers reported that annual exams are not necessary (Table 3, Q13).

Radiation Exposure

As shown in Table 3 (Q14), providers reported significantly more concern than consumers about patient exposure to radiation via various tests (eg, x-rays, mammograms, angiograms). On a 7-point scale, with higher scores corresponding to higher levels of concern, the mean concern level was 4.3 for providers and 3.5 for consumers.

Overall Opinion of New Technology

Providers and consumers differed significantly in their overall opinion of new technology (Table 3, Q15). In regard to overall feelings towards new technology, over half of the providers (806/1406, 57.33%) reported that it must be mastered to stay up-to-date compared with 36.75% (405/1102) of consumers. A higher percentage of consumers reported new technology is exciting (487/1102, 44.19%) compared with providers (548/1406, 38.98%). A subset of consumers reported that new technology is beyond them (166/1102, 15.06%) or scares them (44/1102, 3.99%), whereas, 2% or less of providers endorsed these responses.

Opinion Differences Among Health Care Providers

Differences in response to technology among physicians, medical students, and nurses, nurse practitioners, and physician assistants (collectively grouped as “nurses”) were also examined. Several differences emerged when comparing providers’ response to technology (Table 4). A higher proportion of physicians (137/827, 16.6%) preferred the use of technology for diagnosis when compared to medical students (21/235, 8.9%) and nurses (36/344, 10.5%). Similarly, medical students (154/235, 65.5%) and nurses (231/344, 67.2%) reported that they like technology but preferred that patients seek a professional diagnosis (physicians: 434/827, 52.5%; Table 4, Q1). In regard to genetic testing, physicians (581/827, 70.3%) and nurses (246/344, 71.5%) were less likely to support the use of genetic testing than medical students (185/235, 78.7%) for prolonging the lifespan (Q2). However, when using smartphones for diagnosis, medical students and nurses were less likely to accept an eye exam via a smartphone device (27%) when compared to physicians (274/827, 33.1%; Q4).

In terms of medical record ownership, a higher percentage of physicians (323/827, 39.1%) compared to nurses (78/344, 22.7%) and medical students (30/235, 12.8%) reported that the provider owned the medical record (Table 4, Q6). More doctors and nurses thought that patients should have access to doctors’ notes (physicians: 532/827, 64.3%; nurses: 235/344, 68.3%; medical students: 117/235, 49.8%; Q7). Although all groups agreed that access to electronic health records may increase patient anxiety, more physicians reported that access could lead to better management of health (physicians: 276/827, 33.4%; medical students: 42/235, 17.9%; nurses: 57/344, 16.6%). More nurses (90/344, 26.2%) thought access to electronic health records could lead to unnecessary medical evaluations than doctors or medical students (16%).

Table 4. Comparison of survey results among health care providers.

Survey items	Doctor, n (%)	Medical student, n (%)	Nurses[d], n (%)	*P* value[e]
Q1. Technology [a] (choose one)				.03
Like technology, prefer professional diagnosis	434 (52.5)	154 (65.5)	231 (67.2)	
Like technology for diagnosis	137 (16.6)	21 (8.9)	36 (10.5)	
Uneasy using technology	256 (31.0)	60 (25.5)	77 (22.4)	
Q2. Support genetic testing [b] (% No)				
Having a baby	112 (13.5)	23 (9.8)	37 (10.8)	.58
Diagnose problems in fetus	70 (8.5)	18 (7.7)	29 (8.4)	.89
Treat disease	27 (3.3)	5 (2.1)	7 (2.0)	.98
Disease prevention	56 (6.8)	8 (3.4)	16 (4.7)	.05
Treat Infections	147 (17.8)	23 (9.8)	58 (16.9)	.07
Identify drug side effects	102 (12.3)	21 (8.9)	49 (14.2)	.65
Prolong lifespan	246 (29.7)	50 (21.3)	98 (28.5)	.007
Identify cause of death	132 (16.0)	30 (12.8)	47 (13.7)	.63
Q3. Blood tests using smartphone[b] (% No)	303 (36.6)	92 (39.1)	135 (39.2)	.22
Q4. Send/accept information via smartphone [b] (% No)				
Skin problem	440 (53.2)	124 (52.8)	173 (50.3)	.18
Heart rate/rhythm	321 (38.8)	91 (38.7)	142 (41.3)	.68
Eye exam	562 (68.0)	171 (72.8)	250 (72.7)	.02
Ear exam	553 (66.9)	161 (68.5)	248 (72.1)	.62
Q5. Hesitant due to privacy concerns[b] (% true)	332 (40.1)	66 (28.1)	93 (27.0)	<.001
Q6. Ownership of medical record [a] (choose one)				
Provider owns records	323 (39.1)	30 (12.8)	78 (22.7)	
Patient owns records	314 (38.0)	128 (54.5)	171 (49.7)	
Don’t know who owns records	190 (23.0)	77 (32.8)	95 (27.6)	
Q7. Access to med records [b] (% I/Patient have/has a right to see)				
Patient has right to see all test results	790 (95.5)	223 (94.9)	326 (94.8)	.49
Patient has right to see all doctors’ notes	532 (64.3)	117 (49.8)	235 (68.3)	.03
Q8. Access to electronic health care record information [b] (% No)				
Could lead to feeling anxious about results	75 (9.1)	17 (7.2)	44 (12.8)	.25
Could lead to better management of my health	276 (33.4)	42 (17.9)	57 (16.6)	<.001
Could lead to requesting unnecessary medical evaluations	129 (15.6)	38 (16.2)	90 (26.2)	.02
Q9. Access to lab tests [a] (choose one)				.29
Provider should review	639 (77.3)	181 (77.0)	276 (80.2)	
Patients should have access	122 (14.8)	23 (9.8)	37 (10.8)	
Doctors review results that may cause concern	66 (8.0)	31 (13.2)	31 (9.0)	
Q10. Cost medical procedure[b] (% No)	421 (50.9)	125 (53.2)	172 (50.0)	.94
Q11. Right to know full cost of procedure[b] (% agree)	803 (97.1)	228 (97.0)	333 (96.8)	.89
Q12. Prices charged by different providers[b] (% No)	84 (10.2)	29 (12.3)	20 (5.8)	.16
Q12a. Prepared to compete on price[b] (% No)	362 (43.8)	101 (43.0)	102 (29.7)	<.001

Survey items	Doctor, n (%)	Medical student, n (%)	Nurses[d], n (%)	*P* value[e]
Q13. Annual physicals [a] (choose one)				.55
Annual exam is necessary	479 (57.9)	148 (63.0)	210 (61.0)	
Alternatives for monitoring health	265 (32.0)	76 (32.3)	115 (33.4)	
Annual physical unnecessary	83 (10.0)	11 (4.7)	19 (5.5)	
Q14. Concern about radiation exposure[c], mean (SD)	4.44 (1.7)	3.97 (1.6)	4.10 (1.6)	.003
Q15. Feelings about new technology [b] (choose one)				.31
Must be mastered	478 (57.8)	128 (54.5)	200 (58.1)	
It is exciting	315 (38.1)	99 (42.1)	134 (39.0)	
It is beyond me	26 (3.1)	4 (1.7)	7 (2.0)	
It scares me	8 (1.0)	4 (1.7)	3 (0.9)	

[a]Multinomial probit regression.

[b]Probit regression.

[c]Linear regression.

[d]Nurses, nurse practitioners, and physician assistants.

[d]Age and gender modeled as covariates.

Discussion

Principal Findings

The opinions of consumers and health care providers who completed this survey differ significantly in many scenarios when it comes to new medical technology. Nevertheless, interest in utilizing new technology does exist, with 40% of respondents reporting excitement about using new devices. Although respondents expressed some hesitancy (eg, access to medical records, ownership of records, transmitting information via smartphone, privacy), a majority of individuals from both groups were also in favor of using new medical technology.

A high percentage of consumers and providers agreed on the use of genetic testing to help prevent disease, identify side effects of certain medications, peri-partum diagnostics, and identify cause of death. A recent report of the opinion of parents towards genetic testing early post-partum reinforces our finding of strong consumer support [3].

Similar to consumers, providers believed that patients should have access to lab and diagnostic tests and cost transparency for procedures, but the same was not true for access to office medical notes. Although consumers were willing to use technology for self-diagnosis, providers reported a higher level of unease accepting this information.

Comparison With Prior Research

The largest differences between consumers and providers emerged when assessing access to electronic health records. A marked disparity between health care providers and consumers was noted over concerns that patients would experience an increase in anxiety and request unnecessary health care resources. In contrast to all providers, consumers believed that access would not lead to anxiety, but instead, result in better management of their health care. These perceptions mirror the results of the Open Notes study that found that patients do benefit from access to their medical notes and, although doctors anticipated negative psychological impact, few patients experienced symptoms of anxiety [4]. Prior research suggests that patient access to information generated by new technologies, such as genetic risk information, does not result in an increase in health care utilization [5]. Moreover, there is empirical support for the efficacy of electronic health records access [6] and, in general, patients respond positively to the information [1]. Nonetheless, there is a continued concern that more information via technology will burden physicians and medical resources [2] and that this may have an impact on confidentiality and privacy [1].

Providers were, for the most part, less willing to accept diagnostic information from their patient via smartphone, although that was somewhat information-type dependent with heart rhythm detection being twice as acceptable as diagnostic imaging. The use of camera phones provides another venue of communication, can be a form of empowerment, and can engage the patient in both the diagnostic and management of their own health. Furthermore, instead of hindering rapport, the additional communication and involvement could potentially lead to a stronger doctor-patient relationship [7].

While consumers expressed more privacy concerns for new technology than providers, it was surprising that less than half of the respondents expressed any security concerns. This contrasts with the results of a recent survey of just over 2000 individuals from the Office of the National Coordinator that found nearly 70% of respondents whose providers used electronic health records to be very or somewhat concerned about data privacy, and approximately 75% were concerned about data security [8]. This difference may reflect a higher level of trust in digital data by individuals who routinely used Web-based resources such as WebMD.

The ownership of a medical record was also an area of substantial divergence. Providers believed that they owned their patient's medical record nearly twice as often as did consumers. In contrast, just over 50% of consumers believed that the patient owned their own record. Perhaps surprisingly, a higher proportion of doctors, when compared to other medical students, reported that the patient owned the medical record. Interestingly, a quarter of all respondents did not know who owned the medical record. Another surprising finding was that medical students tended to express more conservative views regarding use of technology in several areas compared with physicians, being more likely to prefer diagnoses to be made by health care providers and less likely to consider patients to own their medical records and to endorse patient access to provider notes.

Limitations

Respondents represent a small proportion of Medscape members and WebMD consumers that elected to participate. Therefore, the results may not represent the larger population of medical providers and consumers. Furthermore, only about 6% of Medscape members and 1% of WebMD consumers who were offered this survey elected to complete it. Thus, our results should be interpreted within the context of two potential biases: (1) membership/visits to these corresponding websites, and (2) a small proportion of eligible respondents. However, a recent report of the results of two non-simultaneous surveys of consumer and provider opinions around digital technology in health care found results consistent with ours [9]. Future studies will benefit from collecting in-depth descriptive statistics and diverse samples to further understand nuanced differences between consumers and providers.

Conclusions

Clinical validation of new digital technologies, with assessment of efficacy, safety, and cost-effectiveness, will be an important part of future research efforts. But understanding the attitudes of patients and physicians may be particularly useful before such validation can occur and especially prior to any widespread potential clinical implementation. The new technologies exemplify the disruption of existing systems of health care—medical information flowing directly to patients, such as with smartphone sensors and lab testing, or with the newfound ability for consumers to perform elements of the physical examination. Our results show that both consumers and health care professionals are generally supportive of these technologies, albeit with sizably greater support and enthusiasm among consumers. Furthermore, the sensitive issue of ownership and access to medical records, where a large gap between consumer and provider expectations exists despite recent clinical validation of transparency, requires considerable further attention. As medicine gets increasingly digitized, the forces favoring democratization will likely be intensified.

Acknowledgments

Financial and material support for the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript were provided by WebMD and the NIH/NCATS Clinical and Translational Science Award UL1TR000109.

Authors' Contributions

DB had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors conducted and are responsible for the data analysis.

Conflicts of Interest

Authors AG, MR, and CS are WebMD employees.

Multimedia Appendix 1

Survey items (consumer survey version).

[PDF File (Adobe PDF File), 141KB - jmir_v17i9e215_app1.pdf]

Multimedia Appendix 2

Survey items (provider survey version).

[PDF File (Adobe PDF File), 141KB - jmir_v17i9e215_app2.pdf]

References

1. Hassol A, Walker JM, Kidder D, Rokita K, Young D, Pierdon S, et al. Patient experiences and attitudes about access to a patient electronic health care record and linked web messaging. J Am Med Inform Assoc 2004;11(6):505-513 [FREE Full text] [doi: 10.1197/jamia.M1593] [Medline: 15299001]
2. Iverson SA, Howard KB, Penney BK. Impact of internet use on health-related behaviors and the patient-physician relationship: a survey-based study and review. J Am Osteopath Assoc 2008 Dec;108(12):699-711. [Medline: 19075034]

3. Waisbren SE, Bäck DK, Liu C, Kalia SS, Ringer SA, Holm IA, et al. Parents are interested in newborn genomic testing during the early postpartum period. Genet Med 2015 Jun;17(6):501-504. [doi: 10.1038/gim.2014.139] [Medline: 25474344]
4. 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]
5. Bloss CS, Schork NJ, Topol EJ. Effect of direct-to-consumer genomewide profiling to assess disease risk. N Engl J Med 2011 Feb 10;364(6):524-534 [FREE Full text] [doi: 10.1056/NEJMoa1011893] [Medline: 21226570]
6. Blumenthal D, Glaser JP. Information technology comes to medicine. N Engl J Med 2007 Jun 14;356(24):2527-2534. [doi: 10.1056/NEJMhpr066212] [Medline: 17568035]
7. Tan L, Hu W, Brooker R. Patient-initiated camera phone images in general practice: a qualitative study of illustrated narratives. Br J Gen Pract 2014 May;64(622):e290-e294 [FREE Full text] [doi: 10.3399/bjgp14X679723] [Medline: 24771843]
8. Patel V, Barker W, Siminerio E. ONC Data Brief, no 20. Washington, DC: The Office of the National Coordinator For Health Information Technology; 2014 Sep. Individuals' Access and Use of their Online Medical Record Nationwide URL: https://www.healthit.gov/sites/default/files/consumeraccessdatabrief_9_10_14.pdf [accessed 2015-08-23] [WebCite Cache ID 6b4hBdql4]
9. Health Research Institute. Healthcare delivery of the future: How digital technology can bridge time and distance between clinicians and consumers. 2014. URL: http://www.pwc.com [accessed 2015-08-25] [WebCite Cache ID 6b4ha6rlU]

Edited by G Eysenbach; submitted 23.03.15; peer-reviewed by A Wong, K Blondon, L Ngo; comments to author 09.07.15; revised version received 16.07.15; accepted 05.08.15; published 14.09.15.

Please cite as:
Boeldt DL, Wineinger NE, Waalen J, Gollamudi S, Grossberg A, Steinhubl SR, McCollister-Slipp A, Rogers MA, Silvers C, Topol EJ
How Consumers and Physicians View New Medical Technology: Comparative Survey
J Med Internet Res 2015;17(9):e215
URL: http://www.jmir.org/2015/9/e215/
doi:10.2196/jmir.4456
PMID:26369254

Original Paper

The Diagnostic Validity and Reliability of an Internet-Based Clinical Assessment Program for Mental Disorders

David Phong Nguyen[1], PhD (Clinical Psychology); Britt Klein[2,3], DPsych (Clinical); Denny Meyer[4], DBL; David William Austin[5], PhD; Jo-Anne M Abbott[1], DPsych (Health)

[1]National eTherapy Centre, Swinburne University of Technology, Hawthorn, VIC, Australia

[2]Centre for Biopsychosocial and eHealth Research & Innovation; Collaborative Research Network; and the School of Health Sciences & Psychology, Federation University Australia, Ballarat, VIC, Australia

[3]National Institute for Mental Health Research, The Australian National University, Canberra, Australia

[4]School of Health Sciences, Swinburne University of Technology, Hawthorn, VIC, Australia

[5]School of Psychology, Deakin University, Burwood, VIC, Australia

Corresponding Author:
David Phong Nguyen, PhD (Clinical Psychology)
National eTherapy Centre
Swinburne University of Technology
Burwood Road
Hawthorn, VIC, 3122
Australia
Phone: 61 402497855
Fax: 61 392145260
Email: dpnguyen@outlook.com

Abstract

Background: Internet-based assessment has the potential to assist with the diagnosis of mental health disorders and overcome the barriers associated with traditional services (eg, cost, stigma, distance). Further to existing online screening programs available, there is an opportunity to deliver more comprehensive and accurate diagnostic tools to supplement the assessment and treatment of mental health disorders.

Objective: The aim was to evaluate the diagnostic criterion validity and test-retest reliability of the electronic Psychological Assessment System (e-PASS), an online, self-report, multidisorder, clinical assessment and referral system.

Methods: Participants were 616 adults residing in Australia, recruited online, and representing prospective e-PASS users. Following e-PASS completion, 158 participants underwent a telephone-administered structured clinical interview and 39 participants repeated the e-PASS within 25 days of initial completion.

Results: With structured clinical interview results serving as the gold standard, diagnostic agreement with the e-PASS varied considerably from fair (eg, generalized anxiety disorder: κ=.37) to strong (eg, panic disorder: κ=.62). Although the e-PASS' sensitivity also varied (0.43-0.86) the specificity was generally high (0.68-1.00). The e-PASS sensitivity generally improved when reducing the e-PASS threshold to a subclinical result. Test-retest reliability ranged from moderate (eg, specific phobia: κ=.54) to substantial (eg, bulimia nervosa: κ=.87).

Conclusions: The e-PASS produces reliable diagnostic results and performs generally well in excluding mental disorders, although at the expense of sensitivity. For screening purposes, the e-PASS subclinical result generally appears better than a clinical result as a diagnostic indicator. Further development and evaluation is needed to support the use of online diagnostic assessment programs for mental disorders.

Trial Registration: Australian and New Zealand Clinical Trials Registry ACTRN121611000704998; http://www.anzctr.org.au/trial_view.aspx?ID=336143 (Archived by WebCite at http://www.webcitation.org/618r3wvOG).

(J Med Internet Res 2015;17(9):e218) doi:10.2196/jmir.4195

KEYWORDS

Internet; online; mental health; validity; reliability; assessment; diagnosis; screening; anxiety; depression

Introduction

The diagnosis of mental disorders has many important roles in clinical practice, research, and administration (eg, communication, treatment planning and evaluation, decision making, classification, policy development) [1]. However, there are various issues that limit the practice and utility of diagnostic assessment in traditional face-to-face settings [2-5]. For example, clinicians typically favor unstructured interviewing despite being prone to bias and error [4], whereas the more reliable structured interviewing format is often overlooked for being cumbersome and costly to administer in everyday practice [5].

The Internet offers various benefits to assist the assessment of mental disorders [6,7]. Internet-based questionnaires can incorporate complex branching and scoring rules, as well as seamlessly present items and feedback in a standardized manner. The Internet also offers minimal ongoing delivery costs, accessibility across diverse population groups, and efficient data collection. Consumer accessibility is typically better than for traditional face-to-face services because it is usually associated with lower cost and greater convenience. Furthermore, the potential anonymity of online assessment facilitates self-awareness and self-disclosure, potentially enabling more valid outcomes [8].

Given these advantages, numerous and diverse online diagnostic assessment tools have been made available. However, published psychometric properties regarding diagnostic outcomes are only available for a small proportion of these. Furthermore, performance varies widely across these reported programs (eg, [9-12]), probably due to differences in program characteristics and study methodologies. For example, Farvolden et al [9] reported on the validity of the Web-Based Depression and Anxiety Test (WB-DAT), a diagnostic screener for depression and anxiety disorders that functions similarly to a structured diagnostic interview based on *Diagnostic and Statistical Manual of Mental Disorders* (Fourth Edition) (*DSM-IV*) criteria. With a clinician-administered Structured Clinical Interview for *DSM-IV* (SCID-IV) as the gold standard, the WB-DAT displayed a high level of diagnostic accuracy in terms of sensitivity (0.71-0.95) and specificity (0.87-0.97). However, results were limited in that participants were recruited from face-to-face clinical trials and may not have represented typical online consumers of the program. Furthermore, the study involved generally low diagnostic base rates that could have biased classification statistics. Nevertheless, the results for the WB-DAT suggest that an online program can achieve a high level of diagnostic sensitivity and specificity.

More recently, Donker et al [10] evaluated the Web Screening Questionnaire (WSQ), which also diagnostically screens for multiple *DSM-IV* disorders (eg, depression, anxiety, and alcohol-related disorders). Unlike the WB-DAT, the WSQ is very brief, with only 1 to 2 items assigned to each disorder and 15 items in total to promote access and completion [10]. In contrast to Farvolden et al's study, participants (N=502) were recruited online and subsequently completed the WSQ remotely to better represent potential program usage. Compared against a telephone-based Composite International Diagnostic Interview (CIDI) as the gold standard, a refined version of the WSQ displayed generally high sensitivity (0.72-1.00). However, the WSQ demonstrated relatively poor specificity (0.44-0.77) and low positive predictive values (PPV=0.11-0.51) with many false positives, probably due to the small item set. Hence, although the WSQ may be diagnostically sensitive and quick to complete, it does so at the expense of specificity when contrasted to a more comprehensive program such as the WB-DAT.

Within the psychometric literature of online diagnostic programs, test-retest reliability seems to be an important, yet underinvestigated, type of reliability given that numerous factors (eg, changes in test-taking attitudes and lack of control in test environment) could vary online performance and subsequent results between sittings [13]. Only one known study has examined the test-retest reliability of an online diagnostic assessment tool. In Lin et al's study, participants comprising Taiwanese visitors to an online mental health website repeated the Internet-based Self-assessment Program for Depression (ISP-D), a 9- to 24-item measure of 3 different depressive presentations [11]. The ISP-D was found to have excellent test-retest reliability within 2 weeks (weighted κ=.80), although performance dropped over longer durations (eg, weighted κ=.45 for 2-4 weeks). Although Lin et al's results are promising, it is unclear whether they can be generalized to programs targeting other disorders and with different population groups.

Given their practical benefits and psychometric evidence, Internet-based diagnostic assessments have been implemented and trialed in "virtual clinics" as a means of rapid assessment and referral to appropriate online interventions [10,14]. One example is the electronic Psychological Assessment and Screening System (e-PASS), which is the focus of this study. Appearing within the Anxiety Online virtual clinic [14] (now renamed as Mental Health Online [15]), the e-PASS predominantly functions as a diagnostic and referral tool for registered users and is the starting point for accessing online treatment programs [14]. For example, a user identified by the e-PASS as having panic disorder would be recommended to complete an online treatment program for panic disorder [14].

Unlike many other diagnostic assessment programs, the e-PASS aims to produce an accurate diagnostic result by incorporating items reflecting diagnostic criteria and severity. The e-PASS also assesses a considerably wider diagnostic breadth, including 21 *DSM-IV* (Text Revision; *DSM-IV-TR*) disorders, compared with most publically available programs, to help accommodate comorbid and lower prevalence disorders. Another distinct attribute of the e-PASS is that it distinguishes the primary diagnosis (ie, the disorder deemed of greatest severity in a presentation) from any secondary disorders. This feature helps users identify their main mental health issue and prioritize treatment recommendations. Finally, the e-PASS focuses on clinical disorders as well as "subclinical" presentations that represent significant symptoms, but do not meet full criteria and severity of a clinical disorder.

Preliminary evaluation has indicated high diagnostic agreement between the e-PASS and community sources (eg, psychologist, counselor, or medical doctor), although results were based on

limited survey data [14]. The e-PASS has also undergone usability testing suggesting it offers distinct benefits and advantages (eg, convenience, anonymity, comprehensiveness) compared to a clinician-administered interview (D Nguyen, unpublished PhD thesis, Victoria: Swinburne University, 2013). In particular, the e-PASS has proven to be highly accessible with more than 22,620 completions between October 2009 and June 2014.

As with any diagnostic assessment tool, it is crucial to formally clarify the psychometric properties of the e-PASS. This need is particularly apparent given the e-PASS' high usage and explicit role in diagnosis and treatment referral as well as outcome measurement in a "virtual" clinic (eg, [14]). Although psychometric evidence for several online assessment programs exist (eg, [9,10]), their findings are limited in reflecting the potential performance of the e-PASS. For example, the e-PASS differs from previously examined programs in terms of identifying a broader range of disorders (including less common disorders such as bulimia nervosa and body dysmorphic disorder) as well as subclinical diagnostic presentations.

Therefore, this study aimed to examine the diagnostic criterion validity and test-retest reliability of the e-PASS involving prospective users completing the e-PASS under relatively naturalistic conditions. This is the first study known to the authors to evaluate both the criterion validity and test-retest reliability of an online multidisorder diagnostic assessment program. This study is also distinct in examining an online diagnostic program that is central to an internationally available open-access "virtual" clinic for mental health disorders. The findings will help facilitate more informed and appropriate use of the e-PASS and further development of the e-PASS and similar online assessment tools.

Methods

Ethical Approval

This study was approved by the Swinburne University Human Research Ethics Committee. The study was conducted as part of a larger trial of the Anxiety Online service, which received trial registration with the Australian New Zealand Clinical Trials Registry (ACTRN12611000704998) [14].

Recruitment

Recruitment targeted prospective e-PASS users. Visitors to the Anxiety Online website who clicked a link to undertake the e-PASS were presented a brief invitation to this research. Those who declined proceeded with the e-PASS per usual, whereas interested individuals were provided with an online plain language statement and consent form. Inclusion criteria required that individuals be 18 years of age or older and residing within Australia (to allow for appropriate follow-up in the advent of participation issues). All clinical populations were welcome, although individuals experiencing acute distress or risk were encouraged to defer participation in the e-PASS study. Recruitment occurred between November 2009 and June 2011. In all, 29 participants were excluded for residing outside of Australia, leaving 616 in the total sample.

The e-PASS

The e-PASS is a comprehensive assessment program that, in addition to diagnostic assessment, measures a range of factors including sociodemographic background, suicide and psychosis risk, past and current treatment, and preferred learning style. The diagnostic component of the e-PASS consists of more than 500 items grouped into modules representing 21 *DSM-IV-TR* disorders [16]: major depressive disorder (MDD), anxiety disorders (eg, panic disorder), body dysmorphic disorder (BDD), eating disorders (eg, bulimia nervosa), sleep disorders (eg, primary insomnia), alcohol and substance dependence (eg, cannabis dependence), pathological gambling, and somatization disorder. Programmed branching rules allow users to automatically skip nonrelevant items. As a result, users typically only complete a subset of all diagnostic items.

Following e-PASS completion, users are presented with detailed feedback, including a primary diagnosis (ie, the disorder rated as most severe) and any secondary disorders identified. Diagnostic severity is based on the extent that symptom criteria are met and rating scores of distress and interference associated with reported symptoms. A "clinical" diagnostic result is given when all symptom criteria are met and rated with at least "mild" to "moderate" distress and interference. A "subclinical" result is assigned when some, but not all, symptom criteria are met or when all symptom criteria are met but overall severity is rated as less than "mild".

Items screening for bipolar disorder and schizophrenia, as well as the potential causal role of a medical condition, substance use, and other notable factors (eg, bereavement in depression symptoms) are also reflected in e-PASS diagnostic feedback (see [14] for a more detailed account).

The Clinical Interview

The clinical diagnostic results of a clinical interview, conducted over telephone, were considered the "gold standard." The use of telephone interviewing for assessing mental health disorders has support in the literature [17-19]. Interviewers were either fully or provisionally registered psychologists undertaking postgraduate clinical training and were blind to participants' e-PASS results. Two interview schedules were predominantly used to reach a diagnosis. All interviews commenced with the administration of the Mini International Neuropsychiatric Interview-Plus (MINI-Plus) structured interview schedule. The MINI-Plus is considered practical, while maintaining high diagnostic reliability and validity with the more cumbersome, but highly regarded, SCID-IV [20]. Participants who endorsed MINI-Plus questions indicating some level of anxiety symptoms were also presented the anxiety disorder modules of the Anxiety Disorders Interview Schedule for *DSM-IV-TR* (ADIS-IV), a "gold standard" semistructured interview with demonstrated reliability [21,22]. Participants who indicated sleep difficulties in response to a screening question were also administered the Insomnia Severity Index, a reliable and valid instrument for identifying clinical insomnia [23].

Procedure

Participants consented by supplying their name, email address, and details of their general practitioner. Participants then

completed the e-PASS, which took a mean 25.0 (SD 5.0) minutes, and received diagnostic feedback as per usual. Between June 2010 and June 2011, all e-PASS participants were sent an email invitation to repeat the e-PASS within 35 days of their initial assessment. Interviewers attempted to call participants within 4 weeks of completing the e-PASS. Due to constraints on the interviewing process (eg, interviewers unavailable), a small minority of the total sample (N=616) were not contacted and, unfortunately, it was not noted who those individuals were. Ultimately, of the 162 participants reached, 158 agreed to interviewing whereas 4 declined due to personal reasons. Interviews were completed a mean of 10.4 (SD 7.0) days after e-PASS and had a mean duration of 48.0 (SD 15.0) minutes.

Interviewers commenced with an introduction then proceeded with administering the MINI-Plus followed by the ADIS-IV and Insomnia Severity Index, where relevant. Interviewers were blind to participants' e-PASS results. Calls ended with participants being invited to other e-PASS–related research activities (eg, qualitative interviewing and online survey of e-PASS experience) not reported in the present study. Following each clinical interview, interviewers completed an assessment summary form including diagnostic outcomes (the presence/absence of a clinical disorder). Interviewers undertook peer supervision and clinical supervision to discuss any clinical concerns and diagnostic issues (eg, differential diagnoses). A random subset of interviews were recorded for interrater reliability testing.

Statistical Analysis

The e-PASS' criterion validity was examined by calculating standard classification statistics including sensitivity, specificity, Cohen's kappa, PPV and negative predictive values (NPV), with diagnostic results of the clinical interview as the criterion (ie, gold standard). Given that classification statistics can be biased by very low diagnostic base rates, only clinical disorders with greater than 4% prevalence according to the clinical interview are reported. Other studies have also reported classification statistics with similarly low base rates (eg, [9,10]).

Sensitivity reflects the proportion of people with a positive clinical interview diagnosis who also received a positive e-PASS diagnosis (ie, true positives). Specificity represents the proportion of those with a negative clinical interview diagnosis who also received a negative e-PASS diagnosis (ie, true negatives). Sensitivity and specificity range from 0 to 1, with higher values indicating better accuracy. Although there are no commonly recommended thresholds for sensitivity/specificity, a minimum sensitivity and specificity of 0.70 was considered acceptable to reflect the priority of screening accuracy [10].

The PPV is the probability of actually having a disorder given a positive diagnosis by the e-PASS, whereas NPV refers to the probability of not actually having a disorder given a negative diagnosis of the disorder by the e-PASS [24]. For sensitivity, specificity, PPV, and NPV, 95% confidence intervals based on the Wilson interval [25] were calculated. Confidence intervals of these statistics reflect potential variability influenced by diagnostic base rates (ie, wider estimates resulting from lower base rates). It is worth noting that previous studies evaluating similar programs (eg, [9,10]) have not included confidence intervals.

Cohen's kappa [26] measures diagnostic agreement beyond that expected by chance [27]. Kappa values were interpreted following guidelines proposed by Landis and Koch [28]: .01-.20=slight, .21-.40=fair, .41-.60=moderate, .61-.80=substantial, and .81-1.00=almost perfect agreement.

Kappa was also used to measure diagnostic agreement between initial and repeated e-PASS results. The McNemar test examined whether there were systematic changes in diagnosis from test to retest. A significant result implies the need to reject the null hypothesis that the clinical diagnosis for a particular disorder has remained consistent between test and retest, and an examination of the contingency table can then show whether the inconsistency reflects a pattern of change from a positive to negative or negative to positive diagnosis from test to retest [29].

Results

Overview

The total sample comprised of 616 people, 443 (71.9%) female and 173 (28.1%) male, with a mean age of 37.7 (SD 12.9) years. The clinical interview sample comprised of 158 people within the total sample. Table 1 shows the sociodemographic characteristics of the total and clinical interview samples. Chi-square tests found no significant differences between the clinical interview sample and the total sample in relation to these sociodemographic variables. A comparison in treatment access showed that a greater proportion were currently accessing treatment within the clinical interview sample (87/158, 55.1%) than the total sample (290/616, 47.1%), but it was not statistically significant (χ^2_1=3.4, P=.06). Furthermore, results indicated cognitive behavioral therapy access was significantly more prevalent among the clinical interview sample (n, 21.2%) than the total sample (n, 14.3%; χ^2_1=6.0, P=.01).

Given so few of the clinical interview subsample (ie, 12 of 158) were eventually recorded, it was decided not to proceed with interrater reliability analysis.

Table 1. Demographic variables of total sample and clinical interview subsample.

Sociodemographic and treatment factors	Total sample, n (%) N=616	Clinical interview subsample, n (%) n=158	χ^2 (*df*)	*P*
Gender			0.2 (1)	.65
Male	173 (28.1)	42 (26.6)		
Female	443 (71.9)	116 (73.4)		
Relationship			0.7 (4)	.94
Married	175 (28.4)	44 (27.8)		
Single	169 (27.4)	44 (27.8)		
De facto	172 (28.0)	46 (29.1)		
Separated or divorced	66 (10.7)	14 (8.9)		
Other	34 (5.5)	10 (6.3)		
Country of birth			3.0 (5)	.70
Australia	453 (73.5)	117 (74.1)		
United Kingdom	53 (8.6)	14 (8.9)		
Asian countries	30 (4.9)	9 (5.7)		
United States	22 (3.6)	2 (1.3)		
European country (except UK)	22 (3.6)	6 (3.8)		
Other	36 (5.8)	10 (6.3)		
Setting			2.6 (3)	.45
Metropolitan	384 (62.3)	104 (65.9)		
Regional	155 (25.2)	36 (22.8)		
Rural	65 (10.6)	13 (8.2)		
Remote	12 (1.9)	5 (3.2)		
Highest schooling			3.7 (3)	.29
Year 9 or less	36 (5.8)	7 (4.4)		
Year 10	70 (11.4)	11 (7.0)		
Year 11	41 (6.7)	12 (7.6)		
Year 12	469 (76.1)	128 (81.0)		
Highest postsecondary education			6.0 (5)	.30
None	89 (14.4)	17 (10.8)		
Current undergraduate	83 (13.4)	15 (9.5)		
Undergraduate	144 (23.4)	40 (25.3)		
Postgraduate	117 (19.0)	38 (24.1)		
Diploma, apprenticeship, trade	92 (14.9)	22 (13.9)		
Certificate	91 (14.8)	26 (16.5)		
Employment			2.6 (6)	.86
Full time	235 (38.1)	65 (41.1)		
Part time	175 (28.4)	42 (26.6)		
Disability, maternity, sick leave	44 (7.1)	10 (6.3)		
Home duties/carer	43 (7.0)	8 (5.1)		
Retired	19 (3.1)	7 (4.4)		
Unemployed	63 (10.2)	17 (10.8)		

Sociodemographic and treatment factors	Total sample, n (%) N=616	Clinical interview subsample, n (%) n=158	χ^2 (*df*)	*P*
Other (eg volunteer, student)	37 (6.0)	9 (5.7)		
Receiving current mental health assistance	290 (47.1)	87 (55.1)	3.4 (1)	.06
Current cognitive behavior therapy access	88 (14.3)	33 (20.9)	6.0 (1)	.01

Diagnostic Validity

Only 10 of the 21 disorders targeted by the e-PASS had sufficient base rates to warrant meaningful classification statistics. Among these, measures of diagnostic accuracy indicated mixed performance (Table 2). Kappa values indicated the e-PASS clinical diagnoses of generalized anxiety disorder (GAD; κ=.37) and obsessive-compulsive disorder (OCD; κ=.39) had fair agreement with the clinical interview. The remaining disorders reflected moderate (bulimia nervosa: κ=.47) to substantial (panic disorder: κ=.62) agreement. Sensitivity ranged from 0.43 (alcohol dependence) to 0.86 (MDD), with half of the disorders falling below the acceptable value of 0.70. When taking into account confidence intervals, sensitivity estimates ranged from as low as 0.16 (OCD, alcohol dependence) to a maximum of 0.94 (MDD). In contrast, specificity varied between 0.68 (GAD) and 1.00 (alcohol dependence), with most values greater than 0.90. Estimated specificity values remained generally greater than 0.70 even after considering confidence intervals.

The PPVs primarily varied between 0.45 (posttraumatic stress disorder; PTSD) and 1.00 (alcohol dependence). The NPVs were consistently higher for most disorders, with the smallest magnitude being 0.80 (social phobia) and the remainder equal to or greater than 0.90. From these predictive values, an e-PASS clinical diagnosis appeared to have a low to moderate likelihood of reflecting a positive clinical diagnosis depending on the disorder, whereas a negative e-PASS diagnosis in general was far more likely to be accurate.

Further analyses examined the extent to which an e-PASS clinical or subclinical diagnostic associated with a clinical interview clinical diagnosis. Again, only 10 disorders were considered because of limited base rates and Table 3 summarizes the resulting classification statistics. When considering both a subclinical and clinical e-PASS result as a positive diagnosis, sensitivity ranged from 0.67 (BDD) to 0.98 (MDD) and equaled or exceeded 0.90 for 5 disorders. Specificity was generally lower and varied between 0.38 (MDD) and 0.89 (bulimia nervosa), with only 5 disorders considered acceptable in terms of exceeding 0.70. Kappa values of the e-PASS subclinical/clinical diagnoses remained significant ($P<.001$) and ranged from .18 (PTSD) to .47 (panic disorder, social phobia), with most considered fair (ie, .20-.40) in diagnostic agreement with a clinical interview clinical diagnosis.

The PPVs were generally smaller than those seen when classification was based on the e-PASS clinical diagnosis alone. Only panic disorder and social phobia maintained moderate PPVs with values of 0.48 and 0.58, respectively. As a result of the lower threshold for a positive e-PASS diagnostic result (ie, subclinical rather than clinical diagnosis), the NPVs accordingly increased for all the disorders, with the majority greater than 0.95. This indicates that an individual with the absence of a relevant clinical disorder is very unlikely to receive a positive e-PASS subclinical or clinical diagnosis for that disorder.

Table 2. Classification statistics of e-PASS clinical diagnoses against clinical interview clinical diagnoses (n=158).

e-PASS diagnosis	Clinical interview, n		κ[a]	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
	Yes	No					
Panic disorder			.62	0.71 (0.55-0.84)	0.91 (0.85-0.95)	0.69 (0.53-0.82)	0.92 (0.86-0.95)
Yes	25	11					
No	10	112					
GAD			.37	0.78 (0.62-0.88)	0.68 (0.59-0.76)	0.45 (0.34-0.57)	0.90 (0.82-0.95)
Yes	31	38					
No	9	80					
Social phobia			.52	0.60 (0.47-0.71)	0.90 (0.84-0.96)	0.77 (0.63-0.87)	0.80 (0.72-0.86)
Yes	34	10					
No	23	91					
PTSD			.52	0.75 (0.47-0.91)	0.92 (0.87-0.96)	0.45 (0.26-0.66)	0.98 (0.94-0.99)
Yes	9	11					
No	3	135					
OCD			.39	0.36 (0.16-0.61)	0.97 (0.93-0.99)	0.56 (0.27-0.81)	0.94 (0.89-0.97)
Yes	5	4					
No	9	140					
MDD			.58	0.86 (0.73-0.94)	0.79 (0.71-0.85)	0.61 (0.46-0.76)	0.94 (0.87-0.97)
Yes	38	24					
No	6	90					
Insomnia			.53	0.78 (0.62-0.88)	0.82 (0.74-0.88)	0.56 (0.42-0.69)	0.93 (0.86-0.96)
Yes	28	22					
No	8	100					
BDD			.51	0.67 (0.39-0.86)	0.94 (0.89-0.97)	0.47 (0.26-0.69)	0.97 (0.94-1.00)
Yes	8	9					
No	4	137					
Bulimia nervosa			.47	0.50 (0.24-0.76)	0.97 (0.92-0.99)	0.50 (0.24-0.76)	0.97 (0.93-0.99)
Yes	5	5					
No	5	143					
Alcohol dependence			.59	0.43 (0.16-0.75)	1.00 (0.98-1.00)	1.00 (0.44-1.00)	0.97 (0.94-0.99)
Yes	3	0					
No	4	151					

[a] All kappa values $P<.001$.

Table 3. Classification statistics of the e-PASS subclinical or clinical diagnoses against clinical interview clinical diagnoses (n=158).

e-PASS diagnosis	Clinical interview, n		κ[a]	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
	Yes	No					
Panic disorder			.47	0.89 (0.74-0.95)	0.72 (0.64-0.79)	0.48 (0.36-0.60)	0.96 (0.89-0.98)
Yes	31	34					
No	4	89					
GAD			.21	0.92 (0.88-0.97)	0.40 (0.31-0.49)	0.34 (0.26-0.44)	0.94 (0.84-0.98)
Yes	37	71					
No	3	47					
Social phobia			.47	0.86 (0.75-0.93)	0.65 (0.56-0.74)	0.58 (0.48-0.68)	0.89 (0.80-0.94)
Yes	49	35					
No	8	66					
PTSD			.18	0.92 (0.65-0.99)	0.62 (0.54-0.70)	0.17 (0.01-0.27)	0.99 (0.94-1.00)
Yes	11	55					
No	1	91					
OCD			.33	0.79 (0.52-0.92)	0.81 (0.73-0.86)	0.28 (0.17-0.44)	0.97 (0.93-0.99)
Yes	11	28					
No	3	116					
MDD			.24	0.98 (0.88-1.00)	0.38 (0.29-0.47)	0.38 (0.29-0.37)	0.98 (0.88-1.00)
Yes	43	71					
No	1	43					
Insomnia			.23	0.97 (0.86-1.00)	0.42 (0.33-0.51)	0.33 (0.25-0.42)	0.98 (0.90-1.00)
Yes	35	71					
No	1	51					
BDD			.35	0.67 (0.39-0.86)	0.88 (0.81-0.92)	0.31 (0.17-0.50)	0.97 (0.92-0.99)
Yes	8	18					
No	4	128					
Bulimia nervosa			.45	0.90 (0.60-0.98)	0.89 (0.82-0.93)	0.35 (0.19-0.54)	0.99 (0.96-1.00)
Yes	9	17					
No	1	131					
Alcohol dependence			.26	0.86 (0.49-0.97)	0.83 (0.77-0.89)	0.19 (0.09-0.36)	0.99 (0.96-1.00)
Yes	6	25					
No	1	126					

[a] All kappa values $P<.001$.

Test-Retest Reliability

Of the 60 participants who repeated the e-PASS, 39 did so within 25 days of initial completion (mean 7.98, SD 6.63) and were included in reliability analyses. Participants received a mean 5.05 (SD 2.83) and 4.70 (SD 2.65) subclinical or clinical diagnoses on their first and second administration, respectively, and the difference was not significant ($t_{38}=1.56$, $P=.13$).

Table 4 presents the cross-tabulation of e-PASS clinical diagnoses between initial completion and retesting, as well as the significance level of the McNemar test, the percentage agreement, and the kappa agreement coefficient. Due to the small sample size, the exact binomial probability of the data was used to calculate the McNemar test [30]. This was not significant ($P>.05$) for all disorders considered, indicating a similar likelihood of change from nonclinical to clinical diagnosis and vice versa between testing and retesting results. However, this could also be a result of an underpowered McNemar test given that the sample size was only n=39.

All kappa values were significant and reflected generally strong diagnostic agreement between test and retest. Kappa was particularly high for bulimia nervosa and panic disorder, each

of which was associated with more than 90% agreement. There was less agreement for insomnia, MDD, and specific phobia, although kappa values were still considered moderate to substantial. An inspection of cases with disagreement found that most involved a change from a subclinical/clinical to clinical/subclinical (respectively) result. For example, 4 of 5 cases of disagreement for specific phobia included a change from a clinical to subclinical diagnosis, whereas the remaining case was of a change from neither a subclinical or clinical diagnosis to a clinical diagnosis of specific phobia.

Table 4. Test-retest reliability of e-PASS clinical diagnoses (n=39).

Test	Retest, n		Agreement, %	*P* [a]	κ[b]
	Yes	No			
Panic disorder			94.9	.50	.83
Yes	6	2			
No	0	31			
Social phobia			87.1	.50	.71
Yes	10	2			
No	3	24			
GAD			84.6	.22	.67
Yes	11	5			
No	1	22			
Specific phobia			87.2	.22	.54
Yes	4	4			
No	1	30			
PTSD			89.8	.63	.61
Yes	4	3			
No	1	31			
MDD			78.5	.73	.57
Yes	11	5			
No	3	20			
Bulimia nervosa			97.4	>.99	.87
Yes	10	1			
No	1	27			
BDD			84.6	.22	.60
Yes	7	5			
No	1	26			
Insomnia			77.0	>.99	.53
Yes	12	4			
No	5	18			

[a] McNemar test *P* values.

[b] All kappa values significant at *P*<.001.

Discussion

The e-PASS is a free, internationally available, online diagnostic assessment (and referral) program for numerous mental disorders. As with any diagnostic tool, particularly one that is highly accessible and can be independently undertaken, there is a need to ensure the e-PASS is valid and reliable. Hence, this study evaluated the psychometric properties of the e-PASS, focusing on its diagnostic criterion validity and test-retest reliability. To enhance the ecological validity of the study findings, participants were recruited online and represented prospective e-PASS users completing the program under generally naturalistic conditions.

The e-PASS was found to have mixed diagnostic agreement with the semistructured clinical interview (ie, the gold standard), varying from fair (eg, OCD) to substantial (eg, panic disorder) agreement. Compared to previously evaluated programs, the e-PASS' diagnostic sensitivity generally exceeded some (eg,

Internet-administered CIDI-Short Form [12]), but not other programs (eg, WB-DAT [9], WSQ [10]). In contrast, the e-PASS' specificity was generally high, resulting in far less false-positive results than certain programs (eg, WSQ [10]). Predictive statistics suggest that a positive e-PASS result had at least a 45% probability of accurately reflecting an actual disorder, whereas a negative e-PASS result for most disorders was correct in more than 90% of cases. The latter suggests a general strength of the e-PASS is its ability to rule out a disorder, which could be beneficial in minimizing burden associated with false-positive clinical diagnoses (eg, stigma, unnecessary follow-up assessment, and treatment).

Among previously reported programs, the e-PASS most closely resembles the WB-DAT [9]. When considering mutual disorders, the e-PASS produced similar psychometrics to the WB-DAT, except in the cases of OCD and PTSD, where the e-PASS clinical result was noticeably less sensitive. It is worth remembering that psychometric results of the WB-DAT [9] were based on a sample recruited from a face-to-face clinic population consisting of generally lower diagnostic base rates compared to those seen in this study. Furthermore, the e-PASS assesses a wider range of disorders than the WB-DAT and most other programs. To the best of the authors' knowledge, this is the first study that has reported the psychometric performance of an online program that identifies BDD and bulimia nervosa.

Although the e-PASS screens particularly well for certain disorders (eg, panic disorder, MDD), it seems lacking for others (eg, OCD) when considering the combination of low sensitivity and diagnostic agreement with the clinical interview. Various factors could help explain these mixed classification statistics (eg, imprecise wording of some e-PASS items or unreliable diagnostic criteria for certain disorders). Given that e-PASS specificity often exceeded sensitivity values, one likely explanation is that the e-PASS' diagnostic threshold was too high for particular disorders. In support of this, additional analyses found that sensitivity values consistently improved and exceeded 90% for some disorders (while maintaining reasonable specificity) when considering an e-PASS "subclinical" or "clinical" result as predictive of an actual clinical disorder. This suggests that the majority of actual clinical disorder cases at least received an e-PASS diagnosis of subclinical, if not clinical severity, which provides some reassurance in terms of notifying e-PASS users of potential mental health issues. Furthermore, the e-PASS is designed so that a subclinical result also prompts access to associated online treatment programs or recommendations of further assessment (eg, face-to-face consultation with a health professional) for follow-up.

Nevertheless, the results of this study suggest one way of improving the e-PASS' screening properties in terms of maximizing sensitivity would be to reduce the diagnostic threshold (eg, so that a subclinical result is identified as a clinical disorder). However, this in turn would increase false-positive results, decreasing specificity. The extent to which diagnostic thresholds should be reduced will depend on the impact on the respective sensitivity and specificity properties, determined using receiver operating characteristic (ROC) analyses (D Nguyen, unpublished PhD thesis, Victoria: Swinburne University, 2013). A further consideration is the broader impact of accurate/inaccurate results (eg, potential burden of diagnosis including financial costs, stigma, and access of ineffective treatment) which further contributes to the overall utility of the e-PASS.

The e-PASS also demonstrates strong test-retest reliability for identifying a clinical disorder (particularly for panic disorder and bulimia nervosa) over an average of approximately 1 week and a maximum of 25 days. Compared to the ISP-D online screener for MDD [11], the e-PASS produced comparable consistency in identifying MDD. The results of this study are the first to document the test-retest reliability of an online diagnostic assessment program for the other reported disorders (eg, anxiety disorders, insomnia, bulimia nervosa). In general, the e-PASS' test-retest reliability measures are comparable to those of a computer-assisted administration of the CIDI [31] and a clinician-administered MINI [32].

In this study, the few e-PASS cases with test-retest discrepancies were just as likely to reflect a diagnostic change from clinical to nonclinical compared with nonclinical to clinical. However, this result may have stemmed from underpowered statistical testing given the smaller than expected sample size. On closer inspection, test-retest discrepancies were generally subtle and tended to involve changes from clinical to subclinical results (and vice versa). This may have reflected actual symptom changes given the instability of certain disorders (eg, MDD) over the retesting period of up to 25 days after initial completion. Unfortunately, the reliability sample was too small to limit the analysis to those with shorter test-retest intervals (eg, 1 week). Overall, e-PASS results appear to be generally stable over the short term, which suggests that the potential variability of the online experience does not pose a significant risk to test-retest reliability.

Several limitations should be considered when interpreting the current findings. Firstly, insufficient clinical interviews were recorded to analyze interrater reliability. Also, the administration order of the e-PASS and clinical interview was not counterbalanced and participants' viewing of e-PASS results in particular may have biased subsequent interview responses. The period between e-PASS and clinical interview completion (mean approximately 10 days) as well as between test and retest of the e-PASS (mean approximately 8 days) may have led to actual symptom changes in some cases. Therefore, the reported validity and reliability statistics could be conservative estimates. Furthermore, the limited number of participants repeating the e-PASS prompts the need for further reliability testing with a larger sample, while also possibly indicating that the e-PASS has low acceptability to some users. Indeed, separate research (D Nguyen, unpublished PhD thesis, Victoria: Swinburne University, 2013) has suggested that some of the e-PASS users were deterred from further use due to certain factors (eg, length, perceived repetition, lack of immediate assistance and support).

Participant recruitment targeted prospective e-PASS users to enhance the ecological validity of findings. Although not reported in this study, the sociodemographic characteristics (eg, gender, employment and marital status, education level) of the approximately 13,000 individuals who completed the e-PASS

between October 2009 and October 2012 largely resemble those of this study sample. Nevertheless, the extent to which results based on this study's sample can be generalized to all e-PASS users requires a more detailed analysis of participant characteristics as well as their potential relationship with psychometric properties. For example, it may be that certain individual characteristics (eg, education level) could be more conducive for e-PASS diagnostic validity or reliability.

With the introduction of *DSM-5* [33], there is a need to revise the e-PASS in-line with new criteria and reevaluate its psychometric properties. Program changes will be minor for most disorder modules (eg, for MDD), although some will require substantial changes (eg, PTSD). Interestingly, the best performing e-PASS diagnoses (eg, MDD and panic disorder) are also those with relatively little criteria change from *DSM-IV-TR* to *DSM-5*. The e-PASS targets 21 disorders, but many of these (eg, anorexia nervosa, pathological gambling, substance disorders) were not examined due to very low diagnostic base rates in the sample. Therefore, further evaluation could involve specific population groups to clarify the e-PASS psychometric properties for these disorders. Additional psychometric evaluation could also consider properties such as the internal reliability of individual e-PASS items, although this would require a much larger sample size as well as modifications to the e-PASS form (eg, removing branching rules) to provide a suitable dataset for analysis.

New means of online diagnostic screening raises the issue of whether to replace, adapt, or supplement Internet-based programs such as the e-PASS. There is potential, for example, to incorporate audiovisual content (eg, [34]) that could enhance accessibility and acceptability. In light of its mixed diagnostic performance, Internet-based screening could also be followed up with clinician interviewing via videoconferencing (eg, [35,36]) or Web chat (eg, [37,38]). Online assessment could also be complemented with mobile-based applications measuring in-the-moment symptoms via questionnaires [39] or audiovisual cues (eg, speech and body language) of the respondent [40].

In contrast to diagnostic screeners, the use of online clinical scales focusing on dimensional measures may prove to offer greater utility in the assessment of mental health disorders [41]. Such programs extend beyond Internet administrations of standard paper-and-pencil measures and are becoming increasingly sophisticated. For example, Batterham et al [42] proposed a hierarchical system commencing with brief online prescreening (eg, K6) followed by an administration of relevant disorder-specific scales. Computer adaptive testing based on item response theory also shows promise in terms of efficiently screening latent traits underlying mental disorders (eg, [43,44]).

In the meantime, given the utility of a diagnosis in clinical practice [1], there is still arguable value in offering Internet-based questionnaires that produce diagnostic results and directly query diagnostic criteria as similar to the approach of gold standard structured clinical interview schedules [5]. As this study shows, an Internet-based diagnostic assessment program can produce diagnostic results that have high test-retest reliability and, at least for certain disorders, high criterion validity. Despite their potential psychometric limitations, these programs could be incorporated into traditional clinical practice alongside other imperfect assessment means (eg, unstructured interviewing) to broaden assessment information and improve overall diagnostic accuracy [3,5]. For many consumers who are unable or unwilling to access traditional services, Internet-based programs could offer a "good enough" alternative for identifying mental health disorders.

In conclusion, this study suggests that the e-PASS has potential for assisting in the diagnosis of mental health disorders and, in doing so, facilitating access to appropriate interventions among other benefits of identifying mental disorders. Nevertheless, further development and evaluation is needed to clarify the full scope of its clinical utility.

Acknowledgments

We would like to thank the Australian Government Department of Health who provided the funding for the development and maintenance of the e-PASS and the broader Anxiety Online service. The funders had no other involvement in this study or report. We also thank all those who were involved in the development of the e-PASS and clinical interviewing for this study.

Conflicts of Interest

None declared.

References

1. First MB. Clinical utility in the revision of the Diagnostic and Statistical Manual of Mental Disorders (DSM). Professional Psychology: Research and Practice 2010;41(6):465-473. [doi: 10.1037/a0021511]
2. Garb HN. Clinical judgment and decision making. Annu Rev Clin Psychol 2005;1:67-89. [doi: 10.1146/annurev.clinpsy.1.102803.143810] [Medline: 17716082]
3. Endicott J. Good diagnoses require good diagnosticians: collecting and integrating the data. Am J Med Genet 2001 Jan 8;105(1):48-49. [Medline: 11424996]
4. Brugha TS, Bebbington PE, Jenkins R. A difference that matters: comparisons of structured and semi-structured psychiatric diagnostic interviews in the general population. Psychol Med 1999 Sep;29(5):1013-1020. [Medline: 10576294]

5. Aboraya A, France C, Young J, Curci K, Lepage J. The validity of psychiatric diagnosis revisited: the clinician's guide to improve the validity of psychiatric diagnosis. Psychiatry (Edgmont) 2005 Sep;2(9):48-55 [FREE Full text] [Medline: 21120108]
6. Barak A, English N. Prospects and limitations of psychological testing on the Internet. Journal of Technology in Human Services 2002 Mar;19(2-3):65-89. [doi: 10.1300/J017v19n02_06]
7. Barak A, Klein B, Proudfoot JG. Defining internet-supported therapeutic interventions. Ann Behav Med 2009 Aug;38(1):4-17. [doi: 10.1007/s12160-009-9130-7] [Medline: 19787305]
8. Barak A, Hen L. Exposure in Cyberspace as Means of Enhancing Psychological Assessment. Cambridge, UK: Cambridge University Press; 2008:129-162.
9. Farvolden P, McBride C, Bagby RM, Ravitz P. A Web-based screening instrument for depression and anxiety disorders in primary care. J Med Internet Res 2003;5(3):e23 [FREE Full text] [doi: 10.2196/jmir.5.3.e23] [Medline: 14517114]
10. Donker T, van SA, Marks I, Cuijpers P. A brief Web-based screening questionnaire for common mental disorders: development and validation. J Med Internet Res 2009 Jul;11(3):e19 [FREE Full text] [doi: 10.2196/jmir.1134] [Medline: 19632977]
11. Lin C, Bai Y, Liu C, Hsiao M, Chen J, Tsai S, et al. Web-based tools can be used reliably to detect patients with major depressive disorder and subsyndromal depressive symptoms. BMC Psychiatry 2007;7:12 [FREE Full text] [doi: 10.1186/1471-244X-7-12] [Medline: 17425774]
12. Carlbring P, Forslin P, Ljungstrand P, Willebrand M, Strandlund C, Ekselius L, et al. Is the Internet-administered CIDI-SF Equivalent to a Clinician-administered SCID Interview? Cognitive Behaviour Therapy 2002 Jan;31(4):183-189. [doi: 10.1080/165060702321138573]
13. Buchanan T, Smith JL. Using the Internet for psychological research: personality testing on the World Wide Web. Br J Psychol 1999 Feb;90 (Pt 1):125-144. [Medline: 10085550]
14. 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. [doi: 10.2196/jmir.1918] [Medline: 22057287]
15. Mental Health Online. URL: https://www.mentalhealthonline.org.au/ [accessed 2015-09-16] [WebCite Cache ID 6bahwRImR]
16. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (4th edition, text revision). Washington, DC: American Psychiatric Association; 2000:0890420254.
17. Aziz MA, Kenford S. Comparability of telephone and face-to-face interviews in assessing patients with posttraumatic stress disorder. J Psychiatr Pract 2004 Sep;10(5):307-313. [Medline: 15361745]
18. Kobak KA, Williams JB, Jeglic E, Salvucci D, Sharp IR. Face-to-face versus remote administration of the Montgomery-Asberg Depression Rating Scale using videoconference and telephone. Depress Anxiety 2008;25(11):913-919. [doi: 10.1002/da.20392] [Medline: 17941100]
19. Rohde P, Lewinsohn PM, Seeley JR. Comparability of telephone and face-to-face interviews in assessing axis I and II disorders. Am J Psychiatry 1997 Nov;154(11):1593-1598. [Medline: 9356570]
20. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59(20):22-33. [Medline: 9881538]
21. Brown TA, Di Nardo PA, Lehman CL, Campbell LA. Reliability of DSM-IV anxiety and mood disorders: implications for the classification of emotional disorders. J Abnorm Psychol 2001;110(1):49-58. [Medline: 11261399]
22. Hilsenroth M, Segal D, Hersen M. Comprehensive Handbook of Psychological Assessment: Personality Assessment. Hoboken, NJ: John Wiley & Sons; 2004.
23. Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med 2001 Jul;2(4):297-307. [Medline: 11438246]
24. Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. Continuing Education in Anaesthesia, Critical Care & Pain 2008 Nov 12;8(6):221-223. [doi: 10.1093/bjaceaccp/mkn041]
25. Brown LD, Cai TT, DasGupta A. Interval estimation for a binomial proportion. Stat Sci 2001;16(2):101-133. [doi: 10.1214/ss/1009213286]
26. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20(1):37-46. [doi: 10.1177/001316446002000104]
27. Streiner D. Diagnosing tests: using and misusing diagnostic and screening tests. J Pers Assess 2003;81(3):209-219. [doi: 10.1207/S15327752JPA8103_03] [Medline: 14638445]
28. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159-174. [Medline: 843571]
29. Field A. Discovering Statistics Using SPSS. London: Sage Publications; 2009.
30. Sheskin D. Handbook of Parametric and Nonparametric Statistical Procedures. Boca Raton, FL: Chapman & Hall; 2000.
31. Wittchen H, Lachner G, Wunderlich U, Pfister H. Test-retest reliability of the computerized DSM-IV version of the Munich-Composite International Diagnostic Interview (M-CIDI). Soc Psychiatry Psychiatr Epidemiol 1998;33(11):568-578. [Medline: 9803825]

32. Lecrubier Y, Sheehan D, Weiller E, Amorim P, Bonora I, Harnett Sheehan K, et al. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. European Psychiatry 1997;12(5):224-231. [doi: 10.1016/S0924-9338(97)83296-8]
33. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (5th edition). Washington, DC: American Psychiatric Association; 2013.
34. van Ballegooijen W, Riper H, Donker T, Martin Abello K, Marks I, Cuijpers P. Single-item screening for agoraphobic symptoms: Validation of a web-based audiovisual screening instrument. PLoS One 2012;7(7):e38480. [doi: 10.1371/journal.pone.0038480]
35. García-Lizana F, Muñoz-Mayorga I. What about telepsychiatry? A systematic review. Prim Care Companion J Clin Psychiatry 2010;12(2):PCC.09m00831 [FREE Full text] [doi: 10.4088/PCC.09m00831whi] [Medline: 20694116]
36. Hyler SE, Gangure DP, Batchelder ST. Can telepsychiatry replace in-person psychiatric assessments? A review and meta-analysis of comparison studies. CNS Spectr 2005;10(5):403-413. [Medline: 15858458]
37. Jowett A, Peel E, Shaw R. Online interviewing in psychology: reflections on the process. Qual Res Psychol 2011;8(4):354-369. [doi: 10.1080/14780887.2010.500352]
38. Mallen MJ, Jenkins IM, Vogel DL, Day SX. Online counselling: an initial examination of the process in a synchronous chat environment. Couns Psychother Res 2011;11(3):220-227. [doi: 10.1080/14733145.2010.486865]
39. Wichers M, Simons CJ, Kramer I. Momentary assessment technology as a tool to help patients with depression help themselves. Acta Psychiatr Scand 2011;124(4):262-272. [doi: 10.1111/j.1600-0447.2011.01749.x] [Medline: 21838742]
40. Joshi J, Dhall A, Goecke R, Breakspear M, Parker G. Neural-net classification for spatio-temporal descriptor based depression analysis. 2012 Presented at: International Conference on Pattern Recognition ICPR2012; Nov 11-15, 2012; Tsukuba, Japan.
41. First MB, Westen D. Classification for clinical practice: how to make ICD and DSM better able to serve clinicians. Int Rev Psychiatry 2007;19(5):473-481. [doi: 10.1080/09540260701563429]
42. Batterham PJ, Calear AL, Sunderland M, Carragher N, Christensen H, Mackinnon AJ. Hierarchical screening for multiple mental disorders. J Affect Disord 2013;151(1):229-236. [doi: 10.1016/j.jad.2013.05.085] [Medline: 23806587]
43. Walter OB, Becker J, Bjorner JB, Fliege H, Klapp BF, Rose M. Development and evaluation of a computer adaptive test for “Anxiety” (Anxiety-CAT). Qual Life Res 2007;16:1-55. [doi: 10.1007/s11136-007-9191-7] [Medline: 17342455]
44. Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): depression, anxiety, and anger. Assessment 2011 Sep;18(3):263-283 [FREE Full text] [doi: 10.1177/1073191111411667] [Medline: 21697139]

Abbreviations

ADIS-IV: Anxiety Disorder Interview Schedule-IV
BDD: body dysmorphic disorder
CIDI: Composite International Diagnostic Interview
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition)
DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition, Text Revision)
DSM-5: Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)
e-PASS: electronic Psychological Assessment Screening System
GAD: generalized anxiety disorder
ISP-D: Internet-based Self-assessment Program for Depression
MDD: major depressive disorder
MINI-Plus: Mini International Neuropsychiatric Interview-Plus
NPV: negative predictive value
OCD: obsessive-compulsive disorder
PPV: positive predictive value
PTSD: posttraumatic stress disorder
SCID-IV: Structured Clinical Interview for DSM-IV
WB-DAT: Web-Based Depression and Anxiety Test
WSQ: Web Screening Questionnaire

Edited by G Eysenbach; submitted 02.01.15; peer-reviewed by P Batterham, J Clarke; comments to author 26.03.15; revised version received 21.05.15; accepted 08.07.15; published 21.09.15.

Please cite as:
Nguyen DP, Klein B, Meyer D, Austin DW, Abbott JAM
The Diagnostic Validity and Reliability of an Internet-Based Clinical Assessment Program for Mental Disorders
J Med Internet Res 2015;17(9):e218
URL: http://www.jmir.org/2015/9/e218/
doi:10.2196/jmir.4195
PMID:26392066

Original Paper

Moving Knowledge Acquisition From the Lecture Hall to the Student Home: A Prospective Intervention Study

Tobias Raupach[1,2], MD, MME; Clemens Grefe[1]; Jamie Brown[2,3], PhD; Katharina Meyer[1]; Nikolai Schuelper[4], MD; Sven Anders[5], PD, MD, MME

[1]University Medical Centre Göttingen, Clinic for Cardiology and Pneumology, Georg-August University Göttingen, Göttingen, Germany
[2]Health Behaviour Research Centre, Department of Epidemiology & Public Health, University College London, London, United Kingdom
[3]Department of Clinical, Educational and Health Psychology, University College London, London, United Kingdom
[4]University Medical Centre Göttingen, Clinic for Haematology and Oncology, Georg-August University Göttingen, Göttingen, Germany
[5]University Medical Centre Hamburg-Eppendorf, Department of Legal Medicine, Hamburg, Germany

Corresponding Author:
Tobias Raupach, MD, MME
University Medical Centre Göttingen
Clinic for Cardiology and Pneumology
Georg-August University Göttingen
Robert-Koch-Strasse 40
Göttingen, D-37075
Germany
Phone: 49 55139 ext 8922
Fax: 49 5513920900
Email: raupach@med.uni-goettingen.de

Abstract

Background: Podcasts are popular with medical students, but the impact of podcast use on learning outcomes in undergraduate medical education has not been studied in detail.

Objective: Our aim was to assess the impact of podcasts accompanied by quiz questions and lecture attendance on short- and medium-term knowledge retention.

Methods: Students enrolled for a cardio-respiratory teaching module were asked to prepare for 10 specific lectures by watching podcasts and submitting answers to related quiz questions before attending live lectures. Performance on the same questions was assessed in a surprise test and a retention test.

Results: Watching podcasts and submitting answers to quiz questions (versus no podcast/quiz use) was associated with significantly better test performance in all items in the surprise test and 7 items in the retention test. Lecture attendance (versus no attendance) was associated with higher test performance in 3 items and 1 item, respectively. In a linear regression analysis adjusted for age, gender, and overall performance levels, both podcast/quiz use and lecture attendance were significant predictors of student performance. However, the variance explained by podcast/quiz use was greater than the variance explained by lecture attendance in the surprise test (38.7% vs 2.2%) and retention test (19.1% vs 4.0%).

Conclusions: When used in conjunction with quiz questions, podcasts have the potential to foster knowledge acquisition and retention over and above the effect of live lectures.

(J Med Internet Res 2015;17(9):e223) doi:10.2196/jmir.3814

KEYWORDS

knowledge; lecture; medical education; podcast; retention

Introduction

In recent years, there has been increasing interest in the use of podcasts as means of knowledge transmission. In broad terms, podcasts can generally be described as audio and/or video files that can be played back on various electronic devices including tablets and smartphones. In fact, the word “podcast,” first used in 2004, is a portmanteau created from the name of one particular device (the iPod) and the word “broadcast.” There is no uniform consensus as to what format or content is required

for an electronic source to be called a "podcast." As a consequence, anything from straight-forward recordings of lectures or conference presentations to complex animated films can be referred to as podcasts. However, some authors have used the term "vodcast" to describe online material containing videos [1] and "enhanced podcast" for audio material supplemented with still images [2]. One common feature of all these formats is that they can be used in an asynchronous manner (ie, at any time, independent of lecture hours).

Within 10 years of their invention, technologies to capture lectures and make them available to students have been embraced by medical teachers involved in both undergraduate and continuing medical education. At the same time, both massive open online courses [3] and scholarly journals [4] now offer a wide range of options to view or listen to material online. User satisfaction is generally high [5], but there is a paucity of data linking podcast use to actual learning outcome. This is in contrast with recent calls for medical school lectures to be moved to online platforms altogether so that classroom time may be used for more efficient teaching activities [6,7]. The underlying assumption is that students viewing course material in preparation of a lecture will retain the content. To our knowledge, this hypothesis has not been tested so far.

A PubMed search combining "medical education" with the terms "podcasts," "lecture video," "online lecture," or "streaming lecture" (search date March 6, 2014) yielded 357 unique citations, and 6 additional articles were identified from reference lists and by contacting experts in the field. Only 78 out of these 363 papers had a specific focus on podcasts. Only 55 of these presented original data, and about half of these (n=27) were related to undergraduate medical education. While half of these (n=13) just reported usage patterns and student satisfaction with podcasts, only 14 original articles assessed the association between podcast use and learning outcome (6 randomized controlled trials, 7 prospective studies, and 1 retrospective analysis). Notably, none of these studies addressed podcast use for preparatory purposes. Instead, podcasts were used to either completely replace or supplement live lectures. In summary, there is currently no scientific data on the effectiveness of using podcasts to stimulate student learning prior to attending a lecture.

The aims of this study were to assess the impact of preparatory podcast use in conjunction with quiz questions versus lecture attendance on short-term and medium-term knowledge retention, and identify significant predictors of short-term and medium-term knowledge retention.

It was hypothesized that students engaging with the material presented in podcasts and submitting answers to quiz questions would retain significantly more knowledge than students not using podcasts. With regard to the second study aim, it was hypothesized that podcast/quiz use would be at least as effective in promoting short-term and medium-term knowledge retention as lecture attendance.

This study did not address any specific psychological framework underlying a potential effect of podcast/quiz use. Instead, it focused on effects elicited by one particular teaching intervention (ie, podcasts and quiz questions) in a "real-world" educational setting.

Methods

Study Design

This study was conducted at Göttingen Medical School. Like most German medical schools, it offers a 6-year undergraduate curriculum comprising 2 preclinical years and 3 clinical years, followed by a practice year. This prospective trial included a cohort of fourth-year medical students who were enrolled in a 6-week cardio-respiratory module in winter term 2013/14. In the preceding summer term, all 37 lectures held during the 6-week module had been recorded using Camtasia Studio 7 (TechSmith). The resulting videos featured the presentation slides used and the lecturer's voice (duration: 35-45 minutes; format: MP4). Following the summer term 2013, the material was reviewed, and the best 10 lecture recordings with regard to sound and image quality were selected to be used in this study. Lecturers were asked to identify key aspects with particular relevance for general internal medicine and to draft free-text questions addressing that content (Table 1).

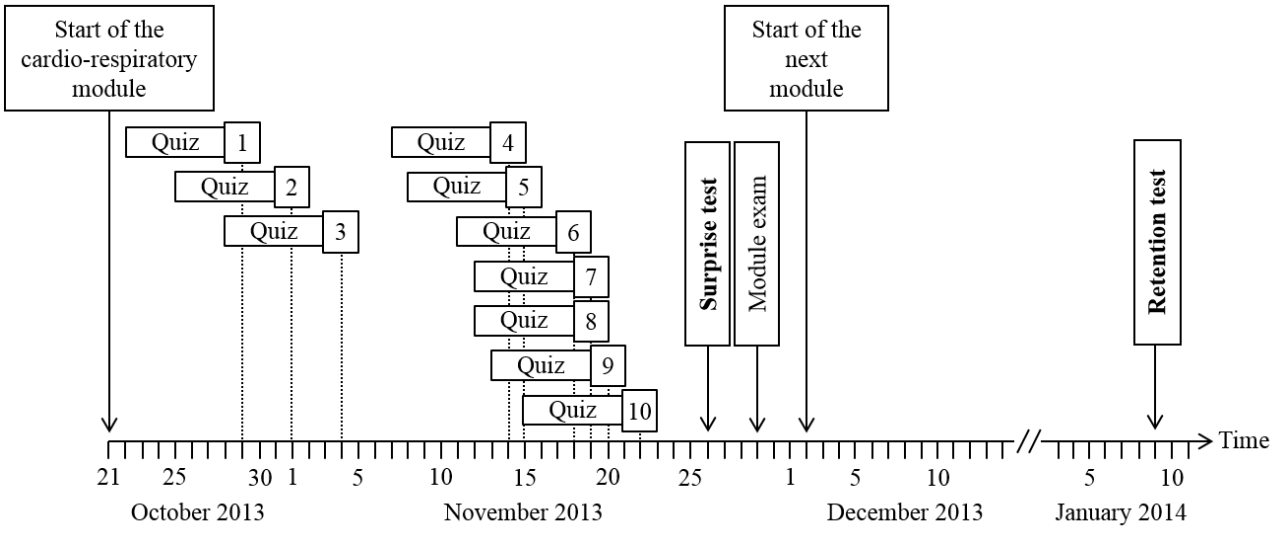
In winter term 2013/14, students enrolled in our module were provided with online access to the 10 selected videos for a period of 7 days before the respective live lectures. A free-text quiz question was linked to each podcast, and students were invited to submit their answers via email until the night before the live lecture. On the day of the lecture, the principal investigator (TR) revealed the correct answer to the entire class and also projected a 1-minute clip from the podcast containing the answer. He then raffled a book voucher (€20) among all students who had submitted a correct answer.

As part of an e-learning session in the final week of the module, students were invited to complete the same 10 quiz questions that had been provided with the podcasts (surprise test). In this session, students were also asked to indicate which lectures they had attended. In order to assess long-term retention, students were invited to answer the same 10 questions again 2 months later in an unannounced retention test. The study outline is summarized in Figure 1.

Table 1. Key aspects covered in podcast lectures and student performance in quizzes as well as in the surprise and retention tests.

Lecture theme	Key aspect	Students with a correct quiz answer n (%)	Students with a correct answer in surprise test n (%)	Students with a correct answer in retention test n (%)
Item 01: Chronic heart failure	Side effects of spironolactone are more pronounced in routine care than in clinical trials due to a lack of potassium monitoring in routine care.	15 (22.4)	13 (19.4)	15 (22.4)
Item 02: Cardiogenic shock	An increase in cardiac output without use of inotropic drugs can be achieved by reducing cardiac afterload.	16 (23.9)	34 (50.7)	34 (50.7)
Item 03: Aortic stenosis	Hallmark symptoms: exertional shortness of breath, angina, and syncope; carotid pulse: prolonged upstroke time.	17 (25.4)	9 (13.4)	11 (16.4)
Item 04: Pacemaker therapy	Effect of placing a magnet over the device: inhibition of shock therapy while pacemaker activity is maintained.	8 (11.9)	21 (31.3)	17 (25.4)
Item 05: Lung function testing	An inhalation test for bronchial hyper-reactivity can be performed only if bronchial obstruction is ruled out in a baseline test.	13 (19.4)	22 (32.8)	23 (34.3)
Item 06: Chronic obstructive pulmonary disease	Neuro-humoral activation is a potential link between intra- and extra-pulmonary manifestations of the disease.	5 (7.5)	14 (20.9)	6 (9.0)
Item 07: Inhaled steroids for asthma	The key to reducing side effects of inhaled steroids was the invention of drugs with high first-pass metabolism.	13 (19.4)	20 (29.9)	5 (7.5)
Item 08: Obstructive sleep apnea	Alcohol intake before going to bed prolongs apneas and causes more pronounced oxygen saturation during sleep.	11 (16.4)	20 (29.9)	16 (23.9)
Item 09: Antibiotics for pneumonia	Ciprofloxacin monotherapy is not recommended as this drug does not target *Streptococcus pneumonia*.	12 (17.9)	27 (40.3)	29 (43.3)
Item 10: Pulmonary fibrosis	New drugs can be assumed to reduce mortality only if this is tested as a primary end point in a randomized trial.	8 (11.9)	15 (22.4)	15 (22.4)

Figure 1. Study outline (numbers in boxes correspond to the 10 lectures used for this study, and vertical dotted lines indicate the date on which live lectures were held). For each lecture, podcasts were available over a period of 7 days leading up to the live lecture. During this time, students were invited to submit their quiz answers.



Student Enrollment and Data Collection

Both the surprise test and the retention test were timed to coincide with scheduled e-learning activity in our institution's computer facilities. Both tests were unannounced in order to avoid confounding by specific preparation, and a time limit of 15 minutes was set for the completion of the 10 quiz questions in both tests. At the beginning of each e-learning session, the study rationale was explained and students were asked to provide written consent to have their data analyzed for study purposes. A total of 10 book vouchers (€20) were raffled among all participants at both the surprise test and the retention test, regardless of test performance.

Questionnaires were created with EvaSys (Electric Paper). In the surprise test, students were asked to provide their age and

gender and to indicate whether they had attended each of the 10 lectures and which podcasts they had watched. The number of quiz answers submitted during the module was derived from the emails sent to the module's administrative staff. The potential impact of recall bias and/or podcast use without answering quiz questions was assessed by comparing the number of students indicating they had watched a particular podcast with the number of students who had submitted an answer to the corresponding quiz question. In order to adjust the analyses for student performance levels, the percent score achieved by each student in the summative end-of-module examination was also obtained (Figure 1). This examination consisted of 25 multiple choice questions addressing factual knowledge on cardiology and pneumology but specifically excluding the content covered by quiz questions as the latter focused on more complex aspects while multiple choice questions were designed to assess basic factual knowledge.

Marking of Quiz Answers

The marking procedure was identical for all three time points (during the module—only students who had submitted an answer via email; surprise test and retention test—all students entering data and consenting to have their data analyzed). After agreeing on corrects answers, 2 raters (TR and CG) independently marked all answers as correct (1) or incorrect (0). Inconsistencies were resolved by discussion. In addition to marking each single question, a sum score (0-10) was calculated, reflecting student attainment in quizzes throughout the module, in the surprise test, and in the retention test.

Statistical Analysis

Unique student identifier codes were used to merge data collected in the surprise and retention tests as well as examination results and data on podcast/quiz use. Data analysis was performed using SPSS Statistics 21 (IBM Corporation). Inter-rater agreement of the marking procedure was assessed by calculating kappa and internal consistency of both tests was assessed by calculating Cronbach alpha.

In order to assess the impact of podcast/quiz use and lecture attendance on short-term and medium-term knowledge retention, the percentage of students providing a correct answer to each question in the surprise and retention tests was calculated. Proportions of students who had/had not submitted a correct quiz answer during the module and those who had/had not attended the corresponding lecture were compared by chi-square tests. Multivariate logistic regression models were run for each of the 10 items with the answer in the surprise/retention test as the dependent variable and controlling for age, gender, and percent score in the end-of-module examination. The comparison between podcast/quiz use and no podcast/quiz use was also adjusted for lecture attendance. Likewise, podcast/quiz use was adjusted for when assessing the impact of lecture attendance on test performance.

A multivariate linear regression analysis was run to identify significant predictors of short-term and medium-term knowledge retention. The dependent variable was the sum score in the surprise/retention test. Age, gender, and percent score in the end-of-module examination as well as the number of submitted correct quiz answers and the number of lectures attended during the module were entered as independent variables.

Group comparisons were performed using chi-square tests (dichotomous variables) and *t* tests (continuous variables). Results of descriptive analyses are presented as percentages and mean with standard deviation (SD), as appropriate. Results of linear regression analyses are reported as unadjusted and adjusted beta values (95% confidence interval) and as the amount of variance explained. Significance levels were set to .05.

Ethical Approval

The local Institutional Review Board (application number 13/12/13) waived ethical approval as the study protocol was not deemed to represent biomedical or epidemiological research. Study participation was voluntary, and all participants signed an informed consent form before entering the study.

Results

Response Rate and Participant Characteristics

Of 130 students enrolled in the module, 126 gave written consent to have their data analyzed for this study. Only students with complete data in both the surprise and the retention tests and the end-of-module examination were included in the final analysis. A total of 101 students attended both tests, but 3 of these did not take the end-of-module examination and another 31 failed to provide complete information on lecture attendance. Thus, complete data of 67 students (24.2 years [SD 2.9]; 39 female) were available. Of these, 34 had submitted at least one correct answer during the module (mean 3.5 [SD 2.6]). On average, students had attended 7.8 (SD 2.3) live lectures. The percentage of students who recalled watching the podcast among those who had submitted a quiz answer was over 80% for all items, suggesting podcast use was not hugely underreported. On the contrary, the proportion of students who had not submitted a quiz answer among those who recalled watching the podcast ranged from 20% to 50%.

Inter-rater Agreement, Item Characteristics, and Results of the Surprise and Retention Tests

Inter-rater agreement for quiz questions and the surprise and retention tests were acceptable (kappa values were .86, .80, and .90, respectively). Cronbach alpha was .68 and .65 in the surprise and retention tests, respectively. The mean number of correct answers in these tests were 2.9 (SD 2.3) and 2.6 (SD 2.0), respectively. As shown in Table 1, performance in all test items was low to moderate. For example, one-third of students were aware that ruling out bronchial obstruction in a baseline lung function test is a prerequisite for bronchial hyper-reactivity testing and only 1 in 5 students displayed adequate knowledge on how to interpret clinical trial reports.

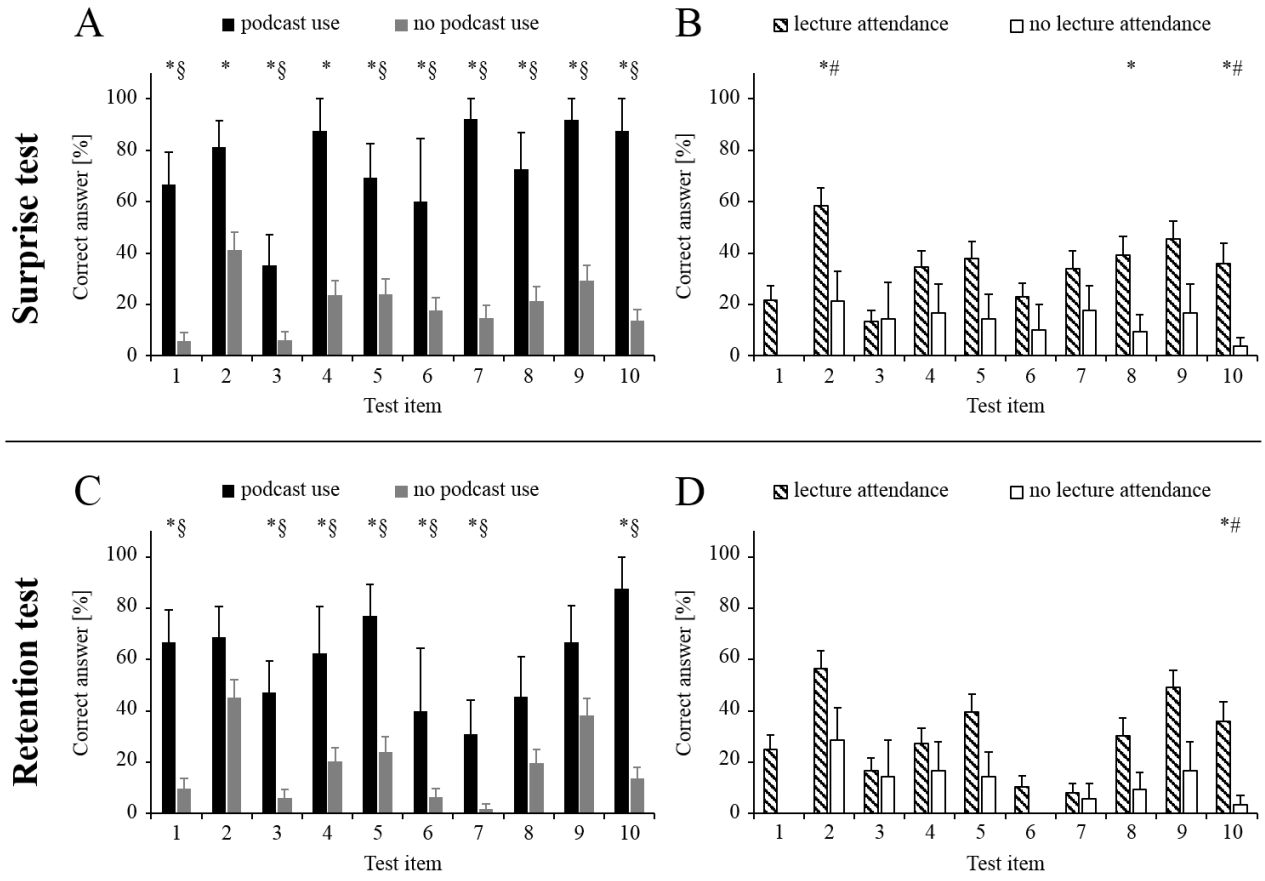
Impact of Podcast Use and Lecture Attendance on Test Performance

Figure 2 presents student performance in the surprise and retention tests as a function of podcast use and lecture attendance.

In the surprise test, podcast use was associated with significantly better knowledge on all test items while such associations with lecture attendance were observed for only 3 items. Similarly, podcast use enhanced knowledge in the retention test for 7 items while there was no such effect of lecture attendance for 9 of 10 items. Adjusting for age, gender, and examination performance attenuated the associations, but the pattern of results was unchanged.

In a sensitivity analysis, percentages of correct answers in the surprise and the retention tests were calculated separately for students who (Group 1) had submitted a quiz answer and recalled having watched the corresponding podcast and students who (Group 2) recalled having watched the podcast but had not submitted a quiz answer. With one exception (Item 2), test performance in the second group was similar to the performance of students who had neither submitted a correct answer nor watched the podcast. The proportion of correct answers in Group 2 was less than half of that observed in Group 1 and was 0% for 3 items in the surprise test and 4 items in the retention test (data not shown).

Figure 2. Student performance in the surprise and retention tests (columns represent the percentage of students providing a correct answer; error bars represent standard errors). * P<.05 for direct comparison (chi-square test); § P<.05 in a logistic regression adjusted for age, gender, exam performance, and lecture attendance; # P<.05 in a logistic regression adjusted for age, gender, exam performance, and podcast use.



Predictors of Short-Term and Medium-Term Knowledge Retention

Results of the linear regression analyses are reported in Table 2. Podcast/quiz use and lecture attendance were significant predictors of student performance in both tests. However, the variance explained by podcast/quiz use was greater than the variance explained by lecture attendance in the surprise test (38.7% vs 2.2%) and retention test (19.1% vs 4.0%).

Table 2. Predictors of student performance in the surprise and retention tests (R^2, variance explained).

Variables	Unadjusted beta (95% CI)	Adjusted beta (95% CI)	R^2
Sum score in the surprise test			
Female gender	0.95 (-0.16 to 2.06)	0.25 (-0.44 to 0.94)	.003
Age in years on the first day of the module	-0.08 (-0.28 to 0.11)	-0.05 (-0.17 to 0.07)	.004
Percent score in the module examination	0.08 (0.04-0.13)	0.06 (0.03-0.09)	.091
Number of correct quiz answers submitted during the module	0.65 (0.50-0.81)	0.60 (0.46-0.73)	.387
Number of lectures attended during the module	0.39 (0.17-0.61)	0.15 (0.004-0.30)	.022
Sum score in the retention test			
Female gender	0.95 (-0.05 to 1.95)	0.33 (-0.47 to 1.13)	.006
Age in years on the first day of the module	-0.17 (-0.34 to 0.01)	-0.13 (-0.27 to 0.02)	.026
Percent score in the module examination	0.07 (0.03-0.11)	0.05 (0.01-0.08)	.062
Number of correct quiz answers submitted during the module	0.44 (0.27-0.61)	0.38 (0.22-0.54)	.191
Number of lectures attended during the module	0.37 (0.17-0.57)	0.19 (0.02-0.36)	.040

Discussion

Principal Findings

This is the first study to examine the impact of podcast use in conjunction with quizzes prior to lecture attendance on knowledge acquisition and retention in undergraduate medical students. Students who engaged with the material before the lecture displayed improved short-term and medium-term retention, regardless of whether they also attended the lecture. The impact of lecture attendance on knowledge retention was considerably weaker despite the correct answers and the decisive part of the podcast being presented to all students in the lecture hall. The most likely explanation for our finding is that—just like interaction during a live lecture [8]—the questions provided with preparation podcasts stirred student alertness, thus facilitating learning [9]. This notion is supported by the results of the sensitivity analysis indicating that watching podcasts without submitting an answer to the corresponding quiz question did not result in improved short-term or medium-term retention. It might be hypothesized that a similar effect could have been observed for lecture attendance if students had been asked to pay attention to a specific detail during the lecture and submit the answer to a related question afterward. However, according to the rationale outlined earlier, one potential use of podcasts could be to partially move the process of knowledge acquisition from the lecture hall to the preparation phase, thereby enabling teachers and learners to explore new and better ways to spend classroom time [10].

Research Context

There has been some debate about the usefulness of podcasts in medical education. While some authors regard them as “toys” [11] and have called for more research into their actual effectiveness, others have argued that students can benefit from exploring novel technologies even in the absence of randomized controlled trials demonstrating their effectiveness [12].

The 14 published reports on the impact of podcast use on learning outcomes in undergraduate medical students vary considerably with regard to study design and outcome measure used. One retrospective analysis detected a small effect of podcast availability on national licensing examination scores that coincided with a national trend for better examination scores [13]. While 2 of the 7 prospective trials found no effect of supplemental podcasts on test scores [14,15], others did find an effect [16,17]. However, some of these effects were either assessed at a very early follow-up (ie, 5 days [18]) or confined to specific student populations, for example, non-native speakers [19]. In one study, students viewing more lectures were even found to score lower in a consecutive examination [20]. Of the 6 randomized trials published so far, 3 [21-23] found a significant effect of podcast use on student examination performance, whereas the other 3 did not [24-26].

In our study, podcasts were used neither to replace nor supplement lectures but as a preparatory tool. In this regard, our results provide some suggestions on *how* this technology might be used to improve learning outcome [27] (as opposed to assessing *whether* it should be used at all [28]). When combined with quiz questions, the provision of podcasts led to a more favorable learning outcome than lecture attendance itself, and this effect was sustained and robust in the adjusted analysis. Given the relatively low uptake observed in our study and previous studies [20], one potential practical implication of our findings could be making the completion of a “preparatory podcast/quiz task” a requirement for course attendance.

Strengths and Limitations

Whereas many previous outcome studies assessed the association between podcast use and overall examination scores, the surprise and retention tests we used were created specifically for this study, and we made every effort to align test questions to the content taught in podcasts and lectures. Inter-rater agreement and internal consistency of the surprise and retention tests were acceptable, but mean scores in both tests were surprisingly low. One potential explanation for this is that these tests were formative in nature, and students might not have made full efforts to achieve a maximum number of correct answers. However, this should apply to all students (regardless

of podcast/quiz use and lecture attendance), and using summative examinations would have had a confounding effect likely to mask any real effect of podcast/quiz use on knowledge levels [29,30]. Another explanation for the low overall scores observed in this study is that quiz questions were related to complex clinical content that—despite being highly relevant for medical practice—is not usually being covered in undergraduate medical textbooks. Moreover, students at our university are not used to open-ended questions as most end-of-course examinations still consist of multiple choice questions. The small amount of variance in surprise and retention test scores explained by the summative multiple choice examination (9.1% and 6.2%, respectively) can be taken as evidence of discriminant validity in that the study-related tests featuring open-ended questions assessed different types of knowledge than the multiple choice questions presented in the end-of-module examination.

One particular strength of our study was the ability to disentangle the effects of podcast/quiz use and lecture attendance in the adjusted analyses presented in Table 2. These data suggest that following podcast use and submitting a correct answer, attending the live lecture had only limited additional benefit in terms of learning outcome. We cannot rule out the possibility that students prepared for lectures with material other than podcasts and/or quiz questions. However, given the marked performance differences between podcast/quiz users and nonusers, any effect of additional preparation would be either confined to podcast/quiz users or too small to detect in students not using podcasts/quizzes.

We excluded a large number of students due to missing information on lecture attendance. This led to a student sample favoring slightly younger (24.2 [SD 2.9] vs 25.4 [SD 2.9]; *P*=.019) and slightly higher-performing students (end-of-module exam scores: 78.6% [SD 11.3] vs 72.8% [SD 13.8]; *P*=.012). The impact of these variables on our results within the final study sample was accounted for by adjusting our analyses accordingly. In addition to selection bias, recall bias is another potential threat to the validity of our findings. However, a great majority of students who had submitted a quiz answer also recalled having watched the corresponding podcast, rendering underreporting of podcast use unlikely. It might be hypothesized that lectures are in fact effective in helping students to acquire and retain knowledge. In order to artificially increase the effect of podcast/quiz use over that of lecture attendance, podcasts users would have had to systematically underreport lecture attendance. However, this was not the case as students submitting at least one correct quiz answer indicated to have attended significantly more lectures than students not submitting any answer: 8.7 (SD 2.0) versus 7.1 (SD 2.5); *P*=.006. In addition, there was a positive correlation between lecture attendance and podcast use (*r*=.252; *P*=.039), hence the need to control for lecture attendance in the analysis of podcast effectiveness and vice versa.

An alternative approach to addressing our research question would have been to conduct a randomized controlled trial. Although this would have yielded higher internal validity, we doubt that we would have been able to restrict podcast use to a specific student group. The aim of this trial was not to test learning processes induced by the availability of podcasts and quiz questions. Instead, this study assessed the effect of one particular teaching intervention in the "real world" of undergraduate medical education. With regard to generalizability, our findings will need to be replicated in other settings. At the same time, there is no reason to believe that using podcasts supplemented with quiz questions as tools to stimulate student learning would be completely ineffective if implemented in a different medical school.

Conclusions

When used in conjunction with quiz questions, lecture podcasts have the potential to foster knowledge acquisition and retention over and above the effect of live lectures. Our findings might help pave the way to move knowledge acquisition from the lecture hall to the preparatory phase, thereby freeing up valuable lecture time for more effective learner-teacher interactions.

Acknowledgments

The authors would like to thank all medical students who agreed to take part in this study. This study was funded by a grant from the Faculty of Medicine, Georg-August-University Göttingen.

Conflicts of Interest

None declared.

References

1. Kurien G, Biron VL, Campbell C, Cote DW, Ansari K. Can a multisensory teaching approach impart the necessary knowledge, skills, and confidence in final year medical students to manage epistaxis? J Otolaryngol Head Neck Surg 2013;42:51 [FREE Full text] [doi: 10.1186/1916-0216-42-51] [Medline: 24479815]
2. Thapa MM, Richardson ML. Dissemination of radiological information using enhanced podcasts. Acad Radiol 2010 Mar;17(3):387-391. [doi: 10.1016/j.acra.2009.10.010] [Medline: 20004598]
3. Liyanagunawardena TR, Williams SA. Massive open online courses on health and medicine: review. J Med Internet Res 2014;16(8):e191 [FREE Full text] [doi: 10.2196/jmir.3439] [Medline: 25123952]
4. Björk B. A study of innovative features in scholarly open access journals. J Med Internet Res 2011;13(4):e115 [FREE Full text] [doi: 10.2196/jmir.1802] [Medline: 22173122]

5. Childers RE, Dattalo M, Christmas C. Podcast pearls in residency training. Ann Intern Med 2014 Jan 7;160(1):70. [doi: 10.7326/L14-5000-2] [Medline: 24573667]
6. Kalmey J. Stop wasting classroom time: embrace the podcast and use the "lecture" to enhance learning. Acad Med 2013 Aug;88(8):1054. [doi: 10.1097/ACM.0b013e31829a6e74] [Medline: 23899846]
7. Prober CG, Heath C. Lecture halls without lectures--a proposal for medical education. N Engl J Med 2012 May 3;366(18):1657-1659. [doi: 10.1056/NEJMp1202451] [Medline: 22551125]
8. Nelson C, Hartling L, Campbell S, Oswald AE. The effects of audience response systems on learning outcomes in health professions education. A BEME systematic review: BEME Guide No. 21. Med Teach 2012;34(6):e386-e405. [doi: 10.3109/0142159X.2012.680938] [Medline: 22578049]
9. Deslauriers L, Schelew E, Wieman C. Improved learning in a large-enrollment physics class. Science 2011 May 13;332(6031):862-864 [FREE Full text] [doi: 10.1126/science.1201783] [Medline: 21566198]
10. Kanter S. To be there or not to be there: is attendance really the question? Acad Med 2012 Jun;87(6):679. [doi: 10.1097/ACM.0b013e31825a5a2a] [Medline: 22643365]
11. Zanussi L, Paget M, Tworek J, McLaughlin K. Podcasting in medical education: can we turn this toy into an effective learning tool? Adv Health Sci Educ Theory Pract 2012 Oct;17(4):597-600. [doi: 10.1007/s10459-011-9300-9] [Medline: 21544550]
12. White J, Sharma N. Podcasting: a technology, not a toy. Adv Health Sci Educ Theory Pract 2012 Oct;17(4):601-603. [doi: 10.1007/s10459-012-9361-4] [Medline: 22426544]
13. Bridge P, Jackson M, Robinson L. The effectiveness of streaming video on medical student learning: a case study. Med Educ Online 2009;14:11 [FREE Full text] [doi: 10.3885/meo.2009.Res00311] [Medline: 20165525]
14. Beale E, Tarwater PM, Lee VH. A retrospective look at replacing face-to-face embryology instruction with online lectures in a human anatomy course. Anat Sci Educ 2014;7(3):234-241. [doi: 10.1002/ase.1396] [Medline: 23959807]
15. Bacro Thierry RH, Gebregziabher M, Fitzharris TP. Evaluation of a lecture recording system in a medical curriculum. Anat Sci Educ 2010;3(6):300-308. [doi: 10.1002/ase.183] [Medline: 20954266]
16. Evans DJR. Using embryology screencasts: a useful addition to the student learning experience? Anat Sci Educ 2011;4(2):57-63. [doi: 10.1002/ase.209] [Medline: 21381215]
17. Burnette K, Ramundo M, Stevenson M, Beeson MS. Evaluation of a web-based asynchronous pediatric emergency medicine learning tool for residents and medical students. Acad Emerg Med 2009 Dec;16 Suppl 2:S46-S50 [FREE Full text] [doi: 10.1111/j.1553-2712.2009.00598.x] [Medline: 20053211]
18. O'Neill E, Power A, Stevens N, Humphreys H. Effectiveness of podcasts as an adjunct learning strategy in teaching clinical microbiology among medical students. J Hosp Infect 2010 May;75(1):83-84. [doi: 10.1016/j.jhin.2009.11.006] [Medline: 20227141]
19. Shaw GP, Molnar D. Non-native English language speakers benefit most from the use of lecture capture in medical school. Biochem Mol Biol Educ 2011;39(6):416-420. [doi: 10.1002/bmb.20552] [Medline: 22081545]
20. McNulty J, Hoyt A, Gruener G, Chandrasekhar A, Espiritu B, Price R, et al. An analysis of lecture video utilization in undergraduate medical education: associations with performance in the courses. BMC Med Educ 2009;9:6 [FREE Full text] [doi: 10.1186/1472-6920-9-6] [Medline: 19173725]
21. Bhatti I, Jones K, Richardson L, Foreman D, Lund J, Tierney G. E-learning vs lecture: which is the best approach to surgical teaching? Colorectal Dis 2011 Apr;13(4):459-462. [doi: 10.1111/j.1463-1318.2009.02173.x] [Medline: 20041922]
22. Mehrpour SR, Aghamirsalim M, Motamedi SMK, Ardeshir LF, Sorbi R. A supplemental video teaching tool enhances splinting skills. Clin Orthop Relat Res 2013 Feb;471(2):649-654 [FREE Full text] [doi: 10.1007/s11999-012-2638-3] [Medline: 23054528]
23. Hansen M, Oosthuizen G, Windsor J, Doherty I, Greig S, McHardy K, et al. Enhancement of medical interns' levels of clinical skills competence and self-confidence levels via video iPods: pilot randomized controlled trial. J Med Internet Res 2011;13(1):e29 [FREE Full text] [doi: 10.2196/jmir.1596] [Medline: 21447471]
24. Spickard A, Alrajeh N, Cordray D, Gigante J. Learning about screening using an online or live lecture: does it matter? J Gen Intern Med 2002 Jul;17(7):540-545 [FREE Full text] [Medline: 12133144]
25. Schreiber B, Fukuta J, Gordon F. Live lecture versus video podcast in undergraduate medical education: A randomised controlled trial. BMC Med Educ 2010;10:68 [FREE Full text] [doi: 10.1186/1472-6920-10-68] [Medline: 20932302]
26. Spickard A, Smithers J, Cordray D, Gigante J, Wofford JL. A randomised trial of an online lecture with and without audio. Med Educ 2004 Jul;38(7):787-790. [doi: 10.1111/j.1365-2929.2004.01824.x] [Medline: 15200403]
27. Cook DA. The failure of e-learning research to inform educational practice, and what we can do about it. Med Teach 2009 Feb;31(2):158-162. [doi: 10.1080/01421590802691393] [Medline: 19330674]
28. Cook D, Levinson A, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. JAMA 2008 Sep 10;300(10):1181-1196. [doi: 10.1001/jama.300.10.1181] [Medline: 18780847]
29. Raupach T, Brown J, Anders S, Hasenfuss G, Harendza S. Summative assessments are more powerful drivers of student learning than resource intensive teaching formats. BMC Med 2013;11:61 [FREE Full text] [doi: 10.1186/1741-7015-11-61] [Medline: 23497243]

30. Raupach T, Hanneforth N, Anders S, Pukrop T, Th J ten Cate O, Harendza S. Impact of teaching and assessment format on electrocardiogram interpretation skills. Med Educ 2010 Jul;44(7):731-740. [doi: 10.1111/j.1365-2923.2010.03687.x] [Medline: 20528994]

Edited by G Eysenbach; submitted 28.08.14; peer-reviewed by M Fischer, E Santoro; comments to author 17.02.15; revised version received 18.07.15; accepted 02.09.15; published 28.09.15.

Please cite as:
Raupach T, Grefe C, Brown J, Meyer K, Schuelper N, Anders S
Moving Knowledge Acquisition From the Lecture Hall to the Student Home: A Prospective Intervention Study
J Med Internet Res 2015;17(9):e223
URL: http://www.jmir.org/2015/9/e223/
doi:10.2196/jmir.3814
PMID:26416467

Original Paper

Online Prediction of Health Care Utilization in the Next Six Months Based on Electronic Health Record Information: A Cohort and Validation Study

Zhongkai Hu[1,2], PhD; Shiying Hao[3], PhD; Bo Jin[2], MS; Andrew Young Shin[3], MD; Chunqing Zhu[2], MS; Min Huang[4], MS; Yue Wang[3], PhD; Le Zheng[3], MS; Dorothy Dai[2], BS; Devore S Culver[5], MBA; Shaun T Alfreds[5], MBA; Todd Rogow[5], MPA; Frank Stearns[2], MHA; Karl G Sylvester[3*], MD; Eric Widen[2*], MHA; Xuefeng Ling[3*], PhD

[1]College of Computer Science and Technology, Zhejiang University, Hangzhou, China

[2]HBI Solutions, Inc, Palo Alto, CA, United States

[3]Stanford University, Stanford, CA, United States

[4]Shanghai Children's Hospital, Shanghai Jiao Tong University, Shanghai, China

[5]HealthInfoNet, Portland, ME, United States

[*]these authors contributed equally

Corresponding Author:
Xuefeng Ling, PhD
Stanford University
S370 Grant Building
300 Pasteur Drive
Stanford, CA,
United States
Phone: 1 650 427 9198
Fax: 1 650 723 1154
Email: bxling@stanford.edu

Abstract

Background: The increasing rate of health care expenditures in the United States has placed a significant burden on the nation's economy. Predicting future health care utilization of patients can provide useful information to better understand and manage overall health care deliveries and clinical resource allocation.

Objective: This study developed an electronic medical record (EMR)-based online risk model predictive of resource utilization for patients in Maine in the next 6 months across all payers, all diseases, and all demographic groups.

Methods: In the HealthInfoNet, Maine's health information exchange (HIE), a retrospective cohort of 1,273,114 patients was constructed with the preceding 12-month EMR. Each patient's next 6-month (between January 1, 2013 and June 30, 2013) health care resource utilization was retrospectively scored ranging from 0 to 100 and a decision tree–based predictive model was developed. Our model was later integrated in the Maine HIE population exploration system to allow a prospective validation analysis of 1,358,153 patients by forecasting their next 6-month risk of resource utilization between July 1, 2013 and December 31, 2013.

Results: Prospectively predicted risks, on either an individual level or a population (per 1000 patients) level, were consistent with the next 6-month resource utilization distributions and the clinical patterns at the population level. Results demonstrated the strong correlation between its care resource utilization and our risk scores, supporting the effectiveness of our model. With the online population risk monitoring enterprise dashboards, the effectiveness of the predictive algorithm has been validated by clinicians and caregivers in the State of Maine.

Conclusions: The model and associated online applications were designed for tracking the evolving nature of total population risk, in a longitudinal manner, for health care resource utilization. It will enable more effective care management strategies driving improved patient outcomes.

(J Med Internet Res 2015;17(9):e219) doi:10.2196/jmir.4976

KEYWORDS

health care costs; electronic medical record; prospective studies; statistical data analysis; risk assessment

Introduction

Health care spending in the United States has grown rapidly since the 1980s [1]. The total health care expenditures for 2012 were more than US $2.8 trillion, accounting for more than 17% of the gross domestic product [2]. It is estimated that this figure will reach US $3.1 trillion in 2014 [3] and become the largest component of the federal budget by 2015 [4]. The trend of increasing spending on health care demands focused attention, which should include analyzing health care resource utilization drivers and predicting future care resource utilization. An effective prediction of future resource utilization can help improve care resource allocation and care delivery, supporting the transition from a volume-based incentive system to a value-based system. To accomplish this transition, health care organizations [5] will need to meet the multiple medical needs of patients while achieving improved outcomes at reduced expense [6].

A variety of factors can affect resource utilization, such as patient age, care providers, medical technologies, and morbidity of patients [7]. Several statistics-based algorithms and methodologies have been developed to forecast future health care expenditures [8-19]. However, many of these studies had limitations caused by incomplete data resources or research targeted only on a particular subgroup of patients, such as age- [14,19] or disease-specific [9,19] populations. Similarly, many of the existing commercial models for resource utilization prediction were constructed using insurance claim data methodologies [20]. Lack of validation of prospective data was another weakness in some studies [19].

The goal of our study was to develop a population-based predictive model to estimate the risk of health care resource utilization in next 6 months for patients in Maine using the state's health information exchange (HIE) electronic medical record (EMR)-based data system. The ability of a health care provider organization to effectively predict health care resource utilization risks using only EMR data is important in the shifting US health care payment system. Provider systems continue to extend their EMR infrastructure throughout their acute, subacute, and physician provider network capturing greater longitudinal clinical patient histories in their EMR. Further, provider systems are entering into more value- and risk-based contracts creating a need for more predictive and proactive care strategies. Using EMR data solely for predictive model development has 2 benefits: (1) it negates the need for integrating claims data from multiple payer sources which is costly and (2) the EMR data are in real time providing more timely information than latent claims data systems and risk models that are typically 60 to 90 days old by the time a provider receives the information. The data for our study were provided by the HIE in Maine, which contains clinical histories and demographic information derived from EMRs for more than 1 million patients covering all payers, all diseases, and all age groups in Maine. The predictive risk model was constructed by statistically learning the correlations between the 6-month total health care utilization and the preceding 12-month demographic and clinical data. It was validated prospectively on both an individual level and a population (per 1000 patients) level. Applications of the model in analyzing and managing health care resource utilization were explored.

We hypothesized that past 12-month EMR-based clinical histories of patients can be used to predict risks of their next 6-month resource utilization via statistical learning from all Maine HIE patient data contained in the statewide HIE of longitudinal patterns. To empower the visualization and exploration of the total population risks of more than 1 million patients in the State of Maine, online applications were architected, aiming to connect in real time, aggregate and centrally integrate data, and to compute the next 6-month risks for population health management. To our knowledge, this study is the first to predict future health care resource utilization using only EMR data at the patient level across an entire state.

Methods

Ethics Statements

This work was done under a business arrangement between HealthInfoNet (HIN) and HBI Solutions, Inc (HBI). Use of the data is governed by the business associate agreement between HIN and HBI. No protected health information was released for the purpose of research. Instead, HBI implemented their application, which was the foundation of the agreement, and then reported on the findings resulting from applying this model to the products that HIN now deploys in the field.

Because this study analyzed deidentified data, the Stanford University Institutional Review Board considered it exempt (October 16, 2014).

Data Warehouse

We constructed a data warehouse consisting of all the Maine HIE's aggregated patient histories. Data elements included patient demographic information, laboratory tests and results, radiographic procedures, medication prescriptions, and diagnoses and procedures, which were coded according to the *International Classification of Diseases, 9th Revision, Clinical Modification* (*ICD-9-CM*). Diagnoses were further clustered into 190 chronic diagnoses according to the Clinical Classifications Software (CCS) developed at the Agency for Healthcare Research and Quality (AHRQ). The CCS-coded diagnosis was used to determine the presence of chronic conditions for each patient. Census data from the US Department of Commerce Census Bureau were integrated into our data warehouse to provide an approximation of patients' socioeconomic status information in terms of the average household mean/median family income and average level of educational attainment distinguished by the zip code of each patient. Missing data handling is described in Multimedia Appendix 1. Our model initially contained 14,680 elements (features) describing the patient's full clinical histories. The

feature dimension was significantly reduced in the subsequent feature selection process for modeling purposes.

Cohort Construction

The study covered patients visiting any HIN-connected facility from January 1, 2009, to December 31, 2013. To qualify for the study, all patients included were alive and resided in Maine.

A retrospective cohort of 1,273,114 patients, represented by clinical information between January 1, 2012, and December 31, 2012, was assembled to develop a model to predict the risk of health care resource utilization between January 1, 2013, and June 30, 2013. This model was validated by a prospective cohort of 1,358,153 patients with clinical information between July 1, 2012, and June 30, 2013, used to predict the health care resource utilization risk from July 1, 2013, to December 31, 2013. Cohort construction details are shown in Multimedia Appendix 2. Patient demographics of the 2 cohorts are shown in Table 1.

Table 1. Cohort characteristics.

Characteristic	Cohort	
	Retrospective (01/01/12-12/31/12) n=1,273,114	Prospective (07/01/12-06/30/13) n=1,358,153
Gender, n (%)		
Female	669,021 (52.55)	710,042 (52.28)
Male	604,093 (47.45)	648,111 (47.72)
Age (years), median (IQR)	43.71 (22.40-60.87)	43.76 (22.83-61.11)
Family income estimate (US $), median (IQR)	59,209 (49,148-68,589)	58,984 (49,148-68,082)
Education, median (IQR)		
Percent high-school graduate or higher	90.50 (87.30-93.20)	90.40 (87.20-92.80)
Percent bachelor's degree or higher	24.40 (17.90-33.00)	23.90 (17.90-31.30)

Next Six-Month Health Care Resource Utilization Scoring Metric Development

The resource utilization depends on much more than just cost. Mean outpatient, emergency department (ED), and inpatient days are more accurate reflection of the trending of health care resource utilization. The national mean cost for different types of resource utilization was used as the weighting mechanism in our computational analysis. Patient's health care utilization was calculated using mean costs associated with specific encounter types (outpatient, ED, and inpatient days) derived from a national database of historical encounter- and inpatient day-based costs [21,22]. This method created an overall single utilization measure per patient across the varying encounter types. This utilization measure was used as the outcome to predict.

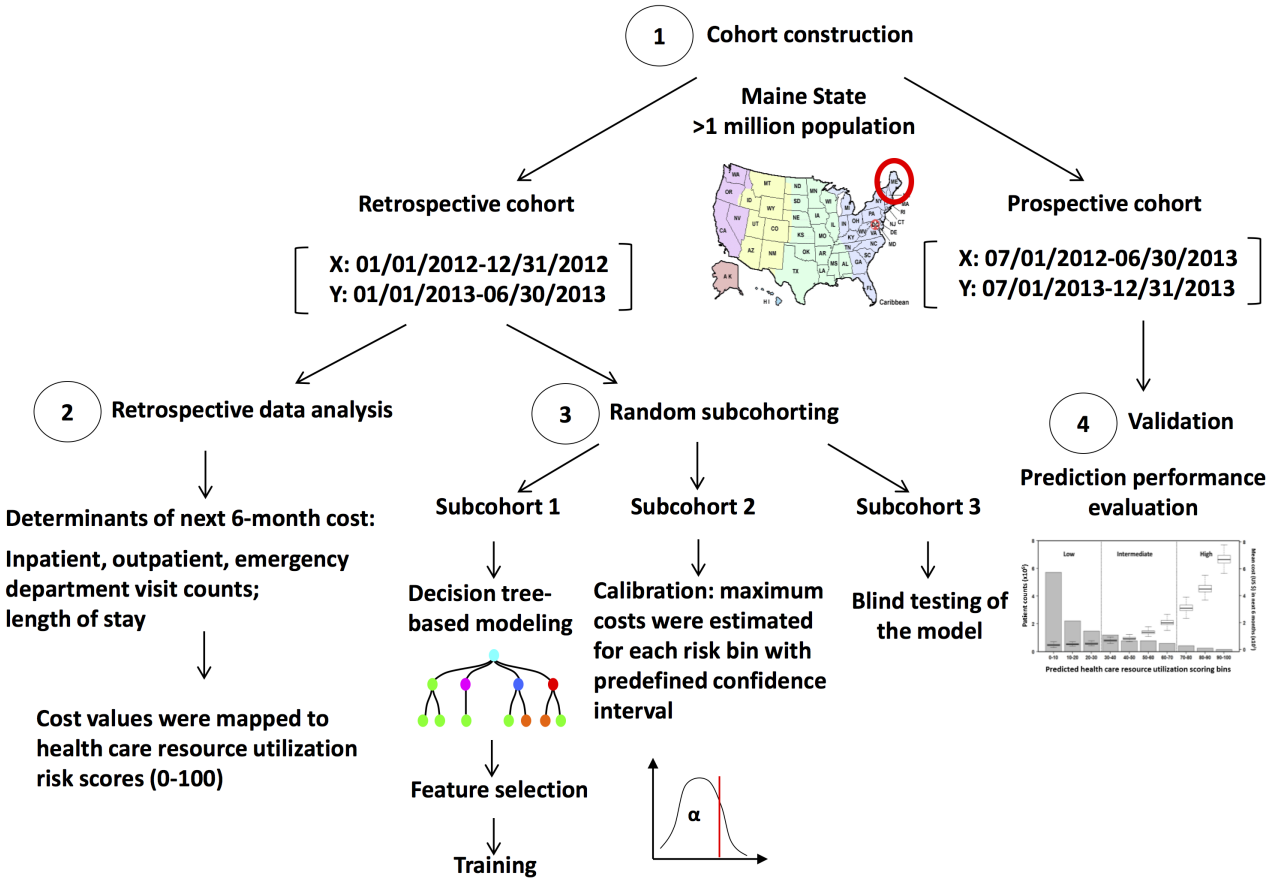
We targeted to predict future risks of patient health care resource utilizations. Future resource utilization distribution analysis on the retrospective cohort revealed a nonlinear correlation between the future resource utilization and the corresponding population sizes. As shown in Multimedia Appendix 3, a small proportion of patients consumed a relatively large amount of health care resources (1.79%, 22,850/1,273,114, of the total population took up 45.96%, US $509.97 million/US $1109.53 million of the next 6-month resource utilization), whereas a large proportion of patients consumed insignificant health care services (62.56%, 796,503/1,273,114, of the population accounted for 2.86%, US $31.74 million/US $1109.53 million of the next 6-month resource utilization). The highly concentrated health care resource utilization distribution of our data indicates that the future utilization characteristics were largely varied among patients of different health statuses, thus revealing the necessity to stratify patients based on the predicted future resource utilization level and treat them separately.

Retrospectively, each patient's next 6-month health care resource utilization was assigned a score ranging from 0 to 100 that correlated with the percentile of the next 6-month resource utilization for that patient in the retrospective cohort. Therefore, a decision tree-based predictive model could be trained with selected preceding 12-month EMR clinical information as independent variables and the next 6-month utilization scores as the dependent variable. The derived predictive algorithm forecasts future risk scores indicative of next 6-month health care resource utilization for each patient.

Model Development

Patients in the retrospective cohort were randomly partitioned into 3 subcohorts (Figure 1) for model training, calibration, and blind testing purposes. The model was then validated with the prospective cohort.

Figure 1. Study design to develop the next 6-month health care resource utilization predictive algorithm. Maine HIE data were split into retrospective and prospective cohorts based on different time frames. A decision tree–based model estimated the health care resource utilization risks in the next 6 months by statistically learning the preceding 12-month clinical histories and were trained, calibrated, and blind tested with the retrospective cohort. The predictive risk model was then validated with the prospective cohort.



Training

Random forest methodology [23-25] was applied to construct 300 decision trees to generate predicted risks based on the preceding 12-month clinical history. Specifically, each of the 300 trees was grown using a randomly selected 63.20% (268,203/424,371) of the patients and one third of the clinical features in the training subcohort. At each node, trees were split by choosing a split feature value producing the minimum sum of square, which equaled the sum of square difference between the risk score of each patient and the mean risk score of all patients on the daughter node to which the patient was assigned. Trees were grown until the size of each terminal node was less than 5. Final decisions were calculated by averaging the predicted risks of each tree.

Calibration

The predictive algorithm derived from the training subcohort was applied on the calibrating subcohort to obtain the risks of patients in that subcohort. The risks were then grouped into 10 bins (0-10, 10-20,..., 90-100), with each bin mapped to a unique health care resource utilization. The value was set as the estimated maximum next 6-month resource utilization of patients falling into a specific bin with a confidence level of alpha. The value of alpha was defined as the proportion of patients in that bin whose next 6-month resource utilization was less than the associated maximum resource utilization value.

The total population was then stratified into 3 main levels indicating an estimation of the level of health care resource utilization for each patient in the next 6 months: high (risk ≥70), intermediate (30≤ risk <70), and low (risk <30). The thresholds (30 and 70) were chosen arbitrarily. Therefore, our analysis produced 2 risk measures: a continuous risk score ranging from 0 to 100 and a categorical risk defined by 3 levels. The former was applied for numerical performance tests whereas the latter was used for stratified analysis in the model validation process, including health care resource utilization analysis and clinical pattern analysis in the model validation process.

Feature Selection

To reduce the calculation complexity of the modeling process, we applied a feature selection process. A total of 2000 features with top variations were first selected from 14,680 initial features to train a random forest model. A list of the features and importance was generated from the random forest model. Second round modeling was done thereafter by using the top 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 features from the feature list. A best ensemble model was chosen according to the performance of sensitivity and confidence level across all

risk levels. Our statistical learning finally identified 70 variables predictive of next 6-month resource utilization risk. These variables were segmented by demographic groups (n=7), different encounter history (n=40), care facilities (n=8), primary and secondary diagnoses (n=14), and outpatient prescription medications (n=1; Table 2).

Table 2. Electronic medical record features used to develop the model.

Feature group	Feature description (12-month clinical history from January 1, 2012-December 31, 2012)
Encounter history (n=40)	Visit counts of different encounter types (emergency/outpatient/inpatient/preadmission)
	The accumulated length of hospitalized stay
	Historical resource utilization
	Counts of historical chronic disease diagnoses
	Counts of total and no redundant total laboratory tests and outpatient prescriptions
Demographics (n=7)	Income, education, payer
	Age group was defined by age on January 1, 2013 (0, 1-5, 6-12, 13-18, 19-34, 35-49, 50-65, and ≥65 years)
Facility (n=8)	Different facilities
Diagnosis (n=14)	Counts for primary diagnosis and secondary diagnosis
Outpatient prescriptions (n=1)	Counts for different outpatient prescriptions

Blind Testing

After calibration, the model's performance was blind tested with the blind testing subcohort. Again, we applied the calibrated model to each patient in that subcohort to derive the predictive risks, grouped all the patients into the 10 bins defined in the calibration process, and identified the proportion of the patients whose resource utilization was less than the estimated maximum resource utilization in each bin.

Model Validation

The predictive algorithm was tested on an independent prospective cohort to validate its effectiveness of risk stratification. The model performance was evaluated on both individual patient level and population level by measuring the confidence levels of prediction and mean resource utilization per person at different risk levels. Clinical patterns of patients at each risk level were analyzed to explore potential applications of the model. A case study was performed to measure the consistency between our risk prediction and resource utilization in a given period (see Multimedia Appendix 4).

Using the Model to Analyze Chronic Disease Care Resource Utilization

The health care resource utilization associated with the presence of chronic disease diagnosed in the preceding 12 months was analyzed using the proposed model. Chronic conditions were defined using the AHRQ Chronic Condition Indicator [26], which provides an effective way to categorize *ICD-9-CM* diagnosis codes into 1 of 2 categories: chronic and nonchronic. The mean resource utilization and population sizes were summarized for all the chronic diseases identified by the CCS coding system. LOESS regression was applied to analyze the correlations between the resource utilization and risk stratification for each chronic disease category.

Online Population Explorer: Statewide Real-Time Surveillance of Population Risks

The risk model and associated online real-time application were designed to track the evolving nature of total population risk of resource utilization in a longitudinal manner. The predictive algorithm was applied to the individual's discriminating feature data extracted from a patient-level database to risk stratify the patients. Individual data were then aggregated for population exploration of resource utilization risks. Results were visualized on an online dashboard. The technical details of our online population explorer implementation are described in Multimedia Appendix 5.

Results

Prospective Performance: Confidence Levels, Mean Resource Utilization, and Risk Stratification

Prospective performance was gauged by the confidence level of prediction and the mean resource utilization per patient in 10 distinctive bins (Table 3). The confidence levels remained at a fairly high level, fluctuating between 0.723 and 0.889 for all risk levels. It illustrated that the predicted risks associated with the estimated maximum resource utilization had an acceptable accuracy for individual patients regardless of their risk levels (ie, resource utilization).

Table 3. Prospective results of our risk model predictive of next 6-month resource utilization (from July 1, 2013, to December 31, 2013).

Result statistics	Predicted risk bin									
	Low			Intermediate				High		
	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Patients, n (%)[a]	571,538 (42.08)	220,746 (16.25)	147,853 (10.89)	119,242 (8.78)	79,152 (5.83)	78,585 (5.79)	59,134 (4.35)	41,711 (3.1)	25,264 (1.86)	14,928 (1.10)
Estimated maximum resource utilization (US $)	0	170	340	510	680	925	1870	2720	4625	13,301
Resource utilization per person[b] (US $), mean (SD)	353.69 (5539.44)	425.86 (7034.20)	449.89 (3194.92)	690.62 (7982.95)	868.47 (11,386.03)	1315.39 (6624.14)	2087.44 (14,347.49)	3211.58 (20,286.97)	4530.99 (12,796.93)	6823.42 (21,814.22)
Confidence level[c]	0.784	0.735	0.805	0.783	0.754	0.723	0.805	0.790	0.796	0.889

[a] Patient percentage of each risk bin is defined as the percentage of patients in that bin of the total prospective population.

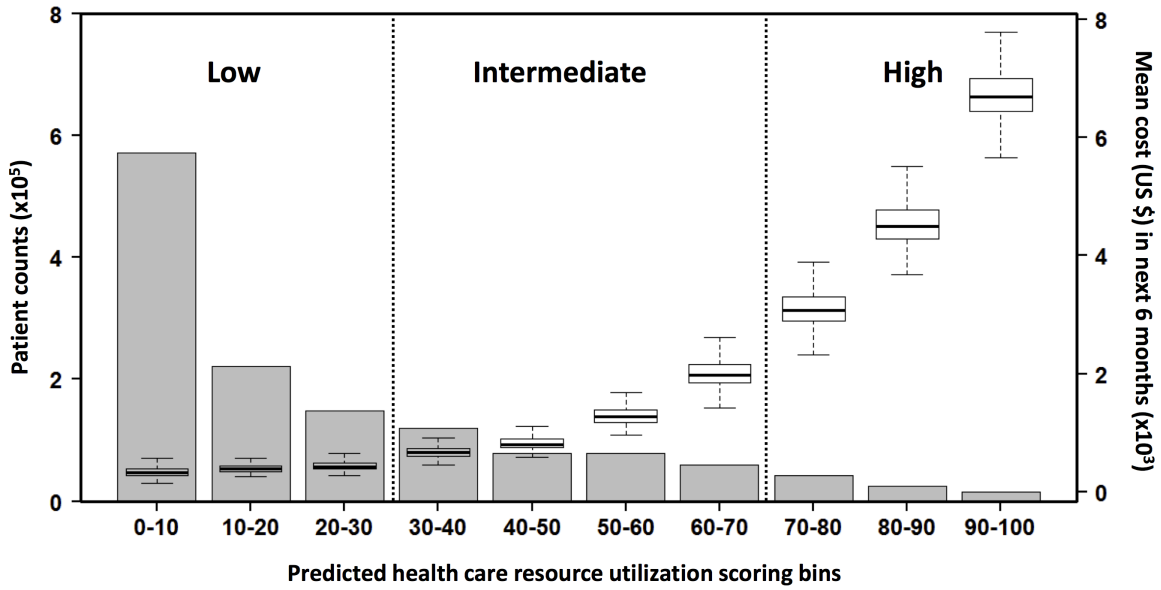
[b] Mean resource utilization per person in each risk bin is defined as the next 6-month mean resource utilization per person in that bin.

[c] Confidence level of each risk bin is defined as the proportion of patients in that bin with next 6-month resource utilization less than the estimated maximum resource utilization.

A monotonic increase was found in mean next 6-month resource utilization per patient from low- to high-risk levels from July 1, 2013, to December 31, 2013 (Table 3). In summary, at low-, intermediate-, and high-risk levels, each patient costs a mean US $385.76, US $1124.33, and US $4276.88 in the next 6 months, respectively. *P* values of resource utilization distributions between these 3 levels were all less than .001 (one-sided Mann-Whitney test). Such findings demonstrate that our predictive model is capable of forecasting the patients who will account for either small or large proportions of next 6-month health care resource utilization.

The distribution of prospective next 6-month resource utilization per 1000 patients in each bin was also analyzed (Figure 2). A monotonic increase in total future spending accompanied by a monotonic decrease in patient counts as the resource utilization risk increases is displayed in the figure. As in the retrospective results, this revealed that a small proportion of patients accounted for a large proportion of the resource utilization: 571,538 patients cost a mean US $353.69, whereas 14,928 patients cost a mean US $6823.42. The box-and-whisker plot also showed that the trend of the total next 6-month resource utilization per 1000 patients correlated with the risk stratification generated by our algorithm: the predicted high resource-utilization population cost more than the predicted low resource-utilization population. It indicates that the division of the population based on the risks produced by the predictive model was intuitive, which supports the use of our model for practical applications focused on identifying patients and populations with high resource utilization for appropriate interventions.

Figure 2. The prospective performance of the model. Prospective validation of the model: the mean next 6-month resource utilization distribution (box-and-whisker plot) and the patient counts (gray bar) versus the predicted risks. The resource utilization distributions were calculated per 1000 patients per 6 months.



Clinical Patterns Associated With Risk Levels

Clinical patterns in the next 6 months of patients in the prospective cohort were summarized based on the health care resource utilization risk levels (see Multimedia Appendix 6). Despite decreases in population and total resource utilization percentages, a monotonic increase of each of the clinical patterns was found from low to high resource-utilization level, including mean resource utilization, percentage of elderly individuals (age ≥ 65 years), percentage of patients making inpatient or ED visits, and percentage of patients with chronic diseases. Patients with high risks of resource utilization were mostly elderly (41.35%, 33,865/81,903), had inpatient (28.55%, 23,381/81,903) or ED visits (45.12%, 36,958/81,903), or had chronic diseases (83.92%, 68,737/81,903). This corresponds with previous research that people who were elderly [27] and had chronic conditions [28] accounted for a large percentage of expenses. Moreover, a fair percentage of high-risk patients had hypertension (30.35%, 24,860/81,903), diabetes (21.69%, 17,764/81,903), heart disease (25.91%, 21,224/81,903), or asthma or chronic obstructive pulmonary disease (14.20%, 11,634/81,903), all of which were reported as major expensive chronic conditions [29]. In all, clinical patterns of high-risk patients identified by our EMR-based algorithm were similar to those revealed by claims data, indicating that reasonable prediction of health care resource utilization can be achieved via EMR and demographics without using any billing information.

Resource Utilization Analysis for Patients with Chronic Diseases

The next 6-month health care resource-utilization patterns were analyzed for patients with chronic diseases (identified by the CCS coding system) in the prospective cohort. There were 178 chronic diseases in total, illustrated as bubbles in Figure 3, in which each bubble represents a patient group sharing the same chronic disease. Patients without any diagnosed chronic disease were grouped separately (the green bubble in Figure 3). The population size for each chronic disease was proportional to the bubble diameter. Figure 3 demonstrates that the algorithm effectively separated the nonchronic from the chronic disease populations. The patients without any chronic disease diagnosis had both the lowest risk (16.1) and lowest next 6-month mean resource utilization (US $490.72). The mean future resource utilization of the chronic disease groups with higher risks tended to be higher than those with lower risks. It should be noted that the outliers in Figure 3 (marked with black circles) represented diseases that only 2-11 patients had in the prospective cohort and thus can be ignored.

The resource utilization and predicted health care resource utilization risks for the 20 most common disease groups are

shown in Figure 4. The LOESS smoothing curve, with 0.95 confidence interval boundary lines, was plotted to estimate the correlation between the mean resource utilization and the corresponding risks for each disease group. The high R^2 value (R^2=.901) and low P value (<.001, calculated by the Gaussian fitting method) indicated that a high linearity was achieved between the next 6-month resource utilization and the mean risk values across the top 20 chronic disease groups. Such linearity validated the effectiveness of our predictive algorithm in risk stratification on the prospective cohort. Furthermore, this semilinear correlation enabled a rough approximation of the future resource utilization for each disease by incorporating the chronic disease diagnosis information into the risk stratification.

Figure 4 illustrates that disorders of lipid metabolism and essential hypertension were 2 chronic diseases with the largest population in our database (n=110,369 and n=113,387, respectively) with mean next 6-month resource utilization of US $2273.22 and US $2512.30 per patient. Conversely, chronic kidney disease, diabetes mellitus with complications, and heart valve disorders were 3 diseases representing the highest mean next 6-month resource utilization (US $5478.18, US $4454.35, and US $4423.01 per patient, respectively), but had relatively small population sizes (n=12,835, n=15,588, and n=13,923, respectively). This confirms that a large amount of the health care resource utilization was consumed by a relatively small percentage of the overall population, most with chronic diseases.

Figure 3. Prospective analysis of next 6-month resource utilizations stratified by chronic diseases. Bubble chart of all 178 chronic diseases (red for diseases with top 20 patient counts and pink for others) stored in our database together with the nonchronic disease group (green). Each bubble represents a chronic disease group, demonstrating mean values of the next 6-month resource utilization and the risks of the patients diagnosed with that disease. The bubble diameter is proportional to the patient counts. Outliers are marked with black circles.

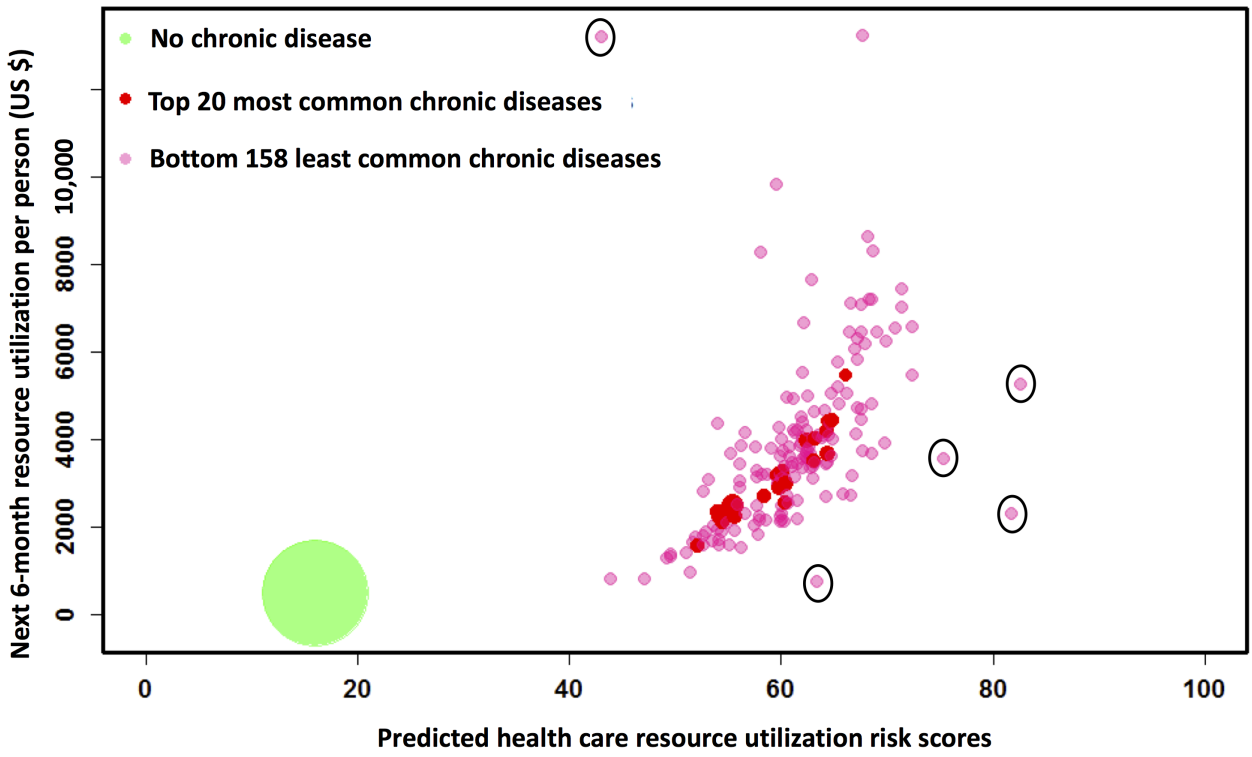
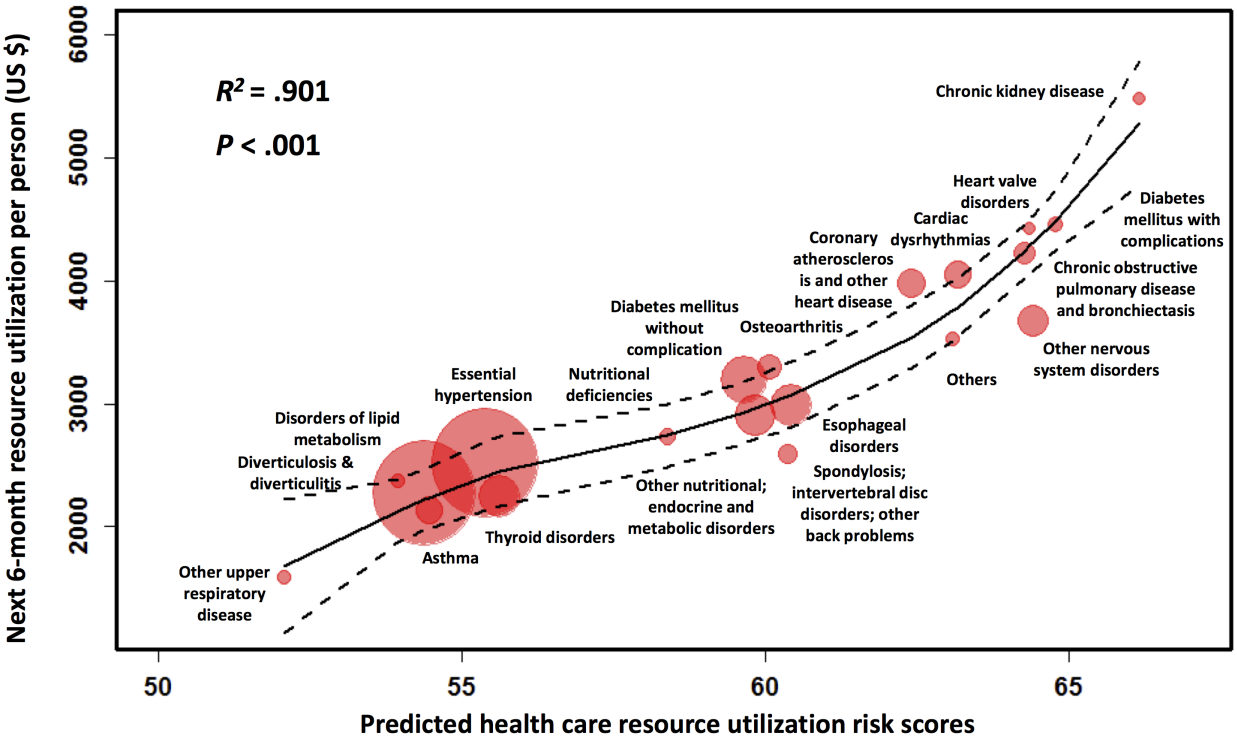


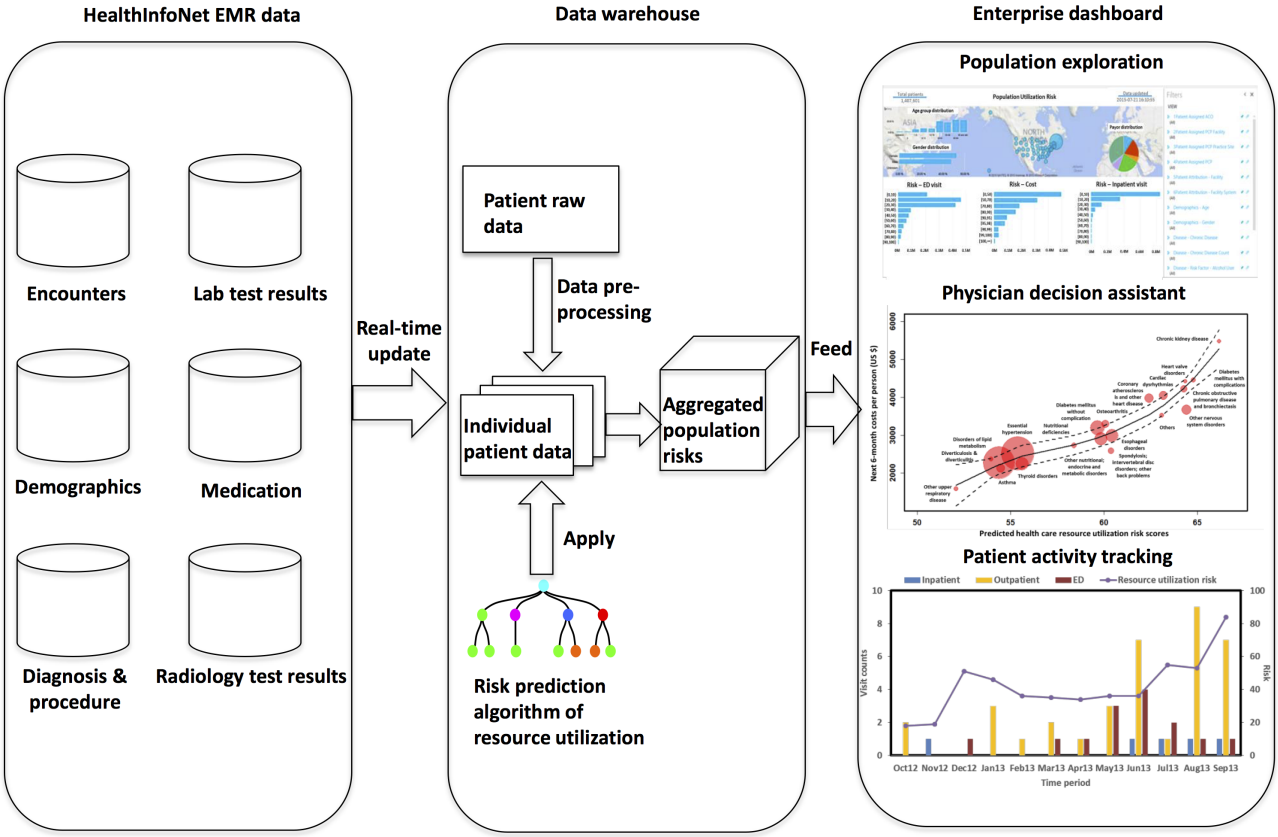
Figure 4. Close examination of the prospective analysis of next 6-month resource utilizations stratified by the top 20 most common chronic diseases. The relationship between the resource utilization and risk score were smoothed by LOESS regression (solid line: the fitting curve; dashed line: the 0.9 confidence level boundaries) showing a good linearity with R-squared=.901 and P <.001.



Online Explorer of Statewide Population Risks of Resource Utilization

Our predictive analytics was integrated into the Maine State HIE system (Figure 5) to allow real-time surveillance of population risks of resource utilization. This online population risk surveillance dashboard (see Multimedia Appendix 7) empowers the Accountable Care Organization field staff and population health managers to visualize the risks derived from each resident's historical medical records in the State of Maine. With our prospectively validated algorithm, our coherent view of population risks of resource utilization can thus be feasible to resolve the barrier of the fragmented nature of population health information to improve public health practice.

Figure 5. Schematic demonstration of data flow and communications of a population risk exploration system, which allows online real-time assessment of population resource utilization risk.



Discussion

Extremely high expenditure rates in health care have become a nationwide issue in the United States. To approach this issue, we developed a model to predict the next 6-month health care resource-utilization risk of patients based on their past 12-month clinical histories. Prospective validation results demonstrated the effectiveness of our model in risk stratification of the future health care spending for patients with sufficient accuracy at all risk levels.

Of significant note in our research is that the proposed risk model was derived using the EMR data from the Maine HIE, the same data source as our previous studies on emergency visit and revisit risks [30,31]. The success of these models demonstrates that, with high dimensional structured data aggregation and statistical learning, past longitudinal clinical patterns can be used to forecast future resource utilization. These studies together establish an approach for developing effective clinical forecasting systems. The approach can be applied to many more use cases including predicting diseases, utilization (inpatient, pharmacy, imaging), and mortality. Integration of HIE data with predictive analytic tools enables cross-analysis and can help construct a comprehensive risk profile of the Maine population.

Our risk model was derived using EMR and demographic data only. Prospective performance (see Multimedia Appendix 6) shows that the high resource-utilization population identified by our EMR-based model had clinical patterns similar to those summarized by claims. Therefore, hospitals can reliably predict future patient resource utilization using their internal EMR and do not necessarily need to incur the resource utilization of integrating insurance claims data to achieve this. Moreover, EMR-based information is generated in near real time, whereas claims-based risk models are typically several months old by the time a provider receives the information [32]. Real-time risk scores facilitate more timely patient care.

The purpose of categorizing risk predictions was to identify high-risk patients (ie, those with relatively high probability of resource utilization). The high-risk group (risk ≥ 70) was more interesting than the other 2 groups because evidence suggests that well-organized interventions targeting high-risk patients can result in a decreased rate of admission or readmission and, therefore, significant resource savings. Further analysis on clinical patterns of high-risk patients could be used for a more personalized or precisely targeted approach to reduce future resource utilization. Furthermore, the online application tool of our model provides a tracking of patients' activities and risks, allowing users to identify patients with increasing risks. Those patients are also of interest to clinicians and early interventions

may prevent them from becoming heavy users of health care resources.

Our algorithm took special focus on the impact of chronic disease history on future resource utilization. As shown in Figure 4, chronic kidney disease, diabetes mellitus with complications, and heart valve disorders were 3 common high-impact diseases having the highest next 6-month resource utilizations and the correlated predicted risks. Such results can be explained by the fact that heart problems, diabetes, and kidney disease are commonly associated with one another and these diseases together with cancer and obesity were reported to comprise the majority of national medical spending every year [33]. Patient populations, grouped by their chronic disease diagnoses, exhibited good linearity between their future resource utilizations and their predicted risks indicating that our model is able to provide each type of chronic disease with a reasonable assessment of future health care expense. In other words, our model can give not only patient-oriented forecasts, but also disease-oriented forecasts of future resource utilization. This feature provides a direct link from our predictive algorithm to health care resource utilization because patients with long-term illness tend to account for a high volume of revisits to either inpatient or ED care settings. Knowing in advance the projected care service usage associated with chronic medical conditions can help providers make more informed decisions on the allocation of care management services with the objective of decreasing unnecessary utilization associated with treating chronically ill patients.

Although HIE data represent an ideal source of communitywide/regional patient data, operational HIEs are not present in all states. Samples collected from HIE might have unexpected bias and not match exactly the nationwide population characteristics. After overcoming these limitations, our predictive model will be improved with a broader applicability in health care globally.

Increasing health care expenditures in the United States have placed significant burdens on the national economy, calling for more cost-effective care strategies. Our study derived an EMR-based prospectively validated model to predict the health care resource utilization in the next 6 months for each of the more than 1 million patients in Maine. This model can assist care providers in applying appropriate care management services aimed at optimizing the resource allocation of caring for high-risk patients. Future studies will focus on integrating payer claims data with the HIE data to get a more accurate and timely prediction of projected future resource utilization. Having this information will assist in budgeting and resource planning in addition to supporting care management programs. The addition of claims data may also improve feature performance in the predictive model as additional or supplemental encounters, medication, diagnosis, and procedure information could be derived from claims data to fill patient longitudinal data gaps in the HIE data. The goal is to utilize the results of this study and future studies to support health care planning, financial management, and public health functions in addition to care management.

Acknowledgments

We thank and express our gratitude to the hospitals, medical practices, physicians, and nurses participating in Maine's HIE. We also thank the biostatistics colleagues at the Department of Health Research and Policy of Stanford University for critical discussions.

Conflicts of Interest

KGS, EW, and XBL are cofounders and equity holders of HBI Solutions, Inc, which is currently developing predictive analytics solutions in health care. The other authors declare that they have no competing interests. From the Departments of Pediatrics, Surgery, and Statistics, Stanford University School of Medicine, Stanford, California, SH, AYS, YW, LZ, KGS, and XBL conducted this research as part of a personal outside-consulting arrangement with HBI Solutions, Inc. The research and research results are not, in any way, associated with Stanford University.

Multimedia Appendix 1

Missing data handling.

[PDF File (Adobe PDF File), 4KB - jmir_v17i9e219_app1.pdf]

Multimedia Appendix 2

Retrospective and prospective cohort construction and inclusion/exclusion criteria.

[PDF File (Adobe PDF File), 90KB - jmir_v17i9e219_app2.pdf]

Multimedia Appendix 3

Relationship between the total next 6-month resource utilization and population size in different bins in the retrospective cohort. Samples were grouped into ten bins based on their next 6-month resource utilization. In each bin, the population size (black bar)

and the resource utilization (white bar) of all the samples were plotted in the form of percentages they took up in the total population or total resource utilization.

[PDF File (Adobe PDF File), 110KB - jmir_v17i9e219_app3.pdf]

Multimedia Appendix 4

Case study of resource utilization patterns.

[PDF File (Adobe PDF File), 85KB - jmir_v17i9e219_app4.pdf]

Multimedia Appendix 5

System implementation of the online population-based risk model for health care resource utilization.

[PDF File (Adobe PDF File), 3KB - jmir_v17i9e219_app5.pdf]

Multimedia Appendix 6

Next 6-month clinical patterns of patients at different risk levels in the prospective cohort.

[PDF File (Adobe PDF File), 5KB - jmir_v17i9e219_app6.pdf]

Multimedia Appendix 7

Online total population resource-utilization risk monitoring dashboard.

[PDF File (Adobe PDF File), 219KB - jmir_v17i9e219_app7.pdf]

References

1. The Unsustainable Cost of Health Care. Washington, DC: Social Security Advisory Board; 2009.
2. National Health Expenditures 2012 Highlights. 2014. URL: https://kaiserhealthnews.files.wordpress.com/2014/04/highlights.pdf [accessed 2015-07-23] [WebCite Cache ID 6aEF7Sj8B]
3. Rau J. NPR. 2013. Health care costs are projected to outpace economic growth URL: http://www.npr.org/sections/health-shots/2013/09/18/223812518/health-care-costs-projected-to-outpace-economic-growth [accessed 2015-07-23] [WebCite Cache ID 6aEFCweDQ]
4. Boccia R, Fraser A, Goff E. The Heritage Foundation. 2013 Aug 20. Federal spending by the numbers, 2013: Government spending trends in graphics, tables, and key points URL: http://www.heritage.org/research/reports/2013/08/federal-spending-by-the-numbers-2013 [accessed 2015-07-23] [WebCite Cache ID 6aEFGf3eM]
5. Dunn L. Becker's Hospital Review. 2013 Dec 18. Predictions for value-based care in 2014 URL: http://www.beckershospitalreview.com/hospital-management-administration/5-predictions-for-value-based-care-in-2014.html [accessed 2015-07-23] [WebCite Cache ID 6aEFLyEEf]
6. Lumeris. 2013 Dec 10. News release: Lumeris releases five predictions for value-based care in 2014 URL: http://lumeris.com/lumeris-news/news-release-lumeris-releases-predictions-value-based-care-2014/ [accessed 2015-07-23] [WebCite Cache ID 6aEFPzNqv]
7. Reinhardt UE, Hussey PS, Anderson GF. Cross-national comparisons of health systems using OECD data, 1999. Health Aff (Millwood) 2002;21(3):169-181 [FREE Full text] [Medline: 12025981]
8. Bertsimas D, Bjarnadóttir MV, Kane MA, Kryder JC, Pandey R, Vempala S, et al. Algorithmic prediction of health-care costs. Operations Research 2008 Dec;56(6):1382-1392. [doi: 10.1287/opre.1080.0619]
9. Virta L, Joranger P, Brox JI, Eriksson R. Costs of shoulder pain and resource use in primary health care: a cost-of-illness study in Sweden. BMC Musculoskelet Disord 2012;13:17 [FREE Full text] [doi: 10.1186/1471-2474-13-17] [Medline: 22325050]
10. Brilleman SL, Purdy S, Salisbury C, Windmeijer F, Gravelle H, Hollinghurst S. Implications of comorbidity for primary care costs in the UK: A retrospective observational study. Br J Gen Pract 2013 Apr;63(609):e274-e282 [FREE Full text] [doi: 10.3399/bjgp13X665242] [Medline: 23540484]
11. Diehr P, Yanez D, Ash A, Hornbrook M, Lin DY. Methods for analyzing health care utilization and costs. Annu Rev Public Health 1999;20:125-144. [doi: 10.1146/annurev.publhealth.20.1.125] [Medline: 10352853]
12. Krop JS, Saudek CD, Weller WE, Powe NR, Shaffer T, Anderson GF. Predicting expenditures for Medicare beneficiaries with diabetes. A prospective cohort study from 1994 to 1996. Diabetes Care 1999 Oct;22(10):1660-1666 [FREE Full text] [Medline: 10526731]
13. Moturu S, Johnson W, Liu H. Predicting future high-cost patients: A real-world risk modeling application. 2007 Presented at: IEEE International Conference on Bioinformatics and Biomedicine, 2007 (BIBM 2007); Nov 2-7, 2007; Silicon Valley, CA. [doi: 10.1109/BIBM.2007.54]

14. Baumeister SE, Dörr M, Radke D, Nauck M, John U, Marschall P, et al. Predictive modeling of health care costs: Do cardiovascular risk markers improve prediction? Eur J Cardiovasc Prev Rehabil 2010 Jun;17(3):355-362. [doi: 10.1097/HJR.0b013e328333a0b7] [Medline: 19907337]
15. Pope GC, Kautter J, Ellis RP, Ash AS, Ayanian JZ, Lezzoni LI, et al. Risk adjustment of Medicare capitation payments using the CMS-HCC model. Health Care Financ Rev 2004;25(4):119-141 [FREE Full text] [Medline: 15493448]
16. Ash AS, Ellis RP. Risk-adjusted payment and performance assessment for primary care. Med Care 2012 Aug;50(8):643-653 [FREE Full text] [doi: 10.1097/MLR.0b013e3182549c74] [Medline: 22525609]
17. Ash AS, Ellis RP, Pope GC, Ayanian JZ, Bates DW, Burstin H, et al. Using diagnoses to describe populations and predict costs. Health Care Financ Rev 2000;21(3):7-28 [FREE Full text] [Medline: 11481769]
18. Fleishman JA, Cohen JW. Using information on clinical conditions to predict high-cost patients. Health Serv Res 2010 Apr;45(2):532-552 [FREE Full text] [doi: 10.1111/j.1475-6773.2009.01080.x] [Medline: 20132341]
19. Mittmann N, Liu N, Porter J, Seung SJ, Isogai PK, Saskin R, et al. Utilization and costs of home care for patients with colorectal cancer: a population-based study. CMAJ Open 2014 Jan;2(1):E11-E17 [FREE Full text] [doi: 10.9778/cmajo.20130026] [Medline: 25077120]
20. Duncan I. Predictive Modelling: Basics and Beyond. Farmington, CT: Solucia Consulting; 2009 Jun. URL: http://www.ehcca.com/presentations/predmodel4/duncan_pc1.pdf [accessed 2015-07-23] [WebCite Cache ID 6aEFTMJfI]
21. Machlin S, Chowdhury S. Medical Expenditure Panel Survey: Statistical Brief #318. Rockville, MD: Agency for Healthcare Research and Quality; 2011 May. Expenses and characteristics of physician visits in different ambulatory care settings, 2008 URL: http://meps.ahrq.gov/mepsweb/data_files/publications/st318/stat318.pdf [accessed 2015-07-23] [WebCite Cache ID 6aEFVnPbX]
22. Pfuntner A, Wier L, Steiner C. Healthcare Cost and Utilization Project: Statistical Brief #146. Rockville, MD: Agency for Healthcare Research and Quality; 2013 Jan. Costs for hospital stays in the United States, 2010 URL: http://www.hcup-us.ahrq.gov/reports/statbriefs/sb146.pdf [accessed 2015-07-23] [WebCite Cache ID 6aEFYefCo]
23. Breiman L. Bagging predictors. Mach Learn 1996;24(2):123-140. [doi: 10.1023/A:1018054314350]
24. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. Random forest: a classification and regression tool for compound classification and QSAR modeling. J Chem Inf Comput Sci 2003;43(6):1947-1958. [doi: 10.1021/ci034160g] [Medline: 14632445]
25. Breiman L. Random forests. Mach Learn 2001:5-32. [doi: 10.1023/A:1010933404324]
26. Healthcare Cost and Utilization Project. Chronic Condition Indicator (CCI) for ICD-9-CM URL: http://www.hcup-us.ahrq.gov/toolssoftware/chronic/chronic.jsp [accessed 2015-07-23] [WebCite Cache ID 6aEFbKuru]
27. Keehan SP, Lazenby HC, Zezza MA, Catlin AC. Age estimates in the national health accounts. Health Care Financ Rev 2004;26(2):1-16 [FREE Full text] [Medline: 25372990]
28. Bodenheimer T, Chen E, Bennett HD. Confronting the growing burden of chronic disease: can the US health care workforce do the job? Health Aff (Millwood) 2009;28(1):64-74 [FREE Full text] [doi: 10.1377/hlthaff.28.1.64] [Medline: 19124856]
29. Druss BG, Marcus SC, Olfson M, Tanielian T, Elinson L, Pincus HA. Comparing the national economic burden of five chronic conditions. Health Aff (Millwood) 2001;20(6):233-241 [FREE Full text] [Medline: 11816664]
30. Hao S, Jin B, Shin AY, Zhao Y, Zhu C, Li Z, et al. Risk prediction of emergency department revisit 30 days post discharge: A prospective study. PLoS One 2014;9(11):e112944 [FREE Full text] [doi: 10.1371/journal.pone.0112944] [Medline: 25393305]
31. Hu Z, Jin B, Shin AY, Zhu C, Zhao Y, Hao S, et al. Real-time web-based assessment of total population risk of future emergency department utilization: statewide prospective active case finding study. Interact J Med Res 2015;4(1):e2 [FREE Full text] [doi: 10.2196/ijmr.4022] [Medline: 25586600]
32. Wilson J, Adam B. The benefit of using both claims data and electronic medical record data in health care analysis. Eden Prairie, MN: Optum; 2012. URL: https://www.optum.com/content/dam/optum/resources/whitePapers/Benefits-of-using-both-claims-and-EMR-data-in-HC-analysis-WhitePaper-ACS.pdf [accessed 2015-07-23] [WebCite Cache ID 6aEFeBaso]
33. Stinson S. Bankrate. 2013. The cost of chronic disease URL: http://www.bankrate.com/finance/insurance/cost-of-chronic-disease-1.aspx [accessed 2015-07-23] [WebCite Cache ID 6aEFgNfXZ]

Abbreviations

AHRQ: Agency for Healthcare Research and Quality
CCS: Clinical Classifications Software
ED: emergency department
EMR: electronic medical record
HIE: health information exchange
HIN: HealthInfoNet
ICD-9-CM: International Classification of Diseases, 9th Revision, Clinical Modification

Edited by G Eysenbach; submitted 24.07.15; peer-reviewed by G Stiglic, S Hollinghurst; comments to author 12.08.15; accepted 02.09.15; published 22.09.15.

Please cite as:
Hu Z, Hao S, Jin B, Shin AY, Zhu C, Huang M, Wang Y, Zheng L, Dai D, Culver DS, Alfreds ST, Rogow T, Stearns F, Sylvester KG, Widen E, Ling X
Online Prediction of Health Care Utilization in the Next Six Months Based on Electronic Health Record Information: A Cohort and Validation Study
J Med Internet Res 2015;17(9):e219
URL: http://www.jmir.org/2015/9/e219/
doi:10.2196/jmir.4976
PMID:26395541

Original Paper

Swab2know: An HIV-Testing Strategy Using Oral Fluid Samples and Online Communication of Test Results for Men Who Have Sex With Men in Belgium

Tom Platteau[1], MSc; Katrien Fransen[1], MSc; Ludwig Apers[1], MD, MPH, PhD; Chris Kenyon[1], MD, PhD; Laura Albers[1], BN; Tine Vermoesen[1], BSc; Jasna Loos[2], MA; Eric Florence[1], MD, PhD

[1]Institute of Tropical Medicine, Department of Clinical Sciences, Antwerp, Belgium

[2]Institute of Tropical Medicine, Department of Public Health, Antwerp, Belgium

Corresponding Author:
Tom Platteau, MSc
Institute of Tropical Medicine
Department of Clinical Sciences
Nationalestraat 155
Antwerp, 2000
Belgium
Phone: 32 3 2476433
Fax: 32 3 2476410
Email: tplatteau@itg.be

Abstract

Background: As HIV remains a public health concern, increased testing among those at risk for HIV acquisition is important. Men who have sex with men (MSM) are the most important group for targeted HIV testing in Europe. Several new strategies have been developed and implemented to increase HIV-testing uptake in this group, among them the Swab2know project.

Objective: In this project, we aim to assess the acceptability and feasibility of outreach and online HIV testing using oral fluid samples as well as Web-based delivery of test results.

Methods: Sample collection happened between December 2012 and April 2014 via outreach and online sampling among MSM. Test results were communicated through a secured website. HIV tests were executed in the laboratory. Each reactive sample needed to be confirmed using state-of-the-art confirmation procedures on a blood sample. Close follow-up of participants who did not pick up their results, and those with reactive results, was included in the protocol. Participants were asked to provide feedback on the methodology using a short survey.

Results: During 17 months, 1071 tests were conducted on samples collected from 898 men. Over half of the samples (553/1071, 51.63%) were collected during 23 outreach sessions. During an 8-month period, 430 samples out of 1071 (40.15%) were collected from online sampling. Additionally, 88 samples out of 1071 (8.22%) were collected by two partner organizations during face-to-face consultations with MSM and male sex workers. Results of 983 out of 1071 tests (91.78%) had been collected from the website. The pickup rate was higher among participants who ordered their kit online (421/430, 97.9%) compared to those participating during outreach activities (559/641, 87.2%; $P<.001$). MSM participating during outreach activities versus online participants were more likely to have never been tested before (17.3% vs 10.0%; $P=.001$) and reported more sexual partners in the 6 months prior to participation in the project (mean 7.18 vs 3.23; $P<.001$). A total of 20 participants out of 898 (2.2%) were confirmed HIV positive and were linked to care. Out of 1071 tests, 28 (2.61%) with a weak reactive result could not be confirmed, and were thereby classified as false reactive results. Most of the 388 participants who completed posttest surveys (388/983, 39.5%) were very positive about their experience. The vast majority (371/388, 95.6%) were very satisfied, while 17 out of 388 (4.4%) reported mixed feelings.

Conclusions: Despite a high yield and a considerable number of false reactive results, satisfaction was high among participants. The project helped us to reach the target population, both in numbers of tests executed and in newly diagnosed HIV infections. Further optimization should be considered in the accuracy of the test, the functionalities of the website (including an online counseling tool), and in studying the cost effectiveness of the methodology.

(J Med Internet Res 2015;17(9):e213) doi:10.2196/jmir.4384

KEYWORDS

HIV; men who have sex with men; MSM; self-sampling; oral fluid; online testing

Introduction

HIV remains an important public health problem. In the European Union, 29,157 new HIV infections were reported in 2013, an incidence of 5.7 per 100,000 inhabitants [1]. A total of 42% of new infections were among men who have sex with men (MSM). Countries with the highest incidence were Estonia (24.6 per 100,000 inhabitants), Latvia (16.8), Portugal (10.4), and Belgium (10.0) [1].

Promoting HIV testing is an integral part of the 90-90-90 Joint United Nations Programme on HIV/AIDS (UNAIDS) plan to end the AIDS epidemic by 2030. In terms of this plan, 90% of all people living with HIV should know their HIV status, 90% should be on treatment, and 90% of these should be virologically suppressed [2]. Part of the rationale for this strategy is that intensified HIV testing contributes to earlier commencement of antiretroviral therapy (ART) which in turn leads to reduced HIV transmission via reducing the HIV viral load [3]. HIV diagnosis also leads to behavioral changes in sexual risk taking in a majority of newly diagnosed persons [4].

Increased HIV testing among those at risk is a key way of achieving the required target of 90% of HIV-infected people knowing their HIV status. The traditional HIV test is offered voluntarily and confidentially by a medically trained health care professional in a health care setting with a strong emphasis on the patient's informed consent [5]. Counseling and test results are provided by trained health care workers during a face-to-face consultation [6]. This approach may remain the standard for most people. However, HIV has reached endemic proportions among MSM in the industrial world, with incidence rates of 2 to 3% per year, and prevalence between 10 and 30% [7]. These men are generally well informed about HIV [8] and in certain groups of MSM, pretest counseling was found to be "repetitive" and "unnecessary" [9]. For these men, alternative HIV testing strategies can be considered. One study found that oral fluid testing is preferred by MSM above giving blood samples [10]. It has also recently been found to be reliable for diagnostic use in groups with an HIV prevalence over 1% [11] like MSM, and several research projects in clinical settings have shown promising results for HIV tests on oral fluid samples [12,13].

Rapid HIV tests, decentralized HIV testing (ie, outreach and community-based testing), and self-testing are additional alternatives. Rapid tests are used in a variety of settings, including primary health care settings [14,15], emergency departments [16,17], and in dental clinics [18]. Their advantage is that clients receive their results at the time of their visit [19]. A major disadvantage is that in low-HIV-prevalence settings they give a relatively high proportion of false positive results [13]. Outreach testing targeting MSM has been implemented in clubs, bars, and bath houses [20-22], as well as at large-scale events, such as Gay Pride festivals [23]. Community-based testing among MSM is increasing in recent years [24,25]. In Europe, an increasing number of community-based testing centers (ie, Checkpoints) have been established [26,27]. Self-tests for HIV that can be ordered through the Internet are the most recent development in this field [28]. The only US Food and Drug Administration (FDA)-licensed oral fluid-based rapid test is OraQuick ADVANCE Rapid HIV-1/2 Antibody test [29]. The FDA approved the use of this test for home use in July 2012 [30], despite its relatively high risk of false positive results [31], especially among lower-risk groups [13]. The quality of other test kits that can be ordered online is largely unknown due to their lack of certification. The advantages of self-testing include increased convenience and heightened privacy [28]. The difficulty of ensuring linkage to care in the case of a positive result is a weakness of these tests. In a recent review, supervised and unsupervised self-testing strategies were found to be highly acceptable and preferred, but all studies lacked an evaluation of posttest linkage to counseling and care [32]. Internet-based testing can therefore be an alternative, where posttest linkage is part of the process. The willingness to use Internet-based HIV-testing strategies was high in recently published quantitative and qualitative studies [33,34]. To address MSM's testing needs, we developed the Swab2know project [35]. This project combines two strategies to increase HIV-testing uptake among MSM: outreach HIV-test sessions and free online testing. In both strategies, samples are collected using oral fluid collection devices and test results are communicated via secured website.

This nonrandomized, prospective descriptive study aimed at detecting new HIV cases among groups at risk for HIV/sexually transmitted infection (STI) acquisition. The secondary objective was to assess the acceptability and feasibility of an HIV-testing strategy with the use of self-administered oral fluid samples collected through outreach and online activities and Web-based delivery of test results.

Methods

Population and Settings

The project targeted two main risk groups for HIV infection in Belgium: MSM and sub-Saharan African migrants (SAM) socializing in community venues. Only the data from MSM were used for this analysis; the findings for SAM will be published elsewhere.

Outreach sessions for MSM were organized in five types of venues, mostly situated in Antwerp: saunas/bathhouses, fetish scene venues, dancing/discotheque venues, outreach sessions organized during the World Outgames in Antwerp targeting athletes and supporters, and large-scale gay events.

Inclusion criteria were that participants had to identify themselves as MSM and be 18 years of age or older. Additional criteria were that participants had to be accepting of oral fluid sampling, sign informed consent forms, provide minimal information, and understand that the test, if positive, would only be strongly indicative of HIV infection. Participants who were not willing to provide a mobile phone number or email address, were under 18 years of age, or were unable to sign the informed

consent form were excluded from this project and redirected to standard testing facilities.

Website

A website, swab2know.be, has been specifically designed for this project [35]. The main aim of the website is to provide a platform where visitors can find information, prevention messages, order test kits, and collect their test results. This website is secured by means of the Secure Sockets Layer protocol, and holds a security certificate provided by Belnet—Belnet is the federal government organization that provides high-bandwidth Internet connection and services to Belgian universities, research centers, and government departments. The certificate confirms the identity of, and encrypts the communication between, the Swab2know Web server and the computer where the information is requested.

All online materials described in the Methods section are available through the study website [35].

Sampling Procedures

Outreach sessions took place in various MSM leisure venues. Field workers of Sensoa, a local prevention organization, announced the Swab2know session at the entrance. If clients decided to participate, they visited the Swab2know team in a separate room. All materials were available in Dutch, English, and French. After being informed and signing the informed consent (IC) form, baseline data using a self-administered pen-and-paper questionnaire (see Multimedia Appendix 1) were gathered from participants and an online account was created. Each account was unique and linked with an email address and phone number. A generated password was sent automatically to the participant's email address. The oral fluid samples were self-collected by the participants under the supervision of study staff. All samples were identified by a unique sample code, which linked the sample with the personal account, the IC, and baseline data. Samples were kept at room temperature and were brought to the laboratory on the next working day.

Online recruitment happened on the website by occasional visitors who created an account and provided their email address and phone number. The project was advertised by prevention organizations and through articles and announcements in dedicated media, including gay-oriented websites and magazines and a Swab2know Facebook account. Participants provided consent by accepting the terms of the study. A sampling kit identified with a unique sample code was sent to the Belgian address of their choice. Participants took the oral fluid sample after having seen a short educational video on the website. Samples were sent to the lab with a prepaid envelope. The participants could also opt to collect their results during a face-to-face consultation.

HIV Test

The validation of the accuracy of the test has previously been evaluated in our AIDS Reference Laboratory at the Institute of Tropical Medicine (ITM) [10]. Each sample underwent a two-step HIV-test procedure. First, all samples were tested for HIV using Genscreen HIV-1/2 Version 2 by Bio-Rad (Marnes-la-Coquette, France) [36]. The results were classified as strong, weak, or nonreactive. In a second step, all nonreactive samples were checked for sample quality using a human IgG detection test. The quality of the oral fluid samples was measured using the IgG enzyme-linked immunosorbent assay (ELISA) quantification kit (Human IgG ELISA Immunology Consultants Laboratory, Inc, Portland, OR, USA). If the sample contained more than 3500 ng total IgG/mL, the nonreactive result was considered valid and ready to be communicated. Prior to uploading them onto the website, each result of the HIV test performed was technically validated by two persons, individually.

Communication of the HIV Test Results

Once the results of the HIV tests were known in the laboratory, they were uploaded onto the website. Upon uploading, participants received an email indicating that their result was available. Participants received one of four standardized messages (for full messages, see Multimedia Appendices 2-5): (1) a strong reactive test result, strongly indicating HIV infection, to be confirmed by a blood sample, (2) a weak reactive result, indicating a probable false positive result or an early infection, to be confirmed by a blood sample, (3) a nonreactive result, indicating the absence of HIV infection, taking into account a window period of 3 months, and (4) an invalid result, with the suggestion to repeat the oral fluid sample or to take a state-of-the-art HIV test. In the case of a reactive result, a mobile phone number was provided for emergency counseling by a trained paramedic.

Participants who did not check their results were contacted by phone or email. All participants with a reactive result were contacted by phone within 24 hours of having read their results. The purpose of the call was to offer counseling and to arrange a further confirmation test and guarantee linkage to care. If confirmation did not take place at the organizing health care center, participants were contacted after the confirmation procedure to collect the final result.

Repeated Testing

Participants with a nonreactive test result, both through outreach and online participation, were offered the possibility to order a sampling package to be delivered to a Belgian address every 4 to 6 months, allowing frequent and repeated testing. For this purpose, a reminder email was sent 4 to 6 months after participation to the email address linked with the personal account.

Acceptability of the Methodology

In the delivery message of the test result, participants were asked to fill out an online self-administered survey. In this survey, participants provided information on their impression of the project as a whole (not good, mixed feelings, good) and whether they would participate again in the future (no, not sure, without hesitation).

Statistical Analysis

Statistical analysis was performed using IBM SPSS version 22. Descriptive and univariate analyses were carried out. Chi-square tests were used for categorical variables and independent

samples *t* tests for continuous variables. A significance level of 5% was applied.

Ethical Considerations

The methodology was conceptualized in close collaboration with community-based prevention organizations targeting MSM and African communities.

Ethical approval was obtained from the Institutional Review Board at the ITM and the University Hospital in Antwerp.

Results

Number of Performed Tests

In a period of 17 months (December 2012-April 2014), 1082 tests were executed on samples collected through outreach activities and online ordering of sampling kits. A total of 11 tests were excluded from the analysis because the participants disclosed their HIV-positive status during the baseline survey. Data from 1071 tests from 898 participants were used for this analysis. A total of 4 persons participated 4 times, 16 participated 3 times, and 129 participated twice in this period. A total of 53 persons participated during outreach activities and ordered a sampling kit later in the project.

During 23 outreach sessions, 553 out of a total 1071 (51.63%) samples were collected. These sessions were organized in saunas/bathhouses (5 sessions), fetish scene venues (4 sessions), dancing/discotheque venues (8 sessions), during the World Outgames in Antwerp targeting athletes and supporters (3 sessions), and at other gay events (3 sessions). Additionally, 88 samples out of 1071 (8.22%) were collected by two partner organizations who used the project's methodology to facilitate HIV testing within their regular activities during face-to-face consultations with MSM and male sex workers. For the analysis, these samples were added to the *outreach* group. These 641 samples out of 1071 (59.85%) were collected from 609 men. From September 1, 2013, we offered people the possibility of ordering a sampling kit through the website. In the subsequent 8-month period, 430 samples out of 1071 (40.15%) were ordered online by 289 participants.

Participant Characteristics

A description of the population with a comparison between the outreach and the online population is presented in Table 1.

Table 1. Characteristics of study participants.

Characteristic	Online participants (n=430), mean (SD) or n (%)	Outreach participants (n=641), mean (SD) or n (%)	*P*
Age in years, mean (SD)	34.3 (10.2)	33.5 (11.4)	.25
Sexual partners, n (%)			.07
Men	397 (92.3)	524 (81.7)	
Men and women	30 (7.0)	52 (8.1)	
Women[a]	3 (0.7)	0 (0)	
Has a general practitioner/family doctor, n (%)	389 (90.5)	558 (87.1)	.10
Never tested for HIV, n (%)	43 (10.0)	111 (17.3)	.001
Number of sexual partners in past 3 months, mean (SD)	3.23 (5.02)	7.18 (13.53)	<.001

[a]All participants reporting sexual contacts with women answered "transgender" on the question for gender.

Results Communication

The vast majority (1057/1071, 98.69%) of test results were delivered through the website. A total of 14 out of 1071 results (1.31%) were delivered either by phone (mainly because the participants did not have access to email or Internet) (8/1071, 0.75%) or during a face-to-face consultation (6/1071, 0.56%). All results that were not communicated through the website stemmed from samples collected through outreach activities.

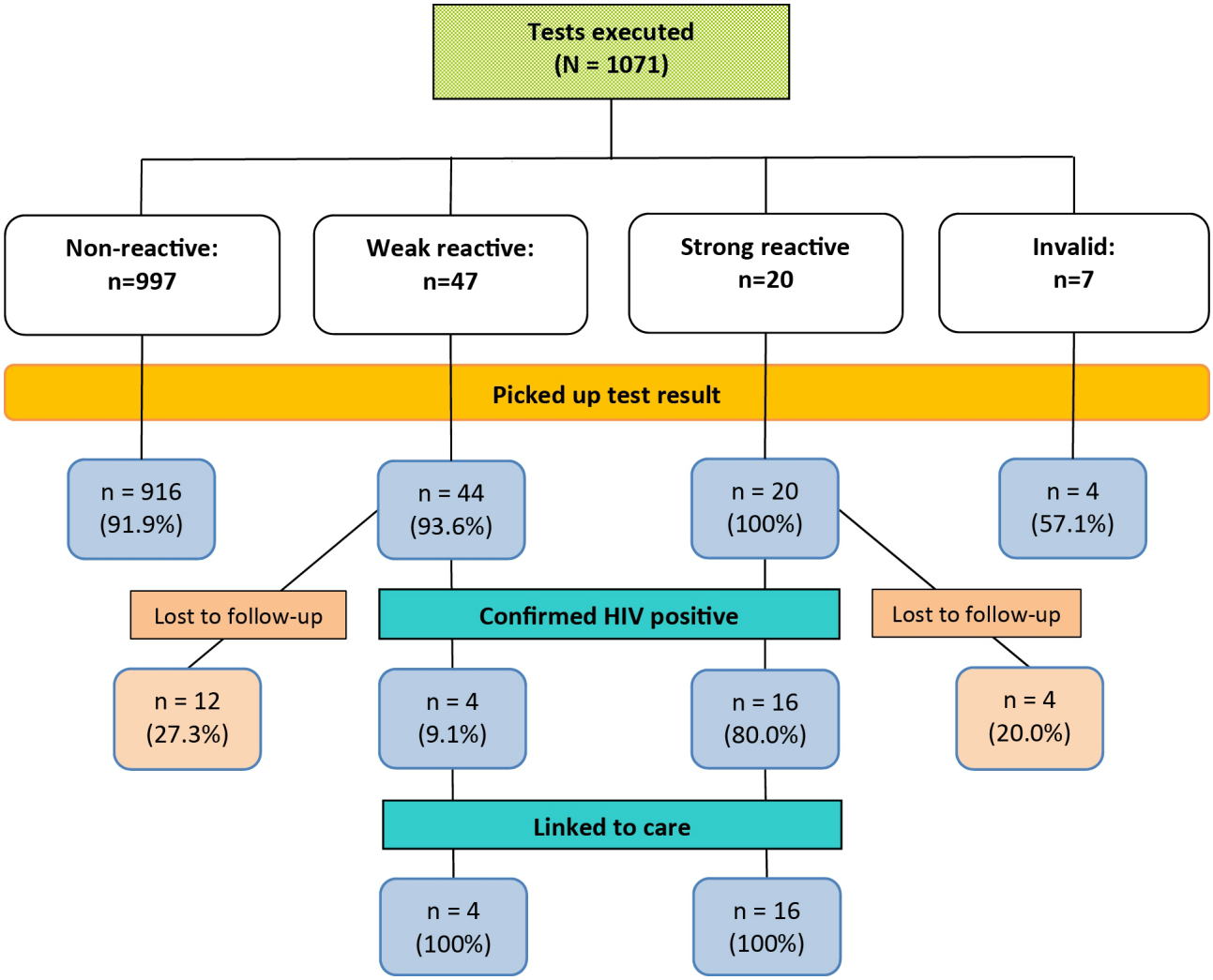
Overall, the results of 983 out of 1071 (91.78%) tests were effectively collected from the website. The pickup rate was significantly higher when the test had been ordered online (421/430, 97.9%) compared to the test performed during outreach activities (559/641, 87.2%; *P*<.001).

Figure 1 shows the number of nonreactive, weak reactive, and strong reactive test results, with confirmation test results from blood and linkage to care.

A total of 28 out of 44 (64%) weak reactive test results were not confirmed HIV positive, and were thereby classified as *false reactive* results.

Overall, 20 participants were confirmed as newly diagnosed with HIV and all of them were linked to care; this represents 2.2% (20/898) of all participants tested. A total of 6 newly diagnosed participants ordered their sampling kit online which put the new HIV infection rate in this group at 2.1% (6/289) while 14 were detected during outreach sessions; the new HIV infection rate for this group was 2.3% (14/609). This difference was not statistically significant (*P*=.83).

Figure 1. Flowchart of test results, confirmation, and follow-up for 1071 tests executed among MSM in the Swab2know study.



Acceptability

Of the 983 participants who collected their test results, 388 (39.5%) completed posttest surveys. The vast majority of participants (371/388, 95.6%) reported being very satisfied with the process while 17 out of 388 (4.4%) experienced mixed feelings taking part in the project. Whereas 48 out of 388 participants (12.4%) reported they would consider taking part again, most of the respondents (337/388, 86.9%) reported they would do so "without hesitation." Of the 388 respondents, 3 (0.8%) reported that they would not participate again in the future. One of them had been diagnosed with HIV through the project, which makes future participation redundant.

Discussion

Principal Findings

Results from recent surveys show that HIV self-testing is gaining momentum within the MSM community across the world [34,37]. This manuscript describes a low-threshold HIV-testing strategy combining oral fluid self-sampling, HIV testing in a specialized AIDS reference laboratory, and result delivery through a secured website with a solid linkage-to-care strategy.

Compared to the bulk of research on HIV self-testing and home-based testing, little has been published on the combination of self-sampling and the remote delivery of the test result. A recent project in the United Kingdom used a similar methodology to our project, but they performed tests on dried blood samples. Negative results were disclosed by text message while positive results were communicated by phone. They found a comparable rate of newly diagnosed HIV infections, successful linkage to care, and participant satisfaction [38].

We specifically chose oral fluid collection devices given their potential advantages: convenient and painless to collect, ideal for self-sampling, and very little risk of contamination during collection and sample transport. The potential problems associated with oral fluid testing are a lower sensitivity especially in detecting early infections and the fact that one cannot perform a confirmation test on the same sample because this needs to be done on a blood sample.

We also decided not to use a rapid test during outreach sessions despite good results described in similar settings [21,22]. This choice was motivated by two reasons. First, when a session is organized in leisure venues, the idea of receiving an HIV diagnosis on the spot could prevent clients from participating. Second, a reactive result requires thorough counseling and support, which are hard to deliver in these types of venues,

especially when other participants are waiting to be tested. Our alternative strategy was to deliver the results via an online tool.

Acceptability and intended uptake of Internet-based HIV/STI-screening programs are high among high-risk groups in various settings [34,39]. Nevertheless using a website to communicate HIV test results has, to the best of our knowledge, not been reported before. It has several advantages over other communication strategies used in similar projects. The participant, not the health care provider, decides when to pick up the result. It is less time consuming and less intrusive than a phone call. It offers the possibility to provide information in addition to the HIV test result, such as information on the test window period, the importance of testing for other STIs, and the need to confirm the result in case of a reactive test. Compared to text messages, website communication opens up possibilities to develop automated counseling strategies tailored to the test results and the patient's profile in the future.

The self-sampling procedure produced samples of acceptable quality. A small minority of participants (all through online testing) provided an invalid sample. Satisfaction with the project was very high among participants; however, given the incomplete response rate to the satisfaction survey, we cannot exclude the possibility of a selection bias.

The project helped us to reach the target population, both in terms of the number of tests executed, and in the number of newly diagnosed HIV infections. The percentage of newly diagnosed participants (2.2%) was higher than expected. As recently reported, 6% of Belgian MSM tested in a variety of nightlife settings are HIV positive [40], of whom 14.3% are unaware of their HIV-positive status. Applying these figures, we expected to diagnose 9 new HIV infections (14.3% of 6% of 1071 tests) compared to the 20 new HIV diagnoses in this project. This may be an indication that we succeeded in attracting the population at highest risk of acquiring HIV. All new cases were successfully linked to HIV care, which is a crucial aspect of the HIV treatment cascade, and a great asset of the project compared to self-testing. Moreover, with a yield of 2.2% of participants (20/898) newly diagnosed with HIV in this project, its method can be considered as cost effective—HIV testing in populations where the prevalence is greater than 0.1% is considered cost effective [41,42].

An additional benefit of the project was that the partners from 3 participants were newly diagnosed in the organizing health center during the course of the project. They were not included in this analysis.

The proportion of participants who were never tested before was considerable. Of 898 participants, 154 (17.1%) answered that they had never been tested for HIV before. This percentage indicates that online and outreach testing may facilitate HIV testing for MSM who experience difficulties in taking a test using the existing structures, and therefore do not get tested.

Limitations

As observed previously with oral fluid testing protocols, the proportion of *false reactive* test results was substantial (2.7%) [13]. Taking into account the impact of a reactive result on people's lives, these *false reactive* results should be minimized. Despite this problem, 4 of 6 participants with a *false reactive* result who provided feedback using the acceptability questionnaire answered that they were "very satisfied" with the project and would "without hesitation" participate again. The other 2 participants reported "mixed feelings" about the project: one said he would not and one said he would consider participating again. It remains crucial that participants with a weak reactive result see a physician to confirm or refute the result. A minority of participants with a weak reactive result were confirmed HIV positive (4/32, 13%).

A considerable number of participants were lost to follow-up in the course of the project. Whereas a loss to follow-up does not mean that participants were not linked to care (some may have visited their general practitioner), we should aim to minimize this proportion.

The yield of this screening project was high; however, contacting and motivating participants to pick up their results required more intensive follow-up than expected. Although the workload for the paramedical staff was much less than with the standard-of-care counseling method, this aspect should not be underestimated in such projects. Further studies should investigate whether such strategies are cost effective in detecting new HIV infections in high-risk groups.

Next Steps

We plan to continue the project in the coming years, with an increased emphasis on Internet-based testing and repeated testing for participants, as well as strong collaboration with community-based and prevention organizations to guide MSM toward the Swab2know project. On the basis of this experience, our methodology will be refined. First, the online counseling tool will be further developed and refined to support participants, with an increased emphasis on those with a reactive result. Comparable e-counseling tools have been developed and implemented in primary care [43]. This could complement, and to some extent replace, the phone counseling, thereby reducing the staff workload. Second, we hope to reduce the number of false reactive results by the use of newly developed point-of-care oral fluid tests. Third, expanding the scope of sexually transmitted infections tests may improve the attraction of the project among MSM. One could consider performing syphilis or hepatitis C serology on oral fluid samples, or nucleic acid tests for the detection of chlamydia and gonorrhea on self-collected urethral and anal samples [44], allowing a comprehensive STI checkup.

From a societal point of view, a legal framework needs to be developed. Self-testing is not officially recognized in Belgium: neither are online testing nor self-sampling activities.

In conclusion, we demonstrated that a low-threshold HIV-testing strategy combining self-sampling with oral fluid and online result delivery was acceptable. The HIV infection rate was higher than expected and the linkage to care was good. This strategy empowers individuals to manage their health, but at this stage it should be reserved for high-risk groups such as MSM where the incidence of HIV is high.

Acknowledgments

We would like to thank all participants and all collaborators and volunteers who joined the team during outreach activities. We want to thank all colleagues from Sensoa, Boysproject, and çavaria for their participation in the field work. We would like to thank Vera Van Boxel for database development and assistance. We would also like to thank our webmaster, Nico Van Aerde, as well as Marcel Wauters and his colleagues from All Directions for developing the website. We want to thank Inch Magazine for providing free advertisement space.

Finally, we want to thank the city of Antwerp for funding the project via a grant. Their input was only financial; there was no involvement in the content of the project, nor in the review or approval of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Baseline questionnaire filled in by all participants.

[JPG File, 680KB - jmir_v17i9e213_app1.jpg]

Multimedia Appendix 2

Full message that was communicated in the case of a strong reactive test result.

[JPG File, 132KB - jmir_v17i9e213_app2.jpg]

Multimedia Appendix 3

Full message that was communicated in the case of a weak reactive test result.

[JPG File, 114KB - jmir_v17i9e213_app3.jpg]

Multimedia Appendix 4

Full message that was communicated in the case of a nonreactive test result.

[JPG File, 125KB - jmir_v17i9e213_app4.jpg]

Multimedia Appendix 5

Full message that was communicated in the case of an invalid test result.

[JPG File, 76KB - jmir_v17i9e213_app5.jpg]

References

1. European Centre for Disease PreventionControl, WHO Regional Office for Europe. HIV/AIDS Surveillance in Europe 2013. Stockholm, Sweden: European Centre for Disease Prevention and Control; 2014. URL: http://ecdc.europa.eu/en/publications/Publications/hiv-aids-surveillance-report-Europe-2013.pdf [accessed 2015-08-27] [WebCite Cache ID 6b6RmWdnG]
2. 90-90-90: An Ambitious Treatment Target to Help End the AIDS Epidemic. Geneva, Switzerland: Joint United Nations Programme on HIV/AIDS (UNAIDS); 2014. URL: http://www.unaids.org/sites/default/files/media_asset/90-90-90_en_0.pdf [accessed 2015-08-27] [WebCite Cache ID 6b6S4dt3P]
3. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, HPTN 052 Study Team. Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med 2011 Aug 11;365(6):493-505 [FREE Full text] [doi: 10.1056/NEJMoa1105243] [Medline: 21767103]
4. Fox J, White PJ, Macdonald N, Weber J, McClure M, Fidler S, et al. Reductions in HIV transmission risk behaviour following diagnosis of primary HIV infection: a cohort of high-risk men who have sex with men. HIV Med 2009 Aug;10(7):432-438. [doi: 10.1111/j.1468-1293.2009.00708.x] [Medline: 19459996]
5. British HIV Association, British Association of Sexual Health and HIV, British Infection Society. UK National Guidelines for HIV Testing 2008. London, UK: British HIV Association; 2008 Sep. URL: http://www.bhiva.org/documents/Guidelines/Testing/GlinesHIVTest08.pdf [accessed 2015-02-06] [WebCite Cache ID 6W8k8Fmh2]

6. WHO Regional Office for Europe. Scaling Up HIV Testing and Counseling in the WHO European Region: As An Essential Component of Efforts to Achieve Universal Access to HIV Prevention, Treatment, Care and Support. Copenhagen, Denmark: World Health Organization Regional Office for Europe; 2010. URL: http://www.euro.who.int/__data/assets/pdf_file/0007/85489/E93715.pdf?ua=1 [accessed 2015-02-06] [WebCite Cache ID 6W8kFbQHP]
7. De Cock KM, Jaffe HW, Curran JW. The evolving epidemiology of HIV/AIDS. AIDS 2012 Jun 19;26(10):1205-1213. [doi: 10.1097/QAD.0b013e328354622a] [Medline: 22706007]
8. Lee S, Sheon N. Responsibility and risk: accounts of reasons for seeking an HIV test. Sociol Health Illn 2008 Mar;30(2):167-181. [doi: 10.1111/j.1467-9566.2007.01038.x] [Medline: 18290930]
9. Spielberg F, Kurth A, Gorbach PM, Goldbaum G. Moving from apprehension to action: HIV counseling and testing preferences in three at-risk populations. AIDS Educ Prev 2001 Dec;13(6):524-540. [Medline: 11791784]
10. Chen MY, Bilardi JE, Lee D, Cummings R, Bush M, Fairley CK. Australian men who have sex with men prefer rapid oral HIV testing over conventional blood testing for HIV. Int J STD AIDS 2010 Jun;21(6):428-430. [doi: 10.1258/ijsa.2010.009552] [Medline: 20606224]
11. Fransen K, Vermoesen T, Beelaert G, Menten J, Hutse V, Wouters K, et al. Using conventional HIV tests on oral fluid. J Virol Methods 2013 Dec;194(1-2):46-51. [doi: 10.1016/j.jviromet.2013.08.004] [Medline: 23969313]
12. Debattista J, Bryson G, Roudenko N, Dwyer J, Kelly M, Hogan P, et al. Pilot of non-invasive (oral fluid) testing for HIV within a clinical setting. Sex Health 2007 Jun;4(2):105-109. [Medline: 17524288]
13. Pant PN, Balram B, Shivkumar S, Martinez-Cajas JL, Claessens C, Lambert G, et al. Head-to-head comparison of accuracy of a rapid point-of-care HIV test with oral versus whole-blood specimens: a systematic review and meta-analysis. Lancet Infect Dis 2012 May;12(5):373-380. [doi: 10.1016/S1473-3099(11)70368-1] [Medline: 22277215]
14. Gennotte A, Semaille P, Ellis C, Necsoi C, Abdulatif M, Chellum N, et al. Feasibility and acceptability of HIV screening through the use of rapid tests by general practitioners in a Brussels area with a substantial African community. HIV Med 2013 Oct;14 Suppl 3:57-60. [doi: 10.1111/hiv.12061] [Medline: 24033907]
15. Gauthier R, Livrozet JM, Prevoteau du Clary F, Taulera O, Bouée S, Aubert JP, et al. Feasibility and acceptability of rapid HIV test screening (DEPIVIH) by French family physicians. Med Mal Infect 2012 Nov;42(11):553-560. [doi: 10.1016/j.medmal.2012.08.005] [Medline: 23026269]
16. d'Almeida KW, Kierzek G, de Truchis P, Le Vu S, Pateron D, Renaud B, Emergency Department HIV-Screening Group. Modest public health impact of nontargeted human immunodeficiency virus screening in 29 emergency departments. Arch Intern Med 2012 Jan 9;172(1):12-20. [doi: 10.1001/archinternmed.2011.535] [Medline: 22025095]
17. Christopoulos KA, Schackman BR, Lee G, Green RA, Morrison EA. Results from a New York City emergency department rapid HIV testing program. J Acquir Immune Defic Syndr 2010 Mar;53(3):420-422 [FREE Full text] [doi: 10.1097/QAI.0b013e3181b7220f] [Medline: 20190591]
18. Hutchinson MK, VanDevanter N, Phelan J, Malamud D, Vernillo A, Combellick J, et al. Feasibility of implementing rapid oral fluid HIV testing in an urban University Dental Clinic: a qualitative study. BMC Oral Health 2012;12:11 [FREE Full text] [doi: 10.1186/1472-6831-12-11] [Medline: 22571324]
19. Hutchinson AB, Branson BM, Kim A, Farnham PG. A meta-analysis of the effectiveness of alternative HIV counseling and testing methods to increase knowledge of HIV status. AIDS 2006 Aug 1;20(12):1597-1604. [doi: 10.1097/01.aids.0000238405.93249.16] [Medline: 16868440]
20. Platteau T, Wouters K, Apers L, Avonts D, Nöstlinger C, Sergeant M, et al. Voluntary outreach counselling and testing for HIV and STI among men who have sex with men in Antwerp. Acta Clin Belg 2012;67(3):172-176. [doi: 10.1179/ACB.67.3.2062651] [Medline: 22897064]
21. Bowles KE, Clark HA, Tai E, Sullivan PS, Song B, Tsang J, et al. Implementing rapid HIV testing in outreach and community settings: results from an advancing HIV prevention demonstration project conducted in seven U.S. cities. Public Health Rep 2008;123 Suppl 3:78-85 [FREE Full text] [Medline: 19172705]
22. Mayer KH, Ducharme R, Zaller ND, Chan PA, Case P, Abbott D, et al. Unprotected sex, underestimated risk, undiagnosed HIV and sexually transmitted diseases among men who have sex with men accessing testing services in a New England bathhouse. J Acquir Immune Defic Syndr 2012 Feb 1;59(2):194-198 [FREE Full text] [doi: 10.1097/QAI.0b013e31823bbecf] [Medline: 22027871]
23. Manavi K, Williams G, Newton R. The uptake of HIV and syphilis testing in a nurse-delivered service during Gay Pride events. Int J STD AIDS 2012 Dec;23(12):887-889. [doi: 10.1258/ijsa.2012.012023] [Medline: 23258830]
24. Thornton AC, Delpech V, Kall MM, Nardone A. HIV testing in community settings in resource-rich countries: a systematic review of the evidence. HIV Med 2012 Aug;13(7):416-426. [doi: 10.1111/j.1468-1293.2012.00992.x] [Medline: 22413900]
25. Champenois K, Le Gall JM, Jacquemin C, Jean S, Martin C, Rios L, et al. ANRS-COM'TEST: description of a community-based HIV testing intervention in non-medical settings for men who have sex with men. BMJ Open 2012;2(2):e000693 [FREE Full text] [doi: 10.1136/bmjopen-2011-000693] [Medline: 22466158]
26. Meulbroek M, Ditzel E, Saz J, Taboada H, Pérez F, Pérez A, et al. BCN Checkpoint, a community-based centre for men who have sex with men in Barcelona, Catalonia, Spain, shows high efficiency in HIV detection and linkage to care. HIV Med 2013 Oct;14 Suppl 3:25-28. [doi: 10.1111/hiv.12054] [Medline: 24033899]

27. Gumy C, Jeannin A, Balthasar H, Huissoud T, Jobin V, Häusermann M, et al. Five-year monitoring of a gay-friendly voluntary counselling and testing facility in Switzerland: who got tested and why? BMC Public Health 2012;12:422 [FREE Full text] [doi: 10.1186/1471-2458-12-422] [Medline: 22682345]
28. Myers JE, El-Sadr WM, Zerbe A, Branson BM. Rapid HIV self-testing: long in coming but opportunities beckon. AIDS 2013 Jul 17;27(11):1687-1695. [doi: 10.1097/QAD.0b013e32835fd7a0] [Medline: 23807269]
29. OraQuick® In-Home HIV Test Summary of Safety and Effectiveness. Silver Spring, MD: US Food and Drug Adminstration; 2012 Jul. URL: http://www.fda.gov/downloads/BiologicsBloodVaccines/BloodBloodProducts/ApprovedProducts/PremarketApprovalsPMAs/UCM312534.pdf [accessed 2015-02-06] [WebCite Cache ID 6W8kN7Q2o]
30. US Food and Drug Administration, FDA Consumer Health Information. Silver Spring, MD: US Food and Drug Administration; 2014 Sep. First rapid home-use HIV kit approved for self-testing URL: http://www.fda.gov/forconsumers/consumerupdates/ucm310545.htm [accessed 2015-08-27] [WebCite Cache ID 6b641mP1V]
31. Centers for Disease Control and Prevention (CDC). False-positive oral fluid rapid HIV tests--New York City, 2005-2008. MMWR Morb Mortal Wkly Rep 2008 Jun 20;57(24):660-665 [FREE Full text] [Medline: 18566566]
32. Pant PN, Sharma J, Shivkumar S, Pillay S, Vadnais C, Joseph L, et al. Supervised and unsupervised self-testing for HIV in high- and low-risk populations: a systematic review. PLoS Med 2013;10(4):e1001414 [FREE Full text] [doi: 10.1371/journal.pmed.1001414] [Medline: 23565066]
33. Gilbert M, Hottes TS, Kerr T, Taylor D, Fairley CK, Lester R, et al. Factors associated with intention to use internet-based testing for sexually transmitted infections among men who have sex with men. J Med Internet Res 2013;15(11):e254 [FREE Full text] [doi: 10.2196/jmir.2888] [Medline: 24240644]
34. Hottes TS, Farrell J, Bondyra M, Haag D, Shoveller J, Gilbert M. Internet-based HIV and sexually transmitted infection testing in British Columbia, Canada: opinions and expectations of prospective clients. J Med Internet Res 2012;14(2):e41 [FREE Full text] [doi: 10.2196/jmir.1948] [Medline: 22394997]
35. Swab2know. URL: https://www.swab2know.be/en/home.aspx [accessed 2015-02-06] [WebCite Cache ID 6W8jXBNBy]
36. Bio-Rad LL. Genscreen™ ULTRA HIV Ag-Ab: Sreening Kit for the Detection of HIV P24 Antigent and Antibodies to HIV-1 and HIV-2 in Human Serum/Plasma by Enzyme Immunoassay. Marnes-la-Coquette, France: Bio-Rad; 2010 Nov. URL: http://www.bio-rad.com/webroot/web/pdf/inserts/CDG/en/883605_EN.pdf [accessed 2015-02-06] [WebCite Cache ID 6W8kSPLgF]
37. Myers JE, Bodach S, Cutler BH, Shepard CW, Philippou C, Branson BM. Acceptability of home self-tests for HIV in New York City, 2006. Am J Public Health 2014 Dec;104(12):e46-e48. [doi: 10.2105/AJPH.2014.302271] [Medline: 25320885]
38. Brady M, Nardone A, Buenaventura E, Qureshi F, Edwardes D, Kelly P, et al. Home HIV sampling linked to national HIV testing campaigns: a novel approach to improve HIV diagnosis. HIV Med 2014;15(Supplement S3):7-8.
39. Zou H, Wu Z, Yu J, Li M, Ablimit M, Li F, et al. Internet-facilitated, voluntary counseling and testing (VCT) clinic-based HIV testing among men who have sex with men in China. PLoS One 2013;8(2):e51919 [FREE Full text] [doi: 10.1371/journal.pone.0051919] [Medline: 23418417]
40. Vanden Berghe W, Nostlinger C, Buvé A, Beelaert G, Fransen K, Laga M. A venue-based HIV prevalence and behavioural study among men who have sex with men in Antwerp and Ghent, Flanders, Belgium, October 2009 to March 2010. Euro Surveill 2011;16(28):1-8 [FREE Full text] [Medline: 21794222]
41. Sanders GD, Bayoumi AM, Sundaram V, Bilir SP, Neukermans CP, Rydzak CE, et al. Cost-effectiveness of screening for HIV in the era of highly active antiretroviral therapy. N Engl J Med 2005 Feb 10;352(6):570-585. [doi: 10.1056/NEJMsa042657] [Medline: 15703422]
42. Walensky RP, Weinstein MC, Kimmel AD, Seage GR, Losina E, Sax PE, et al. Routine human immunodeficiency virus testing: an economic evaluation of current guidelines. Am J Med 2005 Mar;118(3):292-300. [doi: 10.1016/j.amjmed.2004.07.055] [Medline: 15745728]
43. Adamson SC, Bachman JW. Pilot study of providing online care in a primary care setting. Mayo Clin Proc 2010 Aug;85(8):704-710 [FREE Full text] [doi: 10.4065/mcp.2010.0145] [Medline: 20516427]
44. Taylor D, Lunny C, Wong T, Gilbert M, Li N, Lester R, et al. Self-collected versus clinician-collected sampling for sexually transmitted infections: a systematic review and meta-analysis protocol. Syst Rev 2013;2:93 [FREE Full text] [doi: 10.1186/2046-4053-2-93] [Medline: 24112441]

Abbreviations

ART: antiretroviral therapy
ELISA: enzyme-linked immunosorbent assay
FDA: US Food and Drug Administration
IC: informed consent
ITM: Institute of Tropical Medicine
MSM: men who have sex with men
SAM: sub-Saharan African migrants
STI: sexually transmitted infection

UNAIDS: Joint United Nations Programme on HIV/AIDS

Edited by G Eysenbach; submitted 27.02.15; peer-reviewed by C Agusti Benito, M Chen; comments to author 23.07.15; revised version received 04.08.15; accepted 05.08.15; published 01.09.15.

Please cite as:
Platteau T, Fransen K, Apers L, Kenyon C, Albers L, Vermoesen T, Loos J, Florence E
Swab2know: An HIV-Testing Strategy Using Oral Fluid Samples and Online Communication of Test Results for Men Who Have Sex With Men in Belgium
J Med Internet Res 2015;17(9):e213
URL: http://www.jmir.org/2015/9/e213/
doi:10.2196/jmir.4384
PMID:26330138

Letter to the Editor

Twitter-Based Journal Clubs: Additional Facts and Clarifications

Joel M Topf[1], MD; Matthew A Sparks[2], MD; Francesco Iannuzzella[3], MD; Edgar Lerma[4], MD; Thomas Oates[5], MD, PhD; Paul J Phelan[6], MD; Swapnil Hiremath[7], MD, MPH

[1]St John Providence Hospital, Detroit, MI, United States
[2]Duke University and Durham VA Medical Centers, Durham, NC, United States
[3]Arcispedale Santa Maria Nuova, IRCCS, Reggio Emilia, Italy
[4]University of Illinois at Chicago/ Advocate Christ Medical Center, Oak Lawn, IL, United States
[5]Imperial College, London, United Kingdom
[6]Royal Infirmary of Edinburgh, Edinburgh, United Kingdom
[7]Ottawa Hospital Research Institute, Ottawa, ON, Canada

Corresponding Author:
Swapnil Hiremath, MD, MPH
Ottawa Hospital Research Institute
1967 Riverside Dr
Ottawa, ON, K1H7W9
Canada
Phone: 1 613 738 8400
Fax: 1 613 738 8337
Email: shiremath@toh.on.ca

Related Articles:

Comment on: http://www.jmir.org/2015/4/e103/

Comment in: http://www.jmir.org/2015/9/e217/

(J Med Internet Res 2015;17(9):e216) doi:10.2196/jmir.4639

We read the recently published paper on globalization of continuing professional development by Roberts et al with great interest [1]. The authors should be congratulated on their idea as well as their execution of this novel way of evaluating and describing Twitter-based journal clubs. We would like to add to their article by providing some additional advantages and features of a Twitter-based online journal club to provide the reader with a more complete appreciation of their educational potential.

First, we would like to caution the authors from relying on *impressions* as tracked by Symplur for two reasons. First, the impression count is the number of tweets multiplied by the number of followers the participant has. This calculation is performed at the time the analytics are generated, not at the time the participant tweeted. So, if participant X has 30 followers and tweets six times this will add 180 impressions to the analytics. If participant X subsequently gains an additional 970 followers, re-running the analytics will now show that participant X was responsible for 6000 impressions. Since users tend to gain more followers over time this makes early journal clubs look more successful than they actually were. Second, a few highly followed accounts can dramatically influence impressions. Today, #NephJC registers 15.4 million impressions, but this includes 2.4 million impressions from 8 spam accounts that tweeted using the #NephJC hashtag to put their message in front of physicians, but did not meaningfully participate in the chat (see Figure 1). Because of these problems, we advise investigators to be cautious when interpreting *impressions* and focus on the other analytics tracked by Symplur. Unfortunately, there is no easy solution to fix this problem. Regular audits of the hashtag could help to identify such accounts. However, this would require one to manually remove promiscuous tweets using the hashtag of interest.

In the discussion, the authors mention that the "freedom of voluntary participation complicates the establishment of an accurate and efficient record of participation for appropriate ethical acknowledgement for continuing professional development requirements by credentialing authorities." However, registration to a service such as Symplur, allows a Twitter-based journal club to maintain indefinite records of active participation, which we do agree is an essential component of fulfilling continuing professional development requirements by credentialing institutions.

Twitter-based online journal clubs also provide post-publication peer-review (which in the case of #NephJC is captured in a

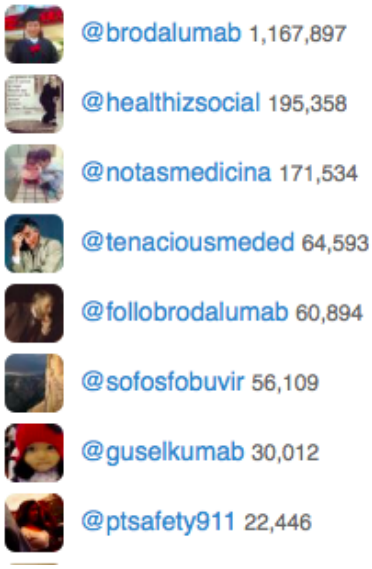
PubMed Commons comment [2] which links to the pre-chat article summary, a complete transcript and a curated Storify of the chat), which is increasingly recognized as a crucial component of knowledge synthesis and critical review.

We agree that the broad based participation (high number of participants) of the international Urology Twitter Journal Club (#UroJC) is indicative of a successful endeavor. We believe that one of the major reasons driving the high number of participants, apart from its longevity (#UroJC started in October 2012), is that the journal club is not conducted as a focused chat but rather is an open period of discussion stretching over a few days. This allows individuals from many different time zones to contribute at a locally convenient time. In contrast, the live nature of other chats (such as #NephJC) provides a vibrant conversational tone, which likely drives a greater number of tweets, but with the trade-off of limiting participation due to inconvenient timing for various time zones.

Lastly, regular updates of a systematic review should be performed when new evidence (usually in the form of new studies or trials) becomes available. Referred to as a "living" systematic review, this concept has been encouraged to keep the literature relevant and to narrow the evidence-practice gap [3]. This aspect is perhaps even more relevant to the present review. Between the period that the present authors conducted their search, and the actual publication, another journal club has come into existence (#RheumJC), and the follower count has changed dramatically in some cases (eg #NephJC from 584 to 1360). Interestingly, the National Library of Medicine is encouraging the archiving of online discussions in medicine, such as these journal clubs, and can serve as a useful resource for researchers in this area [4].

Organizing and producing a Twitter-based medical journal club takes a fair amount of time and effort, especially if one takes into account the background work and post-chat summation, active solicitation of participants, and coordination with authors. These efforts are, currently, not tracked or acknowledged by most academic institutions as scholarly activity [5]. Hence, literature, such as this systematic review, are especially welcome since they can help to validate this work as being of scholarly interest.

Figure 1. The contribution of 8 "spam" accounts that added 2.4 million impressions to the 15.4 million total, thus falsely inflating the impression count.



References

1. Roberts MJ, Perera M, Lawrentschuk N, Romanic D, Papa N, Bolton D. Globalization of continuing professional development by journal clubs via microblogging: a systematic review. J Med Internet Res 2015;17(4):e103 [FREE Full text] [doi: 10.2196/jmir.4194] [Medline: 25908092]
2. Hiremath S, Topf JM. PubMed Commons. Nephrology Journal Club profile URL: http://www.ncbi.nlm.nih.gov/myncbi/nephjc%20-%20nephrology.journal%20club.1/profile [WebCite Cache ID 6YN6F5zmk]
3. Elliott JH, Turner T, Clavisi O, Thomas J, Higgins Julian P T, Mavergames C, et al. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. PLoS Med 2014 Feb;11(2):e1001603 [FREE Full text] [doi: 10.1371/journal.pmed.1001603] [Medline: 24558353]
4. URL: http://www.ncbi.nlm.nih.gov/pubmedcommons/journal-clubs/about/ [WebCite Cache ID 6ZDBDX9sb]
5. Biswas A, Kircherr J. Citations are not enough: Academic promotion panels must take into account a scholar's presence in popular media. URL: http://blogs.lse.ac.uk/impactofsocialsciences/2015/04/09/academic-promotion-scholars-popular-media/ [WebCite Cache ID 6YN6WkIP5]

Edited by G Eysenbach; submitted 08.05.15; peer-reviewed by M Roberts; comments to author 25.05.15; revised version received 11.06.15; accepted 08.07.15; published 16.09.15.

Please cite as:
Topf JM, Sparks MA, Iannuzzella F, Lerma E, Oates T, Phelan PJ, Hiremath S
Twitter-Based Journal Clubs: Additional Facts and Clarifications
J Med Internet Res 2015;17(9):e216
URL: http://www.jmir.org/2015/9/e216/
doi:10.2196/jmir.4639
PMID:26376668

Letter to the Editor

Response to “Twitter-Based Journal Clubs: Some Additional Facts and Clarifications”

Marlon Perera[1,2], BmedSci, MBBS; Matthew Roberts[1,3], MBBS; Nathan Lawrentschuk[2,4,5], MBBS, PhD, FRACS; Damien Bolton[2,4], MBBS, BMedSc, MD

[1]Department of Surgery, School of Medicine, University of Queensland, Toowoomba, Australia

[2]Depeartment of Surgery, Austin Hospital, University of Melbourne, Melbourne, Australia

[3]Centre for Clinical Research, The University of Queensland, Brisbane, Australia

[4]Austin Hospital, Ludwig Institute for Cancer Research, Melbourne, Australia

[5]Division of Cancer Surgery, Peter MacCallum Cancer Centre, Melbourne, Australia

Corresponding Author:
Marlon Perera, BmedSci, MBBS
Department of Surgery
School of Medicine
University of Queensland
Level 5, Anne-Webb Building, Level 5 Pechy St
Toowoomba, 4350
Australia
Phone: 61 416005137
Fax: 61 416005137
Email: marlonlperera@gmail.com

Related Articles:

Comment on: http://www.jmir.org/2015/9/e216/

Comment on: http://www.jmir.org/2015/4/e103/

(J Med Internet Res 2015;17(9):e217) doi:10.2196/jmir.4786

We read with great interest the recent correspondence from Topf et al [1] regarding our recent publication “Globalization of Continuing Professional Development by Journal Clubs via Microblogging: A Systematic Review” [2]. We thank the authors for their interest, opinions, and contribution to the ongoing work evaluating the utility of Twitter-based Journal Clubs in the context of continuing professional development.

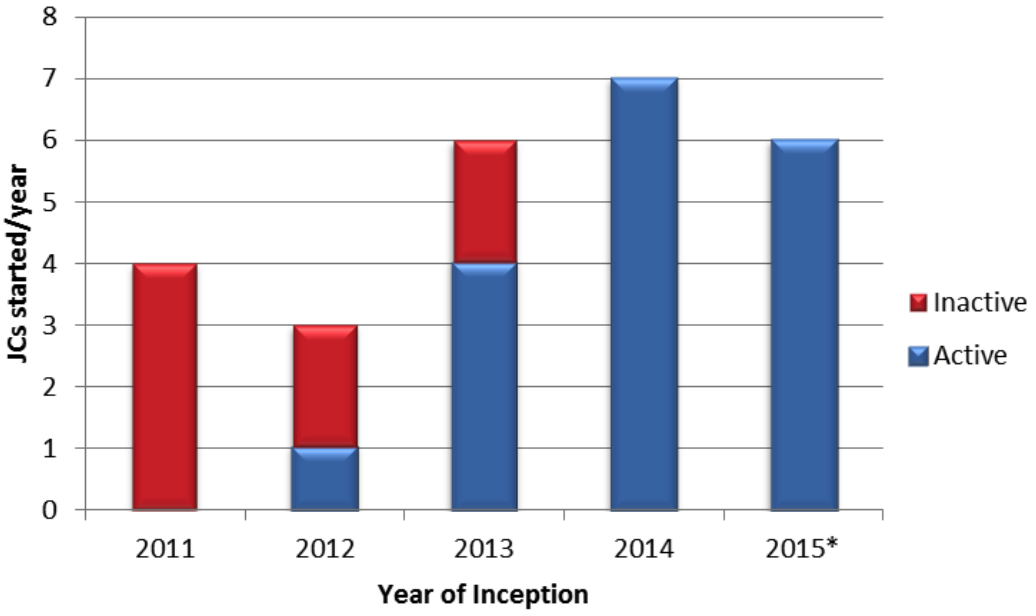
Topf et al note the limitation associated with the dynamic nature of the "*impressions"* data as a reported outcome measure and provide a well-explained example of how this metric is dynamic. Further, they correctly note that "spam" accounts associated with the journal clubs (JC) artificially increase the total impressions for a Twitter journal club. Despite this, when used appropriately, we believe there is some value to impressions as a performance metric given the paucity of comparative outcome measures in the early Twitter-based journal club era. This education tool is unique and traditional analysis methods typically used in systematic reviews and meta-analytical studies are clearly not suitable. The "*impression:tweet ratio"* reported in the initial manuscript was a metric applied only to the Twitter user account to assess the following and “impression” of the journal club, thus reducing the influence of such "spam" accounts to a degree. We believe this modified calculation is a useful quantifiable measure of publicity and potential viewership of the discussion. However, for trend analysis, such as that performed for the top five performing journal clubs, the identification and exclusion of such accounts (eg, @brodalumab) was performed as they were not only statistical outliers but also known spam accounts. This helped us to provide highly accurate data in this analysis.

The dynamic nature of Twitter-based JC was pertinently raised by the authors, as evidenced by the commencement of recent JCs. We support the notion of a "living" systematic review, not currently possible given the publication using traditional peer-reviewed methods and associated delays. The suggested method of “Storifying” the chat is an appealing method for consolidation and formalization of the conversation for later review. The value of these conversations for scholarly activity is gaining momentum, with some institutions promoting Altmetric scores for affiliated publications. Furthermore, Symplur in this context as a real-time aggregate database is an invaluable tool in appreciating the changes in journal club

discussions. We anticipate that with further time and refinement, more sophisticated methods for measuring journal club performance will be devised. The ongoing success of current and future journal clubs will be determined by appropriate identification and recommendation from experienced participants with advice for successes and pitfalls from established JCs.

Given the current opportunity to present updated data six-months following the previous review [2], 6 more Twitter journal clubs have been established and none have become inactive (see Figure 1). These new journal clubs represent diverse groups within the medical field including rheumatology (#rheumJC), radiology (#medradJClub) and neuro-crictical care (#NCSTJC). Additionally, several recent publications regarding the use of social media for medical education, specifically journal clubs, have become apparent. Of these, several represent publication of summaries of a recent Twitter-based journal club discussion [3-5] or narrative reviews on the evolution of Twitter-based journal clubs [6]. Further recent publications have assessed the uptake of Twitter-based journal clubs by respective societies [7,8].

Figure 1. Establishment of JCs per year, comparing active JCs (blue) with inactive JCs (red). 2015 included JCs started prior to May 2015.



References

1. Topf JM, Sparks MA, Iannuzzella F, Lerma E, Oates T, Phelan PJ, et al. Twitter-Based Journal Clubs: Additional Facts and Clarifications. J Med Internet Res 2015;17(9):e216 [FREE Full text] [doi: 10.2196/jmir.4639]
2. Roberts MJ, Perera M, Lawrentschuk N, Romanic D, Papa N, Bolton D. Globalization of continuing professional development by journal clubs via microblogging: a systematic review. J Med Internet Res 2015;17(4):e103 [FREE Full text] [doi: 10.2196/jmir.4194] [Medline: 25908092]
3. Colman R, Anand A. Twitter discussions from a respirology journal club. Lancet Respir Med 2015 Apr;3(4):e10. [doi: 10.1016/S2213-2600(15)00113-7] [Medline: 25890653]
5. Westafer L, Hensley J, Shaikh S, Lin M. Global Emergency Medicine Journal Club: A Social Media Discussion About the Lack of Association Between Press Ganey Scores and Emergency Department Analgesia. Ann Emerg Med 2015 May 20 (forthcoming). [doi: 10.1016/j.annemergmed.2015.04.012] [Medline: 26003003]
6. Topf JM, Hiremath S. Social media, medicine and the modern journal club. Int Rev Psychiatry 2015 Apr;27(2):147-154. [doi: 10.3109/09540261.2014.998991] [Medline: 25906989]
7. Davies N, Murphy DG, van RS, Woo HH, Lawrentschuk N. Online and social media presence of Australian and New Zealand urologists. BJU Int 2015 Apr 24. [doi: 10.1111/bju.13159] [Medline: 25906813]
8. Misrai V, Castagnola C, Descotes JL, Rouprêt M. [Use of social media by French urologists: Results from a study of the National French Urological Association]. Prog Urol 2015 Jun;25(7):428-434. [doi: 10.1016/j.purol.2015.03.008] [Medline: 25921610]

Edited by G Eysenbach; submitted 31.05.15; peer-reviewed by I Thangasamy, S Hiremath; comments to author 01.07.15; revised version received 04.07.15; accepted 08.07.15; published 18.09.15.

Please cite as:
Perera M, Roberts M, Lawrentschuk N, Bolton D
Response to "Twitter-Based Journal Clubs: Some Additional Facts and Clarifications"
J Med Internet Res 2015;17(9):e217
URL: http://www.jmir.org/2015/9/e217/
doi:10.2196/jmir.4786
PMID:26385506

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org