Journal of Medical Internet Research

Journal Impact Factor (JIF) (2022): 7.4
Volume 20 (2018), Issue 8 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

XSL•FO
RenderX

XSL•FO
RenderX

Reviews

Proposal

Corrigenda and Addenda

XSL•FO
RenderX

Original Paper

Effect and Process Evaluation of e-Powered Parents, a Web-Based Support Program for Parents of Children With a Chronic Kidney Disease: Feasibility Randomized Controlled Trial

Wytske W Geense[1], MSc; Betsie GI van Gaal[1,2], PhD; Jaqueline L Knoll[3], RN, MANP; Nienke M Maas[4], PhD; Gerjo Kok[5], PhD; Elisabeth AM Cornelissen[3], MD, PhD; Maria WG Nijhuis-van der Sanden[1], PhD

[1]IQ Healthcare, Radboud Institute for Health Science, Radboud University Medical Center, Nijmegen, Netherlands

[2]Institute of Nursing, HAN University of Applied Science, Nijmegen, Netherlands

[3]Department of Pediatric Nephrology, Amalia Children's Hospital, Radboud University Medical Center, Nijmegen, Netherlands

[4]Department of Medical Psychology, Radboud University Medical Center, Nijmegen, Netherlands

[5]Department of Work & Social Psychology, Maastricht University, Maastricht, Netherlands

Corresponding Author:
Wytske W Geense, MSc
IQ Healthcare
Radboud Institute for Health Science
Radboud University Medical Center
PO Box 9101
6500 HD
Nijmegen,
Netherlands
Phone: 31 24 36 66 868
Email: wytske.geense@radboudumc.nl

Abstract

Background: Parents of children with chronic kidney disease (CKD) experience high levels of stress in the daily management of their child's illness. Parents need continuously available support and information, yet online support programs are lacking. e-Powered Parents was developed to fill this gap; it is an online program consisting of (1) medical information, (2) an interactive part, and (3) four training modules (stress management, setting limits, communication, and coping). Prior to a large-scale evaluation, we conducted a feasibility study that consisted of an effect study and a process evaluation.

Objective: The objectives of our study were to (1) identify the outcome measures that are most likely to capture the potential benefit, (2) evaluate the potential effectiveness and effect size, and (3) evaluate recruitment, reach, the dose received, and context.

Methods: We conducted a feasibility study with a two-armed, wait-list randomized controlled trial (RCT). Prior to baseline, parents (n=146) were randomly allocated to group 1 or group 2. After completing the baseline questionnaire, parents in group 1 were given access to e-Powered Parents, while those in group 2 received usual care. At the 6-month follow-up (T1), all parents received a questionnaire and parents in group 2 were given access to e-Powered Parents as well. After 1.5 years, through an extra measurement (T2), we evaluated the effect of long-term exposure. Outcomes were the child's quality of life (Child Vulnerability Scale), parental stress (Pediatric Inventory for Parents) and fatigue (Multidimensional Fatigue Inventory), self-efficacy in communication with health care professionals (Perceived Efficacy in Patient-Physician Interactions, PEPPI-5), and parental perceptions of family management (Family Management Measure). Floor and ceiling effects and percentage of parents showing no change in scores were calculated. We used linear mixed models to evaluate the potential effectiveness and effect sizes using the intention-to-treat and per-protocol analyses. In the process evaluation, we evaluated recruitment, reach, the dose received, and context using a questionnaire sent to the parents, log-in data, and a focus group interview with health care professionals.

Results: At T1 (n=86) and T2 (n=51), no significant effects were found on any of the five outcomes. The PEPPI-5 showed ceiling effects and high percentages of parents showing no change between the measurement times. The information and interactive part of the intervention were used by 84% (57/68) of the parents in group 1 and 49% (32/65) of the parents in group 2. The information pages were visited most often. Overall, 64% (85/133) of the parents logged in to the training platform and 31% (26/85) actually used the training modules.

XSL•FO
RenderX

Conclusions: We did not observe any significant effect on any of the outcomes. This could possibly be explained by the minimal use of the intervention and by parents' heterogeneity. For continued participation, we recommend a tailored intervention and further studies to find out whether and how online programs could be used to support parents in the management of their child's CKD.

Trial Registration: Netherlands Trial Registry NTR4808; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=4808 (Archived by WebCite at http://www.webcitation.org/719rCicvW)

(J Med Internet Res 2018;20(8):e245) doi:10.2196/jmir.9547

KEYWORDS

child; chronic kidney failure; chronic kidney disease; effect evaluation; health promotion; process evaluation

Introduction

Parents play a key role in the management of their child's illness. However, parents of children with chronic kidney disease (CKD) experience high levels of stress. Complications such as infections, bone diseases, poor growth and development, and kidney failure are frequently seen among such children [1,2]. Mortality among children with CKD remains 30 times higher than that among healthy children, despite renal replacement therapy or kidney transplantation [2]. Moreover, care for these children is complex due to complicated medication schedules, nutritional restrictions, and procedures such as hemodialysis or peritoneal dialysis [2]. It is, therefore, not unusual that parents experience difficulties in balancing the needs of their sick child with their own responsibilities, such as other children, family members, work, and social life [3]. Parents with significant emotional distress of their own and poor family function can negatively affect their child's health outcomes as well as their quality of life [4]. Supporting these parents is, therefore, necessary to help them cope with the difficulties encountered in all the stages of their child's CKD.

In recent years, more attention has been paid to the development of psychoeducational support programs for parents of children with chronic diseases to assist families with the day-to-day management of their child's chronic disease and its consequences [3,5]. Support programs for parents can take many forms, consisting, for example, of a simple provision of information via written materials, computer programs, internet programs, and group interventions. Eccleston et al [5] concluded in their Cochrane review that more psychological interventions are needed that directly target the parents of children with chronic illness. In 2008, Swallow et al [6] described how parents of children with CKD needed continuously available, accessible, and reliable support. Although in recent years, an increasing number of online support programs have been developed for parents of children with chronic diseases, such as diabetes mellitus [7,8], cystic fibrosis [9,10], and asthma [11], online support programs for parents of children with CKD are as yet lacking.

To fill this gap, we developed *e-Powered Parents (Mijn Kinderniernet* in Dutch), an online support program for parents of children with CKD, using intervention mapping. Intervention mapping is a protocol for the systematic development of theory- and evidence-based health promotion interventions, consisting of six different steps [12]. The completion of all these steps serves as a blueprint for designing, implementing, and evaluating an intervention based on the foundation of theoretical, empirical, and practical information [12]. After conducting a needs assessment with parents (consisting of 5 focus group interviews) [13] and with health care professionals (step 1), defining program objectives (step 2), and searching for theories and selecting practical applications (step 3), we developed *e-Powered Parents* (step 4; Multimedia Appendix 1). Subsequently, a plan was designed for the adoption, implementation, and sustainability of *e-Powered Parents* (step 5) [14].

The last step of intervention mapping (step 6) includes the planning for evaluation [15]. We decided to conduct a feasibility randomized controlled trial (RCT) because prior to a large-scale evaluation of developing and testing interventions, feasibility studies are essential [16]. The aims of this feasibility study were to identify outcome measures that are most likely to capture the potential benefit and to evaluate the potential effectiveness and effect size of the program. We also conducted a process evaluation to understand the results of the effect evaluation. According to the Medical Research Council [17], process evaluations are an essential part of designing and testing complex interventions to look inside the so-called "black box" to see what happened in the program and how that could affect program outcomes [18].

Methods

Objectives

This feasibility study consisted of an effect and process evaluation that covered 3 objectives. The objectives of the effect evaluation were (1) to identify outcome measures most likely to capture the potential benefit and (2) to evaluate the potential effectiveness and effect size.

The objective of the process evaluation was (3) to evaluate the recruitment, reach, dose received, and context of *e-Powered Parents*.

Study Design and Randomization

For the effect evaluation, we conducted a feasibility, two-armed RCT at the Pediatric Nephrology Unit of a single university medical center in the Netherlands. In the course of the study, we changed the design to a wait-list RCT (as described below). For the process evaluation, we conducted a quantitative and qualitative study alongside the RCT.

XSL•FO
RenderX

Figure 1. Study design. From June 2015 until December, e-Powered Parents was improved. Group 1: access to e-Powered Parents between T0-T2; Group 2: access to e-Powered Parents between T1-T2.

Website	Part	Topics
Community	*Information part*	• Information and instructions how *e-Powered Parents* (community and training platform) works, including an online manual
		• Organization of pediatric nephrology unit (accessibility and team members)
		• Kidneys, kidney diseases and chronic kidney damage
		• Nutrition (e.g. proteins, phosphates, sodium, potassium, water, energy and recipes)
		• Medication (e.g. blood pressure, iron deficiency, erythropoietin, vitamins, potassium binders, growth hormone, immune suppressants)
		• Hemodialysis (e.g. shunt and dialysis catheter) and peritoneal dialysis (e.g. catheter, infections, continuous ambulatory peritoneal dialysis)
		• Kidney transplantation (e.g. bloodtypes and compatibility, rejection of transplant, living donor, screening and preparation, during and after)
		• Growing up with chronic kidney disease (e.g. transition from a pediatric nephrologist to an adult nephrologist)
		• Education, school, and children's hospital school
		• Finance, laws, and regulations
		• Sport
		• Scientific research, projects, and guidelines
	Interaction part	• Blog
		• Forum (with items such as nutrition, medication, dialysis, transplantation, kidney diseases, school and education, holidays and finance and insurance
		• Chat
		• Private messages
Training platform	*Training modules*	• Welcome
		• Stress management module • Welcome • Symptoms of stress • Coping with stress • Creating a personal stress management plan
		• Setting limits module • Welcome • Why saying no is important • How to say no • Handing over care
		• Communication module • Welcome • Active listening • Asking questions • Being assertive • Communication in practice (healthcare professional, child and other children in family)
		• Coping with your child's CKD module • Coping • Tips that could help you coping with your grief

Randomization, stratified on CKD stages at the family level (see subsection Participants and Recruitment) and performed blind by a statistician using a computer random number generator, was used to allocate equal numbers of parents to the intervention group (group 1) and control group (group 2). After the baseline survey (T0) in January 2015, parents in group 1 were given access to the intervention *e-Powered Parents*, while parents in group 2 received the usual care (Figure 1). After the first follow-up measurement (T1) in June 2015, the Pediatric Nephrology Unit implemented *e-Powered Parents* as part of daily care for children with CKD. Subsequently, parents in group 2 were given access to *e-Powered Parents* as well. We used the feedback from the parents in group 1 at T1 to improve *e-Powered Parents* for a period of 5 months (July to November 2015), adding more information about kidney diseases and supporting videos. After 6 months (May 2016), we conducted an extra measurement (T2) to measure the effect of a longer exposure to the improved program. The design of the study changed thereby from a two-armed RCT to a two-armed, wait-list RCT (Figure 1).

XSL•FO
RenderX

Ethical Considerations

The Medical Ethics Review Committee of the district Arnhem-Nijmegen approved the study (Registration number 2014/302). All parents received written information about the study's content and aim and were only included after providing written informed consent. For the extension of the study, the Medical Ethics Review Committee of the district Arnhem-Nijmegen approved the study once more. This trial was registered under the Dutch Trial registration (NTR4808).

Participants and Recruitment

In this study, Dutch-speaking parents of children aged 0-18 years in 5 different CKD groups were eligible; we included parents of children (1) with hereditary kidney disease (CKD stage I); (2) with nephrotic syndrome (CKD stage I); (3) with chronic kidney failure (CKD stage II-IV); (4) using dialysis (CKD stage V); and (5) with renal transplantation. Both parents of each child were invited to participate. Parents were excluded when their child was not living at home anymore.

Parents were recruited between September and December 2014. They received an information letter, which included an informed consent form. After 3 weeks, reminders were sent and phone calls made, while health care professionals (JK and EC) informed the parents about the study during consultation in the outpatient clinic.

Sample Size

The first objective of this study was to identify the potential outcome measures; for that reason, a formal power calculation for a test comparing the treatment groups was not appropriate. Our aim was to include as many parents as possible.

Intervention and Standard Care

e-Powered Parents consisted of two different components:

An online community consisting of an informative part comprising information and videos about various kidney diseases, treatment possibilities, diets, and (financial) regulations and an interactive part comprising a forum, chat room, and option to send private messages to share parents' experiences with other parents and health care professionals.

A Web-based training platform consisting of 4 different training modules (stress management, setting limits, communication, and coping; Multimedia Appendix 1).

Parents in both groups 1 and 2 received standard care.

Data Collection and Measurements

See Table 1 for data collection regarding the effect evaluation and process evaluation. We collected data after randomization (T0), 6 months (T1), and 1.5 years (T2) through a Web-based questionnaire. Additional data were collected by extracting the log-in data of *e-Powered Parents* as well as through a focus group interview.

Effect Evaluation

We selected five potential main outcomes to identify the outcome measures that were most likely to capture the potential benefit (first objective) and to evaluate the potential effectiveness and effect size (second objective):

1. *Child's quality of life* was measured using the validated Dutch version [19] of the Child Vulnerability Scale (CVS) [20], a proxy instrument measuring the parental perceptions of a child's vulnerability, which is related to the child's health-related quality of life [21]. Each of the 8 items is rated on a 4-point Likert scale (0="definitely true"; 3="definitely false"). The total score ranges from 0 to 24. Higher scores indicate an increased vulnerability.
2. *Pediatric-related parental stress* was measured using the validated Dutch version [22] of the Pediatric Inventory for Parents (PIP) [23], a 42-item questionnaire covering 4 domain scales: "Communication," "Emotional distress," "Medical care," and "Role function." Each item is rated on a 5-point Likert scale (1="not at all"; 5="extremely"). The 4 domain scores are summed up, resulting in a total overall frequency score and a total difficulty score. Higher scores indicate a higher frequency and difficulty.
3. *Parental fatigue* was measured using the validated Dutch version of the Multidimensional Fatigue Inventory (MFI) [24], a 20-item instrument covering 5 dimensions: "General fatigue," "Physical fatigue," "Mental fatigue," "Reduced motivation," and "Reduced activity." Each item is rated on a 5-point Likert scale, ranging from "Yes, that is true" to "No, that is not true." Higher scores indicate a higher degree of fatigue.
4. *Self-efficacy in the communication with health care professionals* was measured using the validated Dutch version [25] of the Perceived Efficacy in Patient-Physician Interactions (PEPPI-5) [26]. Participants rate each of the 5 items on a 5-point Likert scale, with 1 representing "Not at all confident" and 5 representing "Very confident." Total scores are summed up, ranging from 5 to 25; higher scores indicate a higher perceived self-efficacy in patient-physician interactions.
5. *Parental perceptions of the family management of chronic conditions* were measured using the Family Management Measure (FaMM) [27]. The FaMM measures how families manage caring for a child with a chronic condition and the extent to which they incorporate condition management into their everyday family life. The FaMM consists of 45 items, covering 5 dimensions: "Child's daily life," "Condition management ability," "Condition management effort," "Family life difficulty," and "View on condition impact." The sixth dimension, "Parental mutuality" (8 items) is additional for partnered parents. Higher scores on the scale "Child's daily life," "Condition management ability," and "Parental mutuality" indicate a greater ease in managing the child's condition. Higher scores on the other three scales indicate a greater difficulty in managing the condition. Because no validated translation was available for the FaMM, we decided to translate it using forward translation and expert panel back-translation by translators and the members of our team.

Process Evaluation

For the process evaluation (the third objective), we used 4 out of 6 components of the model of Linnan and Steckler [28]: (1)

XSL•FO
RenderX

"recruitment," (2) "reach," (3) "the dose received," and (4) "context." The components "dose delivered" and "fidelity" were not relevant for the Web-based intervention *e-Powered Parents*. We used quantitative research methods for the components "recruitment," "reach," and "the dose received" and a qualitative research method for the component "context."

- *Recruitment* consists of the procedures used to approach, attract, and maintain parents; the number of parents who agreed to participate in the study; and the experienced barriers and reasons for nonparticipation [15,18,28]. We used Excel to register parents who wanted to participate. Furthermore, we added open-ended questions to the Web-based survey at T1 and T2 in order to gain insights into the experienced barriers and reasons for nonparticipation.
- *Reach* is the proportion of parents who actually visited *e-Powered Parents* [28]. In Excel, we registered the parents who had logged in at the community (informative and interactive part) or training platform and had set up an account.
- *The dose received* could be divided into exposure and satisfaction. Exposure is the extent to which the parents actively engaged and interacted with *e-Powered Parents* [15,28]. Satisfaction registers the parents' satisfaction with *e-Powered Parents* [18]. Both the community and the training platform allowed us to extract log-in data, which were used to gain insights into exposure. However, the content and extraction method differed: the log-in data regarding the community included the data of the pages most often visited, number and timing of site visits, page views, and, for example, user device type. We collected these data based on the internet protocol (IP) address in Google Analytics. The log-in data of the training platform included data regarding frequent visits as well as use of the training modules and sessions. Users' data were registered in the program itself based on the email address and could be extracted using Excel. Furthermore, log-in data for the two websites were collected between T0 and T2. To evaluate parents' satisfaction, we added open-ended and Likert scale questions regarding the program in general (such as log-in, navigation on the site, and layout) and the relevance and added value to the Web-based survey at T1 (for group 1) and T2 (for groups 1 and 2).
- *Context* includes aspects of the physical, social, and political environment that may affect the implementation of *e-Powered Parents* [15,28]. To explore the context, we conducted a focus group interview with health care professionals of the Pediatric Nephrology Unit after T2. Pediatric nephrologists, (specialist) nurses, social workers, psychologists, and educational workers were invited to participate. We used purposive sampling to ensure that professionals with and without *e-Powered Parents* experience participated in this study. An experienced external moderator posed open-ended questions about the experiences with and implementation of *e-Powered Parents* and possible improvements for the future. One researcher (WG) acted as an observer in the focus group. The focus group interview took approximately 1.5 hours, was audiotaped, and transcribed verbatim.

Data Analysis

Effect Evaluation

To identify the outcome measures most likely to capture the potential benefit (Objective 1), we calculated the percentages of parents scoring zero (floor effect) or full marks (ceiling effects) on the five outcome measures—CVS, PIP, MFI, PEPPI-5, and FaMM. In this calculation, we considered floor and ceiling effects exceeding 20% to be significant [29]. Additionally, we calculated the percentage of parents showing no change in the score between T0, T1, and T2 [30].

XSL•FO
RenderX

Table 1. Overview of the measurement methods, objectives, and time points.

Research method and specification	Participants	Outcome (time and group)
Quantitative research method		
Effect evaluation (objectives 1 and 2)		
Web-based survey		
CVS[a]	Parents	• Child's quality of life (T0-T2: groups 1 and 2)
PIP[b]	Parents	• Parental stress (T0-T2: groups 1 and 2)
MFI[c]	Parents	• Parental fatigue (T0-T2: groups 1 and 2)
PEPPI-5[d]	Parents	• Self-efficacy in communication (T0-T2: groups 1 and 2)
FaMM[e]	Parents	• Family management (T0-T2: groups 1 and 2)
Process evaluation (objective 3)		
Log-in data of website		
Community[f]	Parents	• *The dose received_Exposure*: Pages visited most often (T0-T1: group 1; T1-T2: groups 1 and 2) • Number and time of site visits (T0-T1: group 1; T1-T2: groups 1 and 2) • Page views (T0-T1: group 1; T1-T2: groups 1 and 2) • Time spent on the site (T0-T1: group 1; T1-T2: groups 1 and 2) • User device type (T0-T1: group 1; T1-T2: groups 1 and 2)
Training platform[g]	Parents	• *The dose received_Exposure*: Frequency and use of the platform (T0-T1: group 1; T1-T2: groups 1 and 2) • Use of training modules and sessions (T0-T1: group 1; T1-T2: groups 1 and 2)
Web-based survey		
Open questions and 4-point Likert scale[h]	Parents	• *Recruitment*: Experienced barriers and facilitators (T0-T1: group 1; T1-T2: groups 1 and 2) • *The dose received_Satisfaction*: Parents' experiences and satisfaction regarding the components of *e-Powered Parent*[i] (T0-T1: group 1; T1-T2: groups 1 and 2)
Own records	Parents	• *Recruitment*: Procedures used to approach and maintain parents (T0-T2: groups 1 and 2) • Number of parents who participate (T0-T2: groups 1 and 2) • *Reach*: Parents who actually visited *e-Powered Parents* (T0-T1: groups 1; T1-T2: groups 1 and 2)
Qualitative research method		
Process evaluation (objective 3)		
Focus group		
Open questions[j]	Health care professionals	• *Context*: Health care professionals' experiences and satisfaction with *e-Powered Parents* program components, use, implementation, and how these components could be improved (after T2)

[a]CVS: Child Vulnerability Scale.

[b]PIP: Pediatric Inventory for Parents.

[c]MFI: Multidimensional Fatigue Inventory.

[d]PEPPI-5: Perceived Efficacy in Patient-Physician Interactions.

[e]FaMM: Family Management Measure.

[f]Informative and interaction part of e-Powered Parents (based on IP address of parents).

[g]Based on email address of parents.

[h]Ranging from "totally disagree" to "totally agree."

XSL•FO
RenderX

[i]Information, newsletter, blog, chat, forum, and training modules.

[j]Asked by a moderator during the focus group.

To evaluate the potential effectiveness and effect size (Objective 2), we used linear mixed models with time of measurement (T1 and T2) and exposure to *e-Powered Parents* (at T0, no exposure; at T1, only parents in group 1 exposed, and at T2, parents in both groups exposed) as fixed variables. Three outcome measures (CVS, PIP, and FaMM) were child specific: participating families that consisted of 2 children with CKD had to fill in these questions twice (for every child each). For the analysis of these three outcomes, we took random child effects into account for potential correlation. For nonchild-specific questionnaires (PEPPI-5 and MFI), we took random family effects into account for potential correlation because both partners participated in the study. Also, we conducted intention-to-treat analysis, followed by per-protocol analysis. Results were considered significant if $P<.05$.

We calculated the standardized effect sizes (Cohen *d*) by dividing the mean difference in the change score between groups 1 and 2 by SD at the baseline. Effect sizes>0.8 were considered as large, between 0.5 and 0.8 as medium, between 0.2 and 0.5 as small, and <0.2 as very small.

Process Evaluation

In this section, we have described the data analysis for the components *recruitment*, *the dose received*, and *context*.

Recruitment

We analyzed the open-ended questions in the questionnaire using thematic analysis in Atlas.ti, a software program used to analyze qualitative data. Experiences of parents were divided into the program components (information, blog, chat, forum, training modules, etc) and subdivided into positive or negative experiences.

The Dose Received

For the community (informative and interactive part), we selected time periods in Google Analytics (eg, January-June 2015) to track the amount of (returning) visitors as well as the information most often read, newsletters, forum topics, chat messages, and private messages sent.

The training platform consisted of 4 modules (stress management, setting limits, communication and coping), with each module consisting of several sessions (Multimedia Appendix 1). Per session, we counted the number of parents finishing that particular session.

For the analysis of the log-in data of the community and training platform, we consulted experts on how best to analyze the data. The 4-point Likert scale questions in the Web-based survey were analyzed in SPSS and merged into 2 categories (Disagree and Agree). Furthermore, frequencies of disagree and agree were counted.

Context

Two researchers independently analyzed the focus group interview with the health care professionals using thematic analysis in Atlas.ti. Themes regarding the components of the program (such as “information,” “communication parents with parents,” “communication parents with health care professionals,” “training platform,” and “layout”) and implementation were labeled while within every theme, we defined negative and positive experiences. The health care professionals who participated in the focus group interview received the conclusions of the analysis to check its authenticity.

Results

Study Population

We assessed a total of 201 families, whose children were under treatment at the Pediatric Nephrology Unit. Regarding eligibility for the study, 22 families were excluded because of not meeting the inclusion criteria and 9 families for other reasons, such as incorrect personal data. The remaining 170 families were asked to participate, of which 81 families declined either due to lack of time or because they did not feel a need for support.

Finally, 146 parents of 89 families who were willing to participate were randomized into group 1 (n=74) and group 2 (n=72). After randomization, 13 parents did not fill in the baseline questionnaire, resulting in a total of 133 parents: 68 parents (43 families) in group 1 and 65 parents (42 families) in group 2 (Figure 2).

Characteristics of Parents and Children

The characteristics of parents in groups 1 and 2 demonstrated similar proportions regarding gender, educational level, and marital status. Regarding children’s characteristics, group 1 had slightly greater proportion of girls and a higher age of children. Overall, the majority of participating parents in both groups were females (83/133, 62%), highly educated (67/133, 50%), married (94/133, 71%), and employed (116/133, 87%). Three families had 2 children with CKD under treatment at the university medical center (Table 2).

Characteristics of Health Care Professionals

We invited 9 health care professionals to take part in the focus group interview. Of them, 5 eventually participated: 1 nurse practitioner, 1 nurse, 2 pediatric nephrologists, and 1 educational worker. All participants were females.

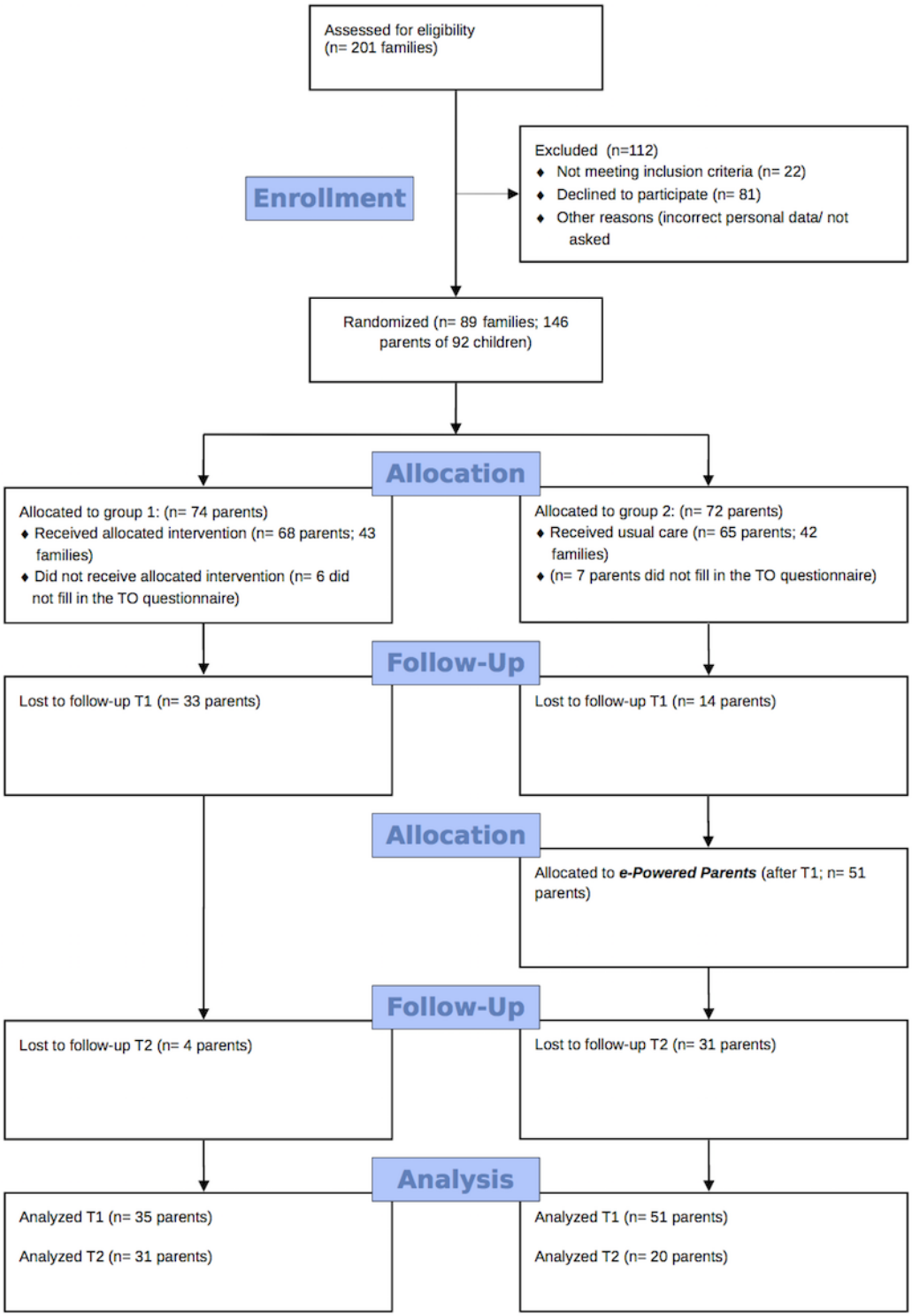
Effect Evaluation

When we assessed the percentage of parents scoring 0 (floor) or full marks (ceiling effects) on the five outcome measures, we noticed significant (>20%) ceiling effects at the PEPPI-5 among parents in group 2—T0 (14/65, 21.5%), T1 (14/51, 27.5%), and T2 (5/20, 25%; Multimedia Appendix 2).

The percentage of parents showing no change was high on the PEPPI-5 as well, with 40.6% (56/138) between T0-T1 and 25.4% (35/138) between T0-T2 (Multimedia Appendix 3).

XSL•FO
RenderX

Figure 2. Flow diagram.



The standardized effect sizes (Cohen *d*) for the five outcome measures were small, ranging in the intention-to-treat analysis from −0.22 to 0.15 and from −0.21 to 0.12 for T0-T1 and T0-T2, respectively (Multimedia Appendix 4). In the per-protocol analysis, these ranged from −0.23 to 0.15 and from −0.20 to 0.12 for T0-T1 and T0-T2 (Multimedia Appendix 5).

In Table 3, we have provided an overview of the mean score on the five outcome measures during T0, T1, and T2. Adjusting for the period of using *e-Powered Parents* (intention-to-treat), the actual use (did parents log in at *e-Powered Parents*: per-protocol), and the experienced stress levels of the parents, no statistically significant differences were found regarding any of these outcomes between parents in groups 1 and 2 between T0-T1 and T0-T2 (Multimedia Appendices 4 and 5).

XSL•FO
RenderX

Table 2. Characteristics of parents and their children at the baseline.

Characteristic	Group 1	Group 2
Parents	**N=68**	**N=65**
Gender, n (%)		
Female	42 (62)	41 (63)
Male	26 (38)	24 (37)
Age in years, mean (SD)	45.2 (6.5)	42.6 (7.3)
Educational level, n (%)[a]		
Low	5 (7)	7 (11)
Medium	27 (40)	26 (41)
High	36 (53)	31 (48)
Marital status, n (%)		
Single	12 (18)	11 (17)
Married	46 (68)	48 (74)
Divorced	1 (1)	4 (6)
Widow	1 (1)	0 (0)
Registered partnership or living together	8 (12)	2 (3)
Housing, n (%)		
With partner and children	65 (96)	62 (95)
With partner	3 (4)	3 (5)
Parent with a chronic disease, n (%)	4 (6)	7 (11)
Job, n (%)	59 (87)	57 (88)
Nanny, n (%)	24 (35)	35 (54)
Experienced stress in last 6 months, n (%)	33 (49)	30 (46)
Children	**N=45**	**N=43**
Gender, n (%)		
Girls	22 (49)	15 (35)
Boys	23 (51)	28 (65)
Child's age in years, mean (SD)	11.1 (4.6)	8.8 (5.2)
CKD[b] stage child, n (%)		
CKD stage I (Hereditary CKD)	11 (24)	10 (23)
CKD stage I (Nephrotic syndrome)	7 (16)	5 (12)
CKD stage II-IV	9 (20)	9 (21)
CKD stage V (using dialysis)	2 (4)	3 (7)
After transplantation	16 (36)	16 (37)
Child also under treatment at other centers, n (%)	24 (53)	21 (49)

[a]One participant did not answer the question.

[b]CKD: chronic kidney disease.

XSL•FO
RenderX

Table 3. Results of the five outcomes.

Outcome measure and group[a]	T0[b], mean (SD)	T1[c], mean (SD)	T2[d], mean (SD)
Child's quality of life (CVS[e], range scale 0-24)			
Group 1	7.4 (4.1)	6.8 (4.4)	5.8 (3.7)
Group 2	7.5 (4.1)	7.1 (4.5)	6.8 (4.8)
Parental stress (PIP[f], range scale 42-210)			
Frequency			
Group 1	99.0 (22.5)	95.7 (24.1)	89.3 (16.6)
Group 2	99.7 (26.3)	97.9 (26.5)	98.0 (21.6)
Difficulty			
Group 1	87.9 (27.2)	85.3 (31.0)	74.6 (21.6)
Group 2	80.8 (25.0)	79.0 (24.6)	78.5 (18.0)
Parental fatigue (MFI[g], range domain scale 4-20)			
General fatigue			
Group 1	11.7 (3.4)	11.5 (3.8)	10.4 (3.1)
Group 2	12.0 (3.8)	11.4 (3.9)	10.8 (4.6)
Physical fatigue			
Group 1	9.7 (2.9)	10.1 (3.4)	9.1 (2.7)
Group 2	9.8 (3.5)	9.7 (3.5)	9.7 (3.8)
Mental fatigue			
Group 1	9.6 (3.0)	9.8 (3.4)	9.0 (3.0)
Group 2	9.9 (3.7)	10.1 (3.8)	9.1 (3.9)
Reduction in motivation			
Group 1	9.5 (2.7)	9.3 (2.8)	8.5 (2.4)
Group 2	9.3 (3.0)	8.9 (2.8)	9.2 (2.9)
Reduction in activity			
Group 1	9.4 (2.7)	9.5 (3.4)	8.7 (2.9)
Group 2	9.0 (3.3)	9.2 (3.4)	8.4 (3.5)
Self-efficacy in communication with health care professionals (PEPPI-5[h], range scale 5-25)			
Group 1	21.3 (2.7)	21.1 (2.2)	21.7 (2.5)
Group 2	21.7 (2.6)	21.8 (2.8)	20.5 (2.9)
Family management (FaMM[i])			
Child's daily life (range 5-25)			
Group 1	17.2 (4.1)	17.4 (4.2)	17.3 (4.2)
Group 2	17.5 (4.4)	18.1 (3.8)	17.7 (4.4)
Condition management ability (range 12-60)			
Group 1	42.6 (4.3)	42.2 (3.7)	44.0 (3.0)
Group 2	42.6 (4.5)	43.2 (4.4)	41.7 (5.0)
Condition management effort (range 4-20)			
Group 1	13.2 (3.0)	12.2 (3.0)	12.3 (3.1)
Group 2	12.6 (3.7)	11.7 (3.9)	11.9 (3.4)
Family life difficulty (range 14-70)			
Group 1	35.2 (9.1)	35.3 (9.1)	32.4 (8.1)

XSL•FO
RenderX

Outcome measure and group[a]	T0[b], mean (SD)	T1[c], mean (SD)	T2[d], mean (SD)
Group 2	34.3 (10.1)	33.5 (10.5)	33.5 (2.1)
Parental mutuality (range 8-40)			
Group 1	32.2 (4.4)	31.9 (4.0)	31.7 (4.1)
Group 2	32.8 (4.0)	32.0 (4.2)	34.5 (4.8)
View on condition impact (range 10-50)			
Group 1	28.4 (4.1)	28.0 (4.5)	27.4 (4.1)
Group 2	27.0 (5.9)	25.9 (5.7)	27.1 (4.2)

[a]Group 1: access to *e-Powered Parents* after T0; Group 2: access to *e-Powered Parents* after T1.

[b]Group 1: n=68; Group 2: n=65.

[c]Group 1: n=35; Group 2: n=51.

[d]Group 1: n=31; Group 2: n=20.

[e]CVS: Child Vulnerability Scale.

[f]PIP: Pediatric Inventory for Parents.

[g]MFI: Multidimensional Fatigue Inventory.

[h]PEPPI-5: Perceived Efficacy in Patient-Physician Interactions.

[i]FaMM: Family Management Measure.

Process Evaluation

Recruitment

In the Methods section, we have described the recruitment of parents for the trial. However, recruitment also included the procedures used to maintain parents' involvement in *e-Powered Parents* (group 1 between T0 and T1 and group 2 after T1). Parents who did not log in to the *e-Powered Parents* received email reminders, including log-in codes and a manual. Furthermore, we regularly posted newsletters on *e-Powered Parents*; parents, who had logged in once, received this newsletter via email.

After group 2 gained access to *e-Powered Parents* as well and *e-Powered Parents* became part of the daily care, JK and MK regularly discussed the program with the parents at the outpatient clinic. Meanwhile, leaflets for parents about *e-Powered Parents* were distributed during the outpatient visits. Apart from the parents, JK also informed the health care professionals involved in the daily care of children with CKD at the Pediatric Nephrology Unit about the program's progress in their weekly multidisciplinary meetings.

Reach

In this study, 133 parents of 89 families participated (response rate, 44.3%), including 31 parents of children with hereditary kidney disease, 17 with nephrotic syndrome, 32 with chronic kidney failure, 9 undergoing dialysis, and 44 with renal transplantation. The community was used by 84% (57/68) of the parents in group 1 (T0-T2) and by 49% of the parents in group 2 (32/65; T1-T2; based on the IP address).

The Dose Received: Exposure

We separately described the dose received for the community and training platform. The majority of parents using the community logged in only once or twice; 22 parents logged in more than 51 times. Although parents in group 2 were also given access to *e-Powered Parents*, the program was most often used between T0 andT1.

The majority of parents visited the information pages on the community. Favorite topics were how the community and the training platform worked (321 and 70 page views, respectively), information about nutrition (47 page views), growing up with CKD (32 page views), and kidney diseases (30 page views; see Figure 3). Specific peritoneal dialysis topics (such as infections), recipes for public holidays, potassium and phosphate binders, and medication for blood pressure were not read at all.

On the interactive part of the community, the most widely read forum messages were about medication (98 views), kidney diseases (95 views), and transplantation (53 views). Parents responded most on topics regarding prednisone use (18 messages) and experiences with school (6 messages). The chat page was visited 109 times, although no one used it. One father wrote a blog about his child's kidney transplantation, which was read 17 times, and 19 conversations took place through private messages. The training platform was used by 64% (85/133) of the parents; 31% (26/85) parents actually followed one or more training modules and 96% logged in only once.

Parents mainly visited the welcome module (n=17), followed by the training module stress management (Figure 4). In 3 of 4 training modules, a decrease was noted in the number of parents per session. Only the parents who followed the training module coping completed the whole training. The third and fourth sessions of the training communication were not visited at all.

The reasons given by parents for not using the community and training platform included no need (because of the stable condition of their child), other priorities, no need for support, or lack of time. Parents also mentioned that their partner was already using the program, or they mentioned a lack of knowledge on how to use the program, which led to difficulties in logging in and setting up an account.

XSL•FO
RenderX

Figure 3. Most widely read information topics on the community between January 2015 and May 2016. CKD: chronic kidney disease; e-PP: e-Powered Parents.

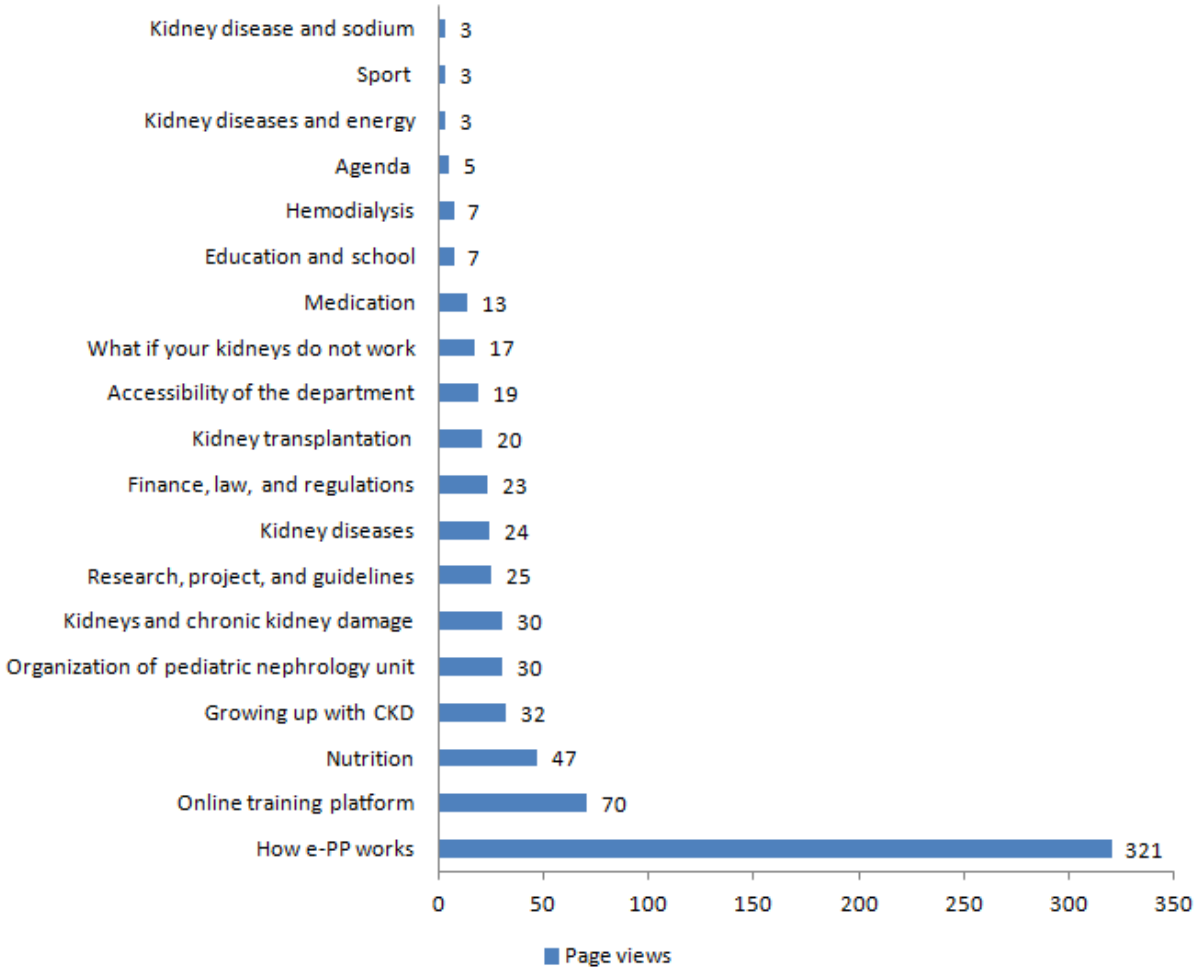
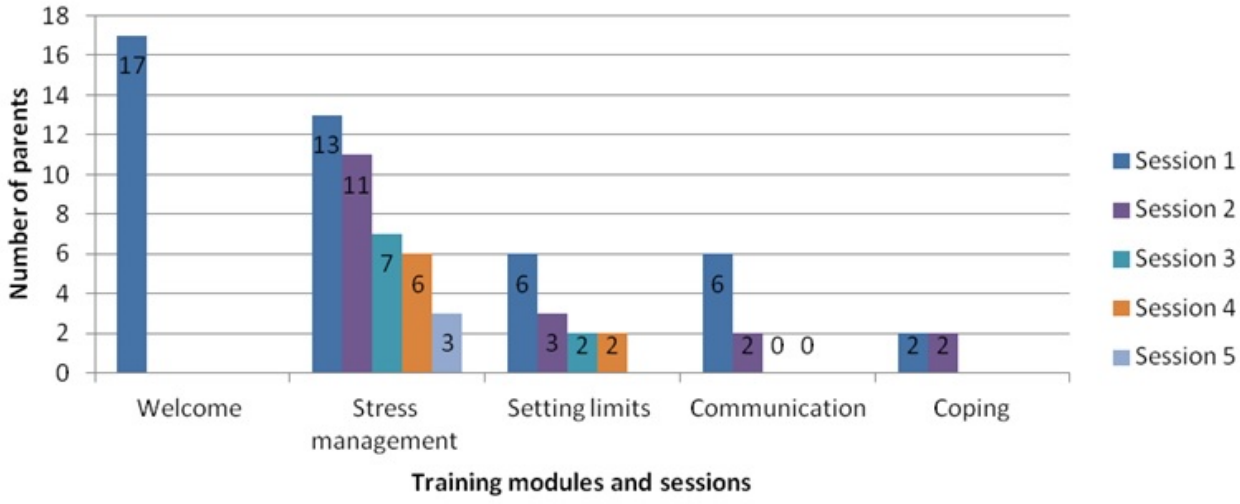


Figure 4. Use of training modules between January 2015 and May 2016.



The Dose Received: Satisfaction

The parents were satisfied with *e-Powered Parents*; they found it to be easy to use (30/36, 83%) and providing relevant information on the information pages and newsletters (31/36, 86% and 31/37, 84%, respectively), which were easy to read (34/36, 94% and 37/37, 100%, respectively). However, some parents did mention that the amount of information was limited (13/36, 36%) or not relevant (5/36, 14%).

The interaction part was underlined by parents as well. However, only a minority of the parents used the Web-based interaction possibilities. Some parents mentioned in the questionnaire that they do not need peer support (because they already use Facebook, among other reasons). Other parents described that

XSL•FO
RenderX

the number of parents participating was limited and that the none response on questions raised was not an incentive to ask further questions.

Parents who did not use the training platform indicated that the modules did not correspond with their request for support. One barrier mentioned by parents was the extra log-in for the Web-based training platform. However, parents who did use the modules were satisfied; they found the instructions to be clear and the modules easy to use.

Context

In the focus group interview, health care professionals indicated that they found it important to keep up with (Web-based) developments. They described *e-Powered Parents* as essential because it provides parents with accessible, reliable, and objective information, which is usually difficult to find. However, the implementation and use of *e-Powered Parents* in daily care (such as by referring parents during their consultation) needs more attention because *"It is not a routine yet."* The extra log-in for parents to use the Web-based training platform was not desirable either. Health care professionals put particular stress on their responsibility for the content of the information, although they also mentioned that it is "*not our core business*," hence, their worry about the continuity of *e-Powered Parents* in the future.

Discussion

Principal Findings

This is the first effect study and process evaluation of an online support program for parents of children with stage I-V CKD. In this feasibility study, the first objective was to analyze the outcome measures most likely to capture the potential benefit. The chosen outcome measures were likely to capture the potential effect, except for self-efficacy in communication with health care professionals, measured by the PEPPI-5 scale; this scale shows ceiling effects and high percentages of parents showing no change between the measurement times. This skewed distribution, meaning that the majority of the parents were already confident in asking questions and discussing their problems with health care professionals, could possibly be explained by the long-standing relationship of the parents with the health care professionals and the small team of health care professionals in pediatric nephrology.

The second objective was to evaluate the potential effectiveness and effect size. Although the parents and health care professionals were enthusiastic about *e-Powered Parents*, no statistically significant effect was found on children's quality of life, parental stress and fatigue, family management, or parents' self-efficacy in communication in both intention-to-treat and per-protocol analyses.

In the process evaluation (third objective), we noticed that the majority of the parents used the community to read the information and that only a few parents actually used the training platform. This could explain the very small effects found in this study: knowledge only does not lead to behavioral changes [31]. Other determinants (such as attitude and self-efficacy) could have been influenced, although strategies to change these determinants were mainly integrated into the training platform and not into the community [14]. On the other hand, parents and health care professionals underlined the importance of evidence-based information on *e-Powered Parents*, providing parents with reliable and up-to-date information about their child's disease and treatment options. The uncertainty about the trustworthiness of the information that parents find on internet is an often-heard reason among parents for not using Web-based information resources [32-35]. However, they do express a need for reliable information to manage their uncertainty, make decisions regarding their child's treatment, and stimulate the dialogue with the care providers of their child [32-34]. The minimal use of the interaction part of the community and the training platform is remarkable. In order to develop this intervention, we conducted an extensive needs assessment consisting of a literature study and 5 focus group interviews with parents [13]. This needs assessment revealed that online peer support was (the most) frequently mentioned need of the parents, and it is often used by parental caregivers for emotional needs [7,8,33]. Even so, the Web-based interaction options of *e-Powered Parents* were not often used. Parents might feel reluctant to share their experiences because of the monitoring by health care professionals. Scharer [34] suggested a continuous clarification of the role of the professionals in online support groups. However, the option to send private messages on *e-Powered Parents* (which could not be monitored by the health care professionals) was not used very much either.

Another possible explanation for the minimal use of *e-Powered Parents* is the heterogeneity of the study population, consisting of fathers and mothers of children with different ages, different kidney diseases and stages (CKD I-V), and treatments. *e-Powered Parents* was not tailored, while the support needs of these heterogeneous groups differed. Swallow et al [36], who developed an online support program for parents of children with stage III-V CKD in the United Kingdom, did not find a significant effect in their feasibility study either.

Strengths and Limitations

We believe that our feasibility study has numerous methodological strengths. By conducting process evaluation, which is recommended when evaluating complex interventions [16], we gained more insight into our recruitment procedures, how many parents we contacted, how actively they engaged, and how satisfied they were. Additionally, we gained more knowledge about the context and how this affected the implementation and use of *e-Powered Parents.* Using the framework of Linnan and Steckler [28], we were able to do this in a structured way. Additionally, using quantitative and qualitative research methods, we increased our understanding of the outcomes, enabling us to improve the intervention for the future [16].

However, some limitations need to be mentioned as well. First, the intervention *e-Powered Parents* consisted of two websites: the community (consisting of the information and interactive part) and the training platform. As described in the Methods section, each website comprised different systems to register and analyze the log-in data and could, therefore, not be interpreted in the same way. Most notably, the use of Google

XSL•FO
RenderX

Analytics (to analyze the community data) was challenging because the log-in data were registered on IP addresses; parents who used *e-Powered Parents* from different locations were consequently registered as different users. Moreover, after T1, when the program became part of the daily care, we were not able to exclude parents who were not a part of the trial. Hence, the presented log-in data of the community could be an overestimation. We tried to correct this by checking the users' account on the community website (did the parents who were part of the trial actually logged in or not?). This problem did not apply to the log-in data of the training platform because this was registered based on the email address of the parents. Parents who did not participate in the trial could easily be excluded from the analysis.

Second, a formal power calculation was not possible in this study, leading us to include as many parents as possible. Although 133 parents filled in the baseline questionnaire, only 38.3% (51/133) parents filled in the questionnaire at T2. This high lost to the follow-up rate could severely comprise the study's validity and reduce the chances of detecting a true effect [37]. Possible explanations for this high lost to follow-up rate are the amount of questions in the questionnaires and an extra unplanned measurement at T2. We decided to conduct an extra measurement, aiming to gain insights into the long-term effects; however, the parents were not aware of this extra measurement at the start of the study.

Finally, we opted for outcomes at a child and parent level, such as quality of life, fatigue, and stress, but not at an organizational level, such as the number of outpatient clinic visits. It would be worthwhile to consider such outcomes in a full-scale RCT.

Conclusions

Parents and health care professionals were very positive about *e-Powered Parents*, and they underlined its importance, yet no significant effect of *e-Powered Parents* was found in this study on the child's quality of life, parental stress and fatigue, parents' self-efficacy in communication, and family management. This could be explained by both the minimal use of *e-Powered Parents* and the heterogeneity of the participants. To continue parents' participation, we recommend a tailored intervention based on the different CKD stages and needs of the parents. Nevertheless, further studies are necessary to determine whether and how online programs can be used to support the parents of children with CKD in the management of their child's disease.

Acknowledgments

We would like to thank the parents and health care professionals for their participation in this study and the Dutch Kidney foundation for their financial support.

Authors' Contributions

JLK, EAMC, and WWG recruited the patients. WWG collected the data and conducted the analysis. BGIvG, GK, NMM, EAMC, and JLK contributed to the interpretation of the results. NM, BvG, and GK supervised the study. WG drafted the manuscript. BGIvG, JLK, NMM, GK, and EC approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Content of e-Powered Parents.

[PDF File (Adobe PDF File), 43KB - jmir_v20i8e245_app1.pdf]

Multimedia Appendix 2

Percentage of parents scoring floor and ceiling effectts.

[PDF File (Adobe PDF File), 37KB - jmir_v20i8e245_app2.pdf]

Multimedia Appendix 3

Percentage of parents showing no change in score between T0, T1, and T2.

[PDF File (Adobe PDF File), 34KB - jmir_v20i8e245_app3.pdf]

Multimedia Appendix 4

Overview of mixed-model analysis intention to treat: T0-T1 and T0-T2.

[PDF File (Adobe PDF File), 30KB - jmir_v20i8e245_app4.pdf]

XSL•FO
RenderX

Multimedia Appendix 5

Overview of mixed-model analysis per protocol: T0-T1 and T0-T2.

[PDF File (Adobe PDF File), 30KB - jmir_v20i8e245_app5.pdf]

Multimedia Appendix 6

CONSORT - EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 1MB - jmir_v20i8e245_app6.pdf]

References

1. Harambat J, van Stralen KJ, Kim JJ, Tizard EJ. Epidemiology of chronic kidney disease in children. Pediatr Nephrol 2012 Mar;27(3):363-373 [FREE Full text] [doi: 10.1007/s00467-011-1939-1] [Medline: 21713524]
2. Tong A, Lowe A, Sainsbury P, Craig JC. Experiences of parents who have children with chronic kidney disease: a systematic review of qualitative studies. Pediatrics 2008 Feb;121(2):349-360. [doi: 10.1542/peds.2006-3470] [Medline: 18245427]
3. Barlow JH, Ellard DR. The psychosocial well-being of children with chronic disease, their parents and siblings: an overview of the research evidence base. Child Care Health Dev 2006 Jan;32(1):19-31. [doi: 10.1111/j.1365-2214.2006.00591.x] [Medline: 16398788]
4. Logan D, Scharff L. Relationships between family and parent characteristics and functional abilities in children with recurrent pain syndromes: an investigation of moderating effects on the pathway from pain to disability. Journal of pediatric psychology 2005 Dec;30(8):698-707. [Medline: 16093517]
5. Eccleston C, Palermo T, Fisher E, Law E. Psychological interventions for parents of children and adolescents with chronic illness. The Cochrane database of systematic reviews 2012;8:1-131. [doi: 10.1002/14651858]
6. Swallow V, Lambert H, Clarke C, Campbell S, Jacoby A. Childhood chronic-kidney-disease: a longitudinal-qualitative study of families learning to share management early in the trajectory. Patient education and counseling 2008 Nov;73(2):354-362.
7. Balkhi AM, Reid AM, McNamara JP, Geffken GR. The diabetes online community: the importance of forum use in parents of children with type 1 diabetes. Pediatr Diabetes 2013 Dec;15(6):408-415. [doi: 10.1111/pedi.12110] [Medline: 24372986]
8. Merkel R, Wright T. Parental self-efficacy and online support among parents of children diagnosed with type 1 diabetes mellitus. Pediatric nursing 2012;38(6):303-308.
9. Fidika A, Herle M, Lehmann C, Weiss C, Knaevelsrud C, Goldbeck L. A web-based psychological support program for caregivers of children with cystic fibrosis: a pilot study. Health and quality of life outcomes 2015 Feb 4;13(11):1-9. [Medline: 25652684]
10. Kirk S, Milnes L. An exploration of how young people and parents use online support in the context of living with cystic fibrosis. Health expectations : an international journal of public participation in health care and health policy 2016 Apr;19(2):309-321. [Medline: 25691209]
11. Shegog R, Bartholomew LK, Sockrider MM, Czyzewski DI, Pilney S, Mullen PD, et al. Computer-based decision support for pediatric asthma management: description and feasibility of the Stop Asthma Clinical System. Health Informatics J 2006 Dec;12(4):259-273. [doi: 10.1177/1460458206069761]
12. Bartholomew L, Parcel GS, Kok G, Gottlieb NH, Fernandez ME. Planning Health Promotion Programs. In: An Intervention Mapping Approach. San Francisco: Jossey-Bass; 2011.
13. Geense W, van GB, Knoll J, Cornelissen E, van Achterberg T. The support needs of parents having a child with a chronic kidney disease: a focus group study. Child: care, health and development 2017;43(6):831-838. [Medline: 28547746]
14. Geense W, van GB, Knoll J, Cornelissen E, Schoonhoven L, Kok G. Online Support Program for Parents of Children With a Chronic Kidney Disease Using Intervention Mapping: A Development and Evaluation Protocol. JMIR research protocols 2016;5(1):1-12 [FREE Full text] [doi: 10.2196/resprot.4837] [Medline: 26764218]
15. Bartholomew LK, Parcel GS, Kok G, Gottlieb NH, Fernandez ME. Chapter nine: Step 6 Planning for evaluation. In: Planning health promotion programs. San Francisco: Jossey-Bass; 2011:509-551.
16. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:1-6 [FREE Full text] [Medline: 18824488]
17. Moore G, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. Bmj 2015;350:1-7.
18. Saunders R, Evans M, Joshi P. Developing a process-evaluation plan for assessing health promotion program implementation: a how-to guide. Health Promot Pract 2005;6(2):134-147. [Medline: 15855283]
19. Houtzager B, Moller E, Maurice-Stam H, Last B, Grootenhuis MA. Parental perceptions of child vulnerability in a community-based sample: Association with chronic illness and health-related quality of life. J Child Health Care 2015;19(4):454-465. [Medline: 24842887]

XSL•FO
RenderX

20. Forsyth B, Horwitz S, Leventhal J, Burger J, Leaf PJ. The child vulnerability scale: an instrument to measure parental perceptions of child vulnerability. Journal of pediatric psychology 1996;21(1):89-101. [Medline: 8820075]
21. Hullmann S, Wolfe-Christensen C, Meyer W, McNall-Knapp R, Mullins LL. The relationship between parental overprotection and health-related quality of life in pediatric cancer: the mediating role of perceived child vulnerability. Quality of life research: an international journal of quality of life aspects of treatment, care and rehabilitation 2010;19(9):1373-1380. [Medline: 20571917]
22. Vrijmoet-Wiersma C, Hoekstra-Weebers J, Margreet DPW, Koopman H, Tissing W, Treffers P, et al. Psychometric qualities of the Dutch version of the Pediatric Inventory for Parents (PIP): a multi-center study. Psychooncology 2010;19(4):368-375. [Medline: 19358212]
23. Streisand R, Braniecki S, Tercyak KP, Kazak AE. Childhood illness-related parenting stress: the pediatric inventory for parents. J Pediatr Psychol 2001;26(3):155-162. [Medline: 11259517]
24. Smets EM, Garssen B, Bonke B, De Haes JC. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. J Psychosom Res 1995 Apr;39(3):315-325. [Medline: 7636775]
25. ten Klooster PM, Oostveen JC, Zandbelt LC, Taal E, Drossaert CH, Harmsen EJ, et al. Further validation of the 5-item Perceived Efficacy in Patient-Physician Interactions (PEPPI-5) scale in patients with osteoarthritis. Patient Educ Couns 2012 Apr;87(1):125-130. [doi: 10.1016/j.pec.2011.07.017] [Medline: 21889864]
26. Maly RC, Frank JC, Marshall GN, DiMatteo MR, Reuben DB. Perceived efficacy in patient-physician interactions (PEPPI): validation of an instrument in older persons. J Am Geriatr Soc 1998 Jul;46(7):889-894. [Medline: 9670878]
27. Knafl K, Deatrick J, Gallo A, Dixon J, Grey M, Knafl G, et al. Assessment of the psychometric properties of the Family Management Measure. Journal of pediatric psychology 2011;36(5):494-505. [Medline: 19451173]
28. Linnan L, Steckler A. Process evaluation for public health interventions and research: an overview. San Francisco: Jossey-Bass; 2002.
29. Hsueh IP, Hsieh CL. Responsiveness of two upper extremity function instruments for stroke inpatients receiving rehabilitation. Clin Rehabil 2002 Sep;16(6):617-624. [doi: 10.1191/0269215502cr530oa] [Medline: 12392336]
30. English C, Hillier SL, Stiller K, Warden-Flood A. The sensitivity of three commonly used outcome measures to detect change amongst patients receiving inpatient rehabilitation following stroke. Clinical Rehabilitation 2006 Jan:52-55. [Medline: 16502750]
31. Fishbein M, Ajzen I. Predicting and changing behavior. In: The reasoned action approach. Abingdon: Taylor & Francis ltd; 2011.
32. Knapp C, Madden V, Wang H, Sloyer P, Shenkman E. Internet use and ehealth literacy of low-income parents whose children have special health care needs. J Med Internet Res 2011;13(3):1-13. [Medline: 21960017]
33. Park E, Kim H, Steinhoff A. Health-related internet use by informal caregivers of children and adolescents: an integrative literature review. J Med Internet Res 2016;18(3):1-10. [Medline: 26940750]
34. Scharer K. Internet social support for parents: the state of science. Journal of child and adolescent psychiatric nursing 2005;18(1):26-35.
35. Walsh A, Hamilton K, White K, Hyde MK. Use of online health information to manage children's health care: a prospective study investigating parental decisions. BMC health services research 2015;15(31):1-10.
36. Swallow V, Carolan I, Smith T, Webb N, Knafl K, Santacroce S, et al. A novel Interactive Health Communication Application (IHCA) for parents of children with long-term conditions: development, implementation and feasibility assessment. Informatics for health & social care 2014:20-46. [Medline: 25119067]
37. Dettori JR. Loss to follow-up. Evidence-Based Spine-Care Journal 2011;2(1):7-10. [Medline: 22956930]

Abbreviations

CKD: chronic kidney disease
CVS: Child Vulnerability Scale
FaMM: Family Management Measure
IP: Internet protocol
MFI: Multidimensional Fatigue Inventory
PEPPI-5: Perceived Efficacy in Patient-Physician Interactions
PIP: Pediatric Inventory for Parents
RCT: randomized controlled trial

XSL•FO
RenderX

Edited by G Eysenbach; submitted 01.12.17; peer-reviewed by H Barahimi, IN Gomez; comments to author 28.12.17; revised version received 30.04.18; accepted 08.05.18; published 01.08.18.

Please cite as:
Geense WW, van Gaal BGI, Knoll JL, Maas NM, Kok G, Cornelissen EAM, Nijhuis-van der Sanden MWG
Effect and Process Evaluation of e-Powered Parents, a Web-Based Support Program for Parents of Children With a Chronic Kidney Disease: Feasibility Randomized Controlled Trial
J Med Internet Res 2018;20(8):e245
URL: http://www.jmir.org/2018/8/e245/
doi:10.2196/jmir.9547
PMID:30068502

©Wytske W Geense, Betsie GI van Gaal, Jaqueline L Knoll, Nienke M Maas, Gerjo Kok, Elisabeth AM Cornelissen, Maria WG Nijhuis-van der Sanden. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 01.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Blended Smoking Cessation Treatment: Exploring Measurement, Levels, and Predictors of Adherence

Lutz Siemer[1,2,3,4], Dipl Psych; Marjolein GJ Brusse-Keizer[3], PhD; Marloes G Postel[2,4], PhD; Somaya Ben Allouch[1], PhD; Angelos Patrinopoulos Bougioukas[2], MSc Psych; Robbert Sanderman[2,5], PhD; Marcel E Pieterse[2], PhD

[1]Research Group Technology, Health & Care, Saxion University of Applied Sciences, Enschede, Netherlands

[2]Centre for eHealth and Well-being Research, University of Twente, Enschede, Netherlands

[3]Medical School Twente, Medisch Spectrum Twente, Enschede, Netherlands

[4]Tactus Addiction Treatment, Enschede, Netherlands

[5]Department of Health Psychology, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

Corresponding Author:
Lutz Siemer, Dipl Psych
Research Group Technology, Health & Care
Saxion University of Applied Sciences
MH Tromplaan 28
Enschede, 7513 AB
Netherlands
Phone: 31 17678025906
Email: l.siemer@saxion.nl

Abstract

Background: Blended face-to-face and Web-based treatment is a promising way to deliver cognitive behavioral therapy. Since adherence has been shown to be a measure for treatment's acceptability and a determinant for treatment's effectiveness, in this study, we explored adherence to a new blended smoking cessation treatment (BSCT).

Objective: The objective of our study was to (1) develop an adequate method to measure adherence to BSCT; (2) define an adequate degree of adherence to be used as a threshold for being adherent; (3) estimate adherence to BSCT; and (4) explore the possible predictors of adherence to BSCT.

Methods: The data of patients (N=75) were analyzed to trace adherence to BSCT delivered at an outpatient smoking cessation clinic. In total, 18 patient activities (eg, using a Web-based smoking diary tool or responding to counselors' messages) were selected to measure adherence; the degree of adherence per patient was compared with quitting success. The minimum degree of adherence of patients who reported abstinence was examined to define a threshold for the detection of adherent patients. The number of adherent patients was calculated for each of the 18 selected activities; the degree of adherence over the course of the treatment was displayed; and the number of patients who were adherent was analyzed. The relationship between adherence and 33 person-, smoking-, and health-related characteristics was examined.

Results: The method for measuring adherence was found to be adequate as adherence to BSCT correlated with self-reported abstinence (P=.03). Patients reporting abstinence adhered to at least 61% of BSCT. Adherence declined over the course of the treatment; the percentage of adherent patients per treatment activity ranged from 82% at the start of the treatment to 11%-19% at the final-third of BSCT; applying a 61% threshold, 18% of the patients were classified as adherent. Marital status and social modeling were the best independent predictors of adherence. Patients having a partner had 11-times higher odds of being adherent (OR [odds ratio]=11.3; CI: 1.33-98.99; P=.03). For social modeling, graded from 0 (=partner and friends are not smoking) to 8 (=both partner and nearly all friends are smoking), each unit increase was associated with 28% lower odds of being adherent (OR=0.72; CI: 0.55-0.94; P=.02).

Conclusions: The current study is the first to explore adherence to a blended face-to-face and Web-based treatment (BSCT) based on a substantial group of patients. It revealed a rather low adherence rate to BSCT. The method for measuring adherence to BSCT could be considered adequate because the expected dose-response relationship between adherence and quitting could be verified. Furthermore, this study revealed that marital status and social modeling were independent predictors of adherence.

Trial Registration: Netherlands Trial Registry NTR5113; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=5113 (Archived by WebCite at http://www.webcitation.org/71BAPwER8).

XSL•FO
RenderX

(J Med Internet Res 2018;20(8):e246) doi:10.2196/jmir.9969

KEYWORDS

blended treatment; smoking; adherence; predictors; tobacco; prevention; cognitive behavioral therapy

Introduction

Smoking Cessation Treatment

As smoking is the leading cause of preventable death, cessation treatment remains pivotal for public health promotion. In past decades, a variety of effective interventions for smoking cessation have become available [1,2], including, more recently, Web-based interventions [3,4] and mobile-phone interventions [5,6]. Currently, both traditional and Web-based modes of delivery are being increasingly merged into blended treatment.

Blended Treatment

Blended treatment is a promising way to deliver behavioral change interventions as it allows combining the strengths of face-to-face treatment (personal attention of a professional, allowing for rich and dynamic synchronous communication) with the unique features of Web-based care (accessibility anytime and anywhere, self-paced asynchronous communication) [7-12]. In the recent past, a growing body of research on blended treatment has emerged [9,13] exploring diverse aspects such as individual and group treatments [14] for a number of health issues, such as depression [15], anxiety [16], and addiction [11]; comparing modes of delivery, such as mainly Web-based [17,18], mainly face-to-face [16,19], and 50-50 [20]; orders of modes of delivery, such as integrated [11] and sequential [18]; and tools used, such as platforms, emails, short message service text messaging, and apps [21,22].

Adherence

While blended treatment may decrease dropout rates [9], it may also increase adherence, which is often low in both Web-based and cessation treatments [23]. Adherence can be defined as the extent to which a person's behavior—taking medication, following a diet, or executing lifestyle changes—corresponds with recommendations from a health care provider [24]. In the context of behavioral change treatments (eg, smoking cessation counseling), issues of adherence are mostly related to premature termination of the treatment and failures to complete between-session tasks and exercises [25]. Low adherence is both an indicator for limited treatment acceptability and a primary determinant of treatment effectiveness [24,26-28] because it leads to suboptimal exposure of patients to evidence-based components of treatment, which in turn—assuming a dose-response relationship—negatively affects treatment outcome [24].

Adherence to Blended Treatment

Until now, little has been known about adherence to blended treatment. In a randomized controlled trial (RCT; N=97), comparing the blended treatment of comorbid mental health and substance use problems with face-to-face treatment, participants were found to be equally able to engage, bond, and commit to treatment [29]. However, in another RCT (N=45), adherence was significantly lower for blended depression treatment than for face-to-face treatment (90.5% vs 95.1%), although both treatments were equally effective [30]. Based on a small sample (N=9) in another blended depression treatment trial, adherence rates were considered promising (ie, 5 of 7 patients who started blended treatment completed 90% of it) [20]. This initial evaluation study also revealed that discontinuing blended treatment appeared to be unrelated to the blended nature of the treatment and, unsurprisingly, having internet access and a functional computer at home was indispensable. Finally, a case report on blended treatment for antepartum depression also showed good adherence [31].

In the context of smoking cessation, to the best of our knowledge, adherence to blended treatment has not been assessed. For smoking cessation treatment in general, adherence rates widely vary between studies (5%-96%), which can be explained by differences in the interventions used, adjunctive support, and populations studied [24]. Typically, in a smoking cessation treatment, adherence rapidly declines over the initial weeks of treatment, followed by a more gradual decrease in the later stages, resulting in rather low adherence rates (<40%) [24].

Predictors of Adherence

As adherence is pivotal for treatment effectiveness [24], predicting adherence becomes relevant because it may increase treatment efficacy. Adherence, in general, is determined by provider behaviors, health system factors, and personal characteristics [24]. In particular, the latter have been examined as predictors of adherence to traditional interventions [32]. However, similar studies on adherence to blended treatment appear to be lacking. Within the context of smoking cessation treatment—including both face-to-face and Web-based treatments—several person-, smoking-, and health-related predictors of adherence have been examined. The likelihood of being adherent increases with a higher age [33,34], male gender [34], higher internet skills [35,36], negative attitude toward smoking and higher motivation to quit at baseline [37,38], higher self-efficacy at baseline [38], early success in quitting after the start of the treatment [26,33,39], and lower nicotine dependency at baseline and fewer withdrawal symptoms after quitting [34,37]. The question arises whether these predictors apply to blended treatment as well.

Measurement of Adherence

To examine the predictors of adherence, valid measurement of adherence becomes a prerequisite. Taking into account both the novelty and diversity of blended treatments, one can understand that established measures for adherence to blended treatment in particular are still lacking. Therefore, for the purpose of this study, a customized measure was constructed based on a combination of parameters used for face-to-face and Web-based interventions. In face-to-face treatment, adherence is often operationalized as completion of tasks assigned during the treatment or the number of completed or attended treatment sessions [40]. In Web-based treatment, the measures of

XSL•FO
RenderX

adherence often comprise log-ins to programs, module completion, time spent online, (self-reported) completion of predefined activities such as use of an Web-based tool, posts made, pages viewed, replies to emails, forum visits, or print requests made [23]. Aiming to increase precision and accuracy, the adherence measure developed for this study was primarily based on objective or direct adherence indicators, such as whether or not a patient attended a face-to-face session, responded to a counselor's message, or used a certain Web-based treatment tool (eg, "goal setting" or "think differently"). Using observable and digitally traceable patient activities, limitations in terms of reliability and validity of self-report data [41] can be largely avoided.

Thresholds Defining "Adequate" and "Inadequate" Adherence

Finally, in addition to measuring adherence as a continuous variable, applying a categorical measure based on a threshold for "adequate" and "inadequate" adherence may be useful for clinical purposes [32]. However, justifications for the operationalizations of thresholds are a common issue; a recent review [42] on adherence to eHealth revealed that 28 of 62 studies described thresholds, but only 6 reported a justification for the threshold. In line with Carolan [43], in this study, we have defined the threshold in relation to the treatment outcome (ie, quitting smoking). To the best of our knowledge, neither for blended smoking cessation treatment (BSCT), in particular, nor for blended treatment, in general, have the thresholds for adherence been explored until now. In the context of smoking cessation, intervention thresholds for classifying participants as adherent or nonadherent range from 75% to 100% use of intervention components offered [24].

Objectives

In view of all that has been mentioned so far, the objectives of this exploratory study were as follows:

1. To develop a method to measure adherence to a BSCT by selecting traceable activities of the patients and to determine whether this method is adequate by comparing the degree of adherence with the quitting success (ie, verifying the expected dose-response relationship between adherence and quitting).
2. To define an adequate degree of adherence to be used as a threshold to detect adherent patients by examining the minimum degree of adherence of the patients who reported abstinence.
3. To estimate adherence to BSCT in three ways: by calculating the number of adherent patients for certain treatment activities; by displaying how the degree of adherence changes over the course of the treatment; and by reporting the proportion of adherent patients according to the threshold for adherence.
4. To explore the possible predictors of adherence to BSCT by examining the relationship between being adherent or nonadherent and 33 person-, smoking-, and health-related characteristics assessed at baseline.

Methods

Study Participants

In this study, we used a subset of an RCT on the effectiveness of BSCT versus face-to-face treatment as usual [11]. Patients were referred to the outpatient smoking cessation clinic at the Medical Spectrum Twente hospital (Enschede or The Netherlands) by the treating physicians of the hospital or by the patients' general practitioners. Inclusion criteria included (1) being at least 18 years old, (2) currently smoking (at least one cigarette a day), (3) having access to email and internet, (4) being able to read and write Dutch. For the adherence analysis, we used the RCT data of the first 75 patients of the BSCT who attended an initial treatment session from May 2015 to December 2016. In line with the Dutch Medical Research Ethics Committee (MREC) guidelines, the study was approved by the accredited MREC Twente (P14-37/NL50944.044.14). Before initiation, the study was registered in the Dutch Trial Registration (NTR5113). All patients had to sign an informed consent form before they were randomized.

Blended Smoking Cessation Treatment

The BSCT examined in this study is a combination of face-to-face treatment and Web-based sessions blended into one integrated smoking cessation treatment, which is delivered in routine care settings. BSCT consists of 5 face-to-face sessions at the outpatient clinic and 5 Web-based sessions delivered via the Web-based treatment platform. Table 1 shows the order, timing, main features, and mode of delivery of the sessions.

The following are the characteristic features of BSCT:

1. High-intensity treatment: BSCT comprises 10 sessions (20 minutes each, except the first one, which is of 50 minutes); it covers the majority of evidence-based behavior change techniques [44]. It is derived from the Dutch Guideline Tobacco Addiction [45], fulfilling the requirements of the Dutch care module for smoking cessation [46]; the counselors are registered in the Dutch quality register of qualified smoking cessation counselors.
2. Supports three quitting strategies: At the start, patients choose to (1) stop at once, (2) change gradually by increasing the number of daily activities that are performed smoke-free, or (3) decrease smoking at regular intervals (scheduled smoking reduction, eg, 100%->75%, 75%->50%). The chosen quitting strategy does not influence the course of the treatment in general, that is, the order, pace, duration, and intensity are the same for all strategies.
3. A 50-50 balance between face-to-face and Web-based treatments: The focus of the treatment is neither on face-to-face nor on Web-based treatment; in addition, the treatment is constantly alternating and there is interactive use of face-to-face and Web-based treatments.

A detailed description of the treatment can be found in the protocol article of the RCT [11].

XSL•FO
RenderX

Table 1. Order, timing, main features, and mode of delivery of blended smoking cessation treatment.

Session	Week	Main features	Mode of delivery
1	1	Goal setting, prompt smoking diary, measure CO[a]	Face-to-face
2	3	Measures for self-control	Web-based
3	5	Dealing with withdrawal	Face-to-face
4	7	Breaking habits	Web-based
5	9	Dealing with triggers	Face-to-face
6	11	Food for thought	Web-based
7	14	Think differently, measure CO	Face-to-face
8	18	Do differently	Web-based
9	22	Action plan, measure CO	Face-to-face
10	26	Closure	Web-based

[a]CO: carbon monoxide.

Data Collection

Patients' Characteristics and Smoking Status

As part of the RCT, 33 person-, smoking-, and health-related characteristics were assessed with the intake measurement using a Web-based questionnaire. A detailed description of these characteristics is available in the protocol article of the RCT [11]. In addition, both the 3-month and 6-month follow-up measurements of the trial were used to examine the self-reported smoking status.

Measuring Adherence to Blended Smoking Cessation Treatment

Two data sources were screened to determine which treatment activities of the patients could be traced after the first treatment session:

1. The patients' record from the Web-based treatment platform. These records provided, on the one hand, a section where patients and counselors communicated via messages and, on the other hand, a section with therapeutic Web-based tools that were used by the patient. Both sections interact with each other. Here is a typical example of this interaction: The counselor sends a message with instructions to use a therapeutic Web-based tool, such as "goal setting." With this message, the counselor also unblocks the goal setting Web-based tool and sets a date for executing this task. After receiving this message, the patient uses the unblocked Web-based tool to elaborate goals. What the patient fills in can then be reviewed by the counselor, who also has access to the tool. The counselor then usually responds to what the patient filled in via a message and leads into the following face-to-face session.
2. Patients' records from the outpatient cessation clinic, which were maintained by the counselors. These records provided additional information about patients' activities, such as adhering to a stop-date or measurement of CO.

After comparing the treatment manual with the data available in the two data sources, 18 activities of patients were selected to score adherence after the first treatment session. The selection of activities was based on the following three considerations:

1. The activity had to refer directly to a relevant evidence-based behavior change technique [44] (eg, goal setting, action plan) that represented the main feature of the sessions, so that adherence to each of the 10 sessions of the treatment was separately measurable.
2. The activities had to trace both face-to-face and Web-based behaviors of patients (eg, attending face-to-face treatment sessions as in "Think differently [face-to-face]" or completion of predefined Web-based tasks as in "Think differently [Web]"), so that adherence to the constant interaction between face-to-face and Web-based treatments—and by this, the blended nature of BSCT—was covered.
3. The data used had to be objective (eg, receiving a message, unblocking a Web-based tool, filling in a minimal number of data in a Web-based tool) to avoid the limitations of self-reported data [41].

The majority of the selected activities reflected the course of the blended treatment, starting with "Goal setting (face-to-face)" at the end of session 1 and finalizing with "Action plan (Web)" in session 10. Three activities were not session dependent as they had to be executed several times ("Measurement of CO [face-to-face]") or across several sessions ("smoking diary [days; Web]"; "smoking diary [moments; Web]"). A detailed description of these activities showing how each activity was operationalized to indicate adherence or nonadherence is provided in Table 2.

Based on data sources, for each patient, adherence to each of the 18 activities was assessed and graded adherent or nonadherent by trained research assistants. Finally, for each patient, an adherence score from 0 (adherent to no activity after the first treatment session) to 18 (adherent to all activities) as well as subscores for Web-based versus face-to-face and session-dependent versus session-independent activities were available.

XSL•FO
RenderX

Table 2. Activities, operationalization, and patients' adherence to blended smoking cessation treatment (N=75).

Activity and mode of delivery	Operationalization	Adherent patients, n (%)
Session-dependent activities		
Session 1: Goal setting		
Face-to-face treatment	The patient was introduced to "goal setting" and received a message with the prompt to use the Web-based goal setting tool.	62 (82)
Web-based treatment	The patient used the Web-based goal setting tool.	44 (58)
Session 2: Measures for self-control		
Face-to-face treatment	The patient was introduced to "measure for self-control" and received a message with information about measures for self-control.	39 (52)
Web-based treatment	The patient reacted to the measure for self-control message. (Note: a response was not obligatory; patients could read only without responding)	27 (36)
Session 3: Dealing with withdrawal		
Face-to-face treatment	The patient was introduced to "dealing with withdrawal" and received a message with information about dealing with withdrawal.	31 (41)
Web-based treatment	The patient reacted to the dealing with withdrawal message. (Note: a response was not obligatory; patients could read only without responding)	20 (26)
Session 5: Dealing with tempters		
Web-based treatment	The patient received a message with information about dealing with tempters.	22 (29)
Session 6: Food for thought		
Face-to-face treatment	The patient was introduced to "food for thought" and received a message with information about food for thought.	17 (23)
Web-based treatment	The patient reacted to the food for thought message (Note: a response was not obligatory; patients could read only without responding)	8 (11)
Session 7: Think differently		
Face-to-face treatment	The patient was introduced to "think differently" and received a message with the prompt to use the Web-based think differently tool.	14 (19)
Web-based treatment	The patient used the Web-based think differently tool.	14 (19)
Session 8: Do differently		
Face-to-face treatment	The patient was introduced to "do differently" and received a message with the prompt to use the Web-based do differently tool.	12 (16)
Web-based treatment	The patient used the Web-based do differently tool.	14 (19)
Session 9: Action plan		
Face-to-face treatment	The patient was introduced to "action plan" and received a message with the prompt to use the Web-based action plan tool.	13 (17)
Session 10: Action plan		
Web-based treatment	The patient used the Web-based action plan tool.	13 (17)
Session-independent activities		
Measurement of CO[a] (face-to-face treatment)	The counselor reported at least 2 CO measurements.	34 (45)
Smoking diary (days; Web)	The patient used the Web-based smoking diary tool registering cigarettes smoked for at least 3 days.	26 (35)
Smoking diary (moments; Web)	The patient used the Web-based smoking diary tool describing at least 3 moments with an urge to smoke.	31 (41)

[a]CO: carbon monoxide.

XSL•FO
RenderX

Patients' Characteristics

Patients' person-, smoking-, and health-related characteristics were reported as means with SDs for normally distributed continuous variables and as medians with interquartile ranges (IQRs) for not-normally distributed continuous variables. Categorical variables were reported as numbers with corresponding percentages.

Dose-Response Relationship Between Adherence and Quitting

To explore the association between the degree of adherence and quitting success, the median number of adherence activities was compared between quitters (based on self-reported smoking status) and smokers at 3 and 6 months after the start of the treatment and tested using Mann-Whitney-U test.

Threshold to Detect Adherent Patients

To define a threshold for an adequate degree of adherence, the minimum number of adherence activities of quitters (6 months after the start of the treatment) was examined and displayed as number (%) of activities for BSCT overall and separately for both face-to-face and Web-related activities.

Adherence to Blended Smoking Cessation Treatment

Adherence per Activity

To examine the degree of adherence to each of the BSCT activities, the number (%) of patients fulfilling each activity was examined and displayed separately.

Adherence Over the Course of the Treatment

To show changes in adherence over the course of the treatment, the number of patients who were adherent to session-dependent activities was displayed in a bar chart.

Adherence Based on the Threshold

The number (%) of patients who were adherent and nonadherent based on the determined threshold to detect adherent patients was cross-tabulated for both the face-to-face and Web-based modes. The number of adherent patients was compared between face-to-face and Web-based treatments and tested using Pearson chi-square test.

Predictors of Adherence

To identify the predictors of adherence within the 33 person-, smoking-, and health-related patient characteristics, *t* tests or Mann-Whitney-U tests were performed as appropriate for continuous variables; Pearson chi-square or Fisher's exact test were performed for categorical variables. Variables with a significance $P<.15$ were considered as the candidates for multivariate logistic regression analyses and were entered after checking for multicollinearity. Forward stepwise logistic regression analyses were performed. Variables were entered step for step and were eliminated when the model fit was not significantly increased by adding the variable (based on −2 log likelihood). In case of multicollinearity, the variable with the best model fit was selected for logistic regression analyses.

Statistical Analysis

All analyses were performed using SPSS version 24.

Results

Patients' Characteristics

Patients' person-, smoking-, and health-related characteristics are shown in Multimedia Appendix 1.

Dose-Response Relationship Between Adherence and Quitting

A subsample of patients' self-reported smoking status at 3 months (n=25) and 6 months (n=17) after the start of the treatment was available to explore the relationship between adherence and quitting. As can be seen by the numbers of activities for adherence tabulated in Table 3, there is a dose-response relationship between adherence to BSCT and self-reported smoking status at 3 and 6 months after the start of the treatment. The median number of activities for adherence is significantly higher among quitters at 6 months after the start of the treatment ($P=.03$).

Threshold to Detect Adherent Patients

Patients with self-reported abstinence at 6 months after the start of the treatment (n=17) were adherent to at least 61% (11/18) activities of BSCT. Because BSCT is built on a 50%-50% relation for both modes of delivery, a 61% threshold was applied to both modes of delivery to detect adherent patients (ie, patients were defined as adherent if 5 of the 8 face-to-face activities as well as 6 of the 10 Web-based activities were fulfilled).

Adherence to Blended Smoking Cessation Treatment

Adherence per Activity

Table 2 shows the number (%) of adherent patients for each activity. Of all, 17.3% (13/75) patients were adherent to none of the activities, which indicates that these patients did not fully complete the first face-to-face session that closes with the patient being introduced to "goal setting" and receiving a message with the prompt to use the Web-based goal setting tool. None of the patients were adherent to all activities.

Adherence per Activity Over the Course of Treatment

To show the change in adherence over the course of the treatment, the number of adherent patients per session-dependent activity (excluding the 3 session-independent activities) is displayed in a bar chart (Figure 1). The number of adherent patients was highest at the start of the treatment (62/75, 82%, patients adherent to "Goal setting [face-to-face]"). Adherence then decreased to 11% (8/75) for the activity "Food for thought (Web)," which is in the 6thtreatment session (Table 2), staying at a low level for the rest of the treatment (varying between 12/75, 16%, and 14/75, 19%).

XSL•FO
RenderX

Table 3. Adherence to blended smoking cessation treatment and self-reported smoking status at 3 and 6 months after the start of the treatment.

Time point	Median number of adherence activities (IQR[a])		*P* value
	Quitter	Smoker	
3 months (n=25)	14.5 (9.5-15.8)	9.0 (6.0-14.0)	.08
6 months (n=17)	15.0 (11.8-16.0)	9.0 (6.5-14.5)	.03

[a]IQR: interquartile range.

Figure 1. Adherence over the course of the treatment. f2f: face-to-face.

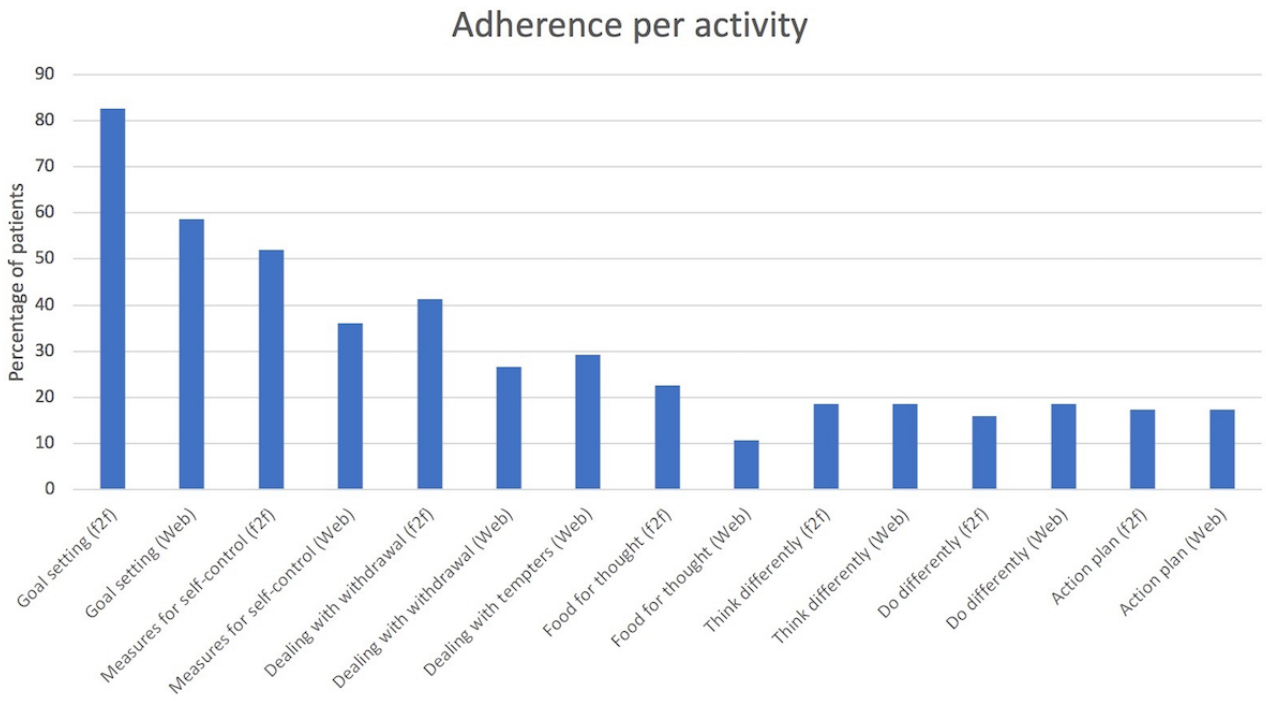


Table 4. Adherence to blended smoking cessation treatment based on the 61% threshold (N=75). Percentages are based on the overall N value.

Face-to-face	Web-based		
	Adherent n (%)	Nonadherent n (%)	Total n (%)
Adherent	14 (18)	5 (7)	19 (25)
Nonadherent	3 (5)	53 (70)	56 (75)
Total	17 (23)	58 (77)	75 (100)

Adherence Based on the Threshold

Based on the 61% threshold for both modes of delivery, 18% (14/75) patients were adherent to both modes of delivery and, therefore, to BSCT as a whole (Table 4). Of all, 25% (19/75) patients were adherent to the face-to-face treatment compared with 23% (17/75) adherent to the Web-based treatment (*P*=.70); 70% (53/75) patients were nonadherent to both modes of delivery. Furthermore, 5% (3/75) patients were adherent to the Web-based mode but not to the face-to-face mode, while 7% (5/75) patients were adherent to the face-to-face mode but not to the Web-based mode.

Predictors of Adherence

The 33 person-, smoking-, and health-related characteristics of the 75 patients, stratified by the 61% adherence or nonadherence threshold, are shown in Multimedia Appendix 1.

The following predictors were univariately associated with adherence (*P*<.15): sex (male = more adherent); marital status (with partner = more adherent); main income (wage or own company = more adherent); social modeling (less smokers in the social environment = more adherent) and use of alcohol (higher alcohol consumption = more adherent); use of other medication (user = more adherent); health-related complaints (as per Maudsley Addiction Profile Health Symptoms Scale [MAP HSS]), smoking-related complaints, and health- and smoking-related complaints (less complaints = more adherent). Due to multicollinearity between health-related complaints (MAP HSS), smoking-related complaints, and health- and smoking-related complaints, only the variable with the best model fit could be included in multivariate regression analysis, which was health- and smoking-related complaints.

XSL•FO
RenderX

Multivariate regression analyses revealed that marital status and social modeling—accounting for 25% of the variance (Nagelkerke R Square)—were independent predictors of whether patients were adherent to BSCT. Patients having a partner had 11 times higher odds of being adherent, although the extremely wide CI indicates considerable uncertainty for this odds ratio (OR=11.3; CI: 1.33-98.99; *P*=.03). For social modeling, graded from 0 (=partner and friends are not smoking) to 8 (=both partner and nearly all friends are smoking), each unit increase was associated with 28% lower odds of being adherent (OR=0.72; CI: 0.55-0.94; *P*=.02).

Discussion

Principal Findings

This study is the first to explore adherence to a blended face-to-face and Web-based smoking cessation treatment (BSCT). Based on a substantial group of participants (N=75), the study revealed a rather low adherence rate to BSCT among this sample of outpatients in a regional hospital in the Netherlands. Applying a 61% threshold for adherence to both face-to-face and Web-based modes of delivery, only 18% (14/75) of the participants were classified as adherent. A dose-response relationship was found between the level of adherence and the likelihood of quitting smoking, corroborating the adequacy of the adherence measure developed for this purpose. Furthermore, several baseline characteristics, in particular marital status and social modeling of nonsmoking, were found to be predictive of adherence to BSCT.

So far, the established measures for adherence to blended treatment are still lacking. Therefore, the first aim of this study was to develop an adequate method for measuring adherence to BSCT and to determine whether this method is adequate. Using data from the hospital's patient records and data logged by patients and counselors on the Web-based treatment platform, a composite score of adherence to 18 distinct treatment activities from both the modes of delivery was calculated. By relying on observed behavior, an objective approach was applied, thus, avoiding bias due to self-report [41]. Adequacy of the adherence measure was confirmed by the observed dose-response relationship between adherence and likelihood of quitting, which is consistent with smoking cessation literature [24,26-28].

A justified threshold for "adequate" and "inadequate" adherence may be useful for clinical purposes [32]. Hence, the second aim of the study was to define an adequate level of adherence to be used as a threshold to detect patients adherent to BSCT. We used a 61% threshold, which is derived from the minimum level of adherence of the patients reporting abstinence. Although this 61% threshold is considerably lower than the commonly applied thresholds of >80% in the previous studies [33,47], it seems realistic if the design of BSCT is taken into consideration. As BSCT is designed as a 10-session, high-intensity treatment, its completion might pose a challenge to the patients of the outpatient clinic. Furthermore, BSCT fosters quitting around month 3 of the treatment while focusing on stabilizing abstinence in the remaining 3 months. Although patients are informed that relapse prevention is pivotal in the later parts of treatment, there seems to be a tipping point at around 60% of the treatment course at which some of the patients who have been successful until then, decide to abandon treatment, thinking themselves "over the hump." This may also be explained by the bidirectional causality between quitting and adherence: early quitting success predicts adherence [33], while, in turn, adherence predicts (long-term) abstinence [24,26-28].

Because, until now, little has been known about adherence to blended treatment, the third aim of the study was to estimate adherence to BSCT. Not surprisingly, we found a notable decrease in adherence over the course of the treatment. Only a small proportion of patients (12/75, 16%, to 14/75, 19%) was adherent to the last 4 of the 10 BSCT sessions. This is in line with the <40% adherence rates for the later stages of smoking cessation treatment in general [24].

Based on the 61% threshold derived from self-reported abstinence, we found that only 18% (14/75) of the patients were adherent to BSCT. This adherence rate seems to be notably lower than the adherence rates reported in smoking cessation literature, which—while applying even higher thresholds—range from, for example, 50% for Web-based treatment [47] to 70% for face-to-face smoking cessation treatment [34]. Explanations for the low adherence may be negative user experience (eg, due to too demanding or time-consuming features) or lack of persuasive elements [48] in the Web-based treatment.

We also found no significant difference in adherence to Web-based and face-to-face modalities. This is in line with earlier findings, showing that discontinuing the blended treatment is unrelated to the blended nature of the treatment [20].

Since predicting adherence may increase treatment efficacy, the fourth aim of the study was to find the predictors of adherence to BSCT. We found being adherent to be significantly ($P \leq .05$) related to male gender, having a partner, wage or own company as the source of main income, and having a low number of smokers in the social environment. Except for gender [34], we could not confirm predictors earlier reported in the literature, such as age [33,34], internet skills [35,36], attitude toward smoking and motivation to quit [37,38], self-efficacy [38], and nicotine dependency and withdrawal symptoms [34,37]. Furthermore, we found potential predictors of adherence to BSCT ($P \geq .05$-.15) not earlier reported in the context of smoking cessation treatment, namely use of alcohol, use of other medication, and health- and/or smoking-related complaints. Two predictors—having a partner and having a low number of smokers in the social environment—were found to be unique, independent predictors in the multivariate predictor model. Having a partner gave higher odds of being adherent than living alone. Although not earlier reported in the context of smoking cessation, this seems to be consistent with a meta-analysis about marital status and adherence to medical treatment in general, which found 1.27 (CI: 1.12-1.43) times higher odds of adherence in married than in unmarried patients [49]. The most common explanation for this effect of marital status is the social support that a partner may provide to the patient. However, within this study, we did not find a predictive effect of social support, which included the partner as an important other. It should be noted, however, that the measure for social support was specified for

XSL•FO
RenderX

smoking cessation and not for adherence. Future research is needed to clarify these inconsistencies. The remarkably high OR of 11 for marital status should be interpreted cautiously because the CI for this ratio was very wide, probably due to limited statistical power in our data. For social modeling, graded from 0 (=partner and friends are not smoking) to 8 (=both partner and nearly all friends are smoking), each unit increase gave 28% lower odds of being adherent. Although not previously reported as a predictor of adherence, social modeling is well known as one of the main determinants of relapse [50-52]. Apparently, patients with more smokers in their environments have a higher probability to relapse; consequently, they drop out of treatment, resulting in lower adherence. This would also be in line with the bidirectional causality between quitting and adherence mentioned above. Looking at social modeling as a relevant predictor of adherence to BSCT, as a clinical implication, the treatment could offer more normative influence (eg, showing videos of peers praising quit attempts of others to patients who report high social modeling of smoking in their environment).

Limitations

A major limitation of this study is that the results found for BSCT have not been compared with either a face-to-face only or a Web-based only treatment. Therefore, it remains undecided whether the results are specific for the blended nature of the treatment or rather for smoking cessation treatment in general.

In addition, the statistical power in this study—especially for predictor analysis—is rather low. This implies, in particular, that false-negative results may have occurred for predictors with small to medium observed effect sizes. As the purpose of this study is primarily exploratory, the risk of false-negative findings was reduced by also considering marginally significant effects as "potential" predictors. However, caution should be taken here, and replication of these findings in future studies is needed.

Furthermore, the adequacy of the adherence measure can be questioned on three aspects. First, 3 of the 10 Web-based activities, "Measure for self-control," "Dealing with withdrawal," and "Food for thought," involved written messages sent by counselors prompting a response by, for example, asking a question at the end of the message. Only when a patient responded and left a traceable action, the activity was scored as completed. Yet, patients may still have read the received messages without explicitly responding to the counselor. This may have led to an underestimation of adherence because the patient may have been exposed to some components without notable traces. Second, the data from the hospital's patient records and the Web-based treatment platform partly depend on activities of the counselors because they maintain these patients' records and also act on the Web-based treatment platform (eg, sending messages to patients, unblocking treatment tools). One can argue, therefore, that adherence measurement is affected by treatment fidelity of counselors. Fidelity of counselors was not evaluated—a common omission in adherence studies [53]. Given that, in this particular case, BSCT was new to the counselors, and their potential unfamiliarity with BSCT may have led to not following the treatment protocol strictly. This may also be increased by therapist drift [54], a common phenomenon in face-to-face cognitive behavioral therapy, which involves a shift from "doing therapies" to "talking therapies." Third, although striving for objective measures, whether an activity was fulfilled by a patient was doubtful in some cases; for example, the use of the Web-based goal setting tool was questionable if only one word to set a goal (eg, "health") was sufficient. These cases had to be discussed in the research team, which adds a subjective factor to the measurement. However, this was only the case for a very small number of participants, mainly in the starting phase of data collection.

Finally, the comparability of the observed adherence rates and thresholds across studies is limited due to variety in treatment demands and operationalizations of adherence [32].

Implication for Future Work

To assess whether the results found in this study are specific for the blended nature of BSCT, the results should be compared with either face-to-face only or Web-based only treatment. Furthermore, future adherence studies should preferably include a measure of fidelity as well, enabling analyses that control for provider-mediated effects on adherence. In addition, from a clinical perspective, the question arises as to how the low adherence rate to BSCT can be increased by, on the one hand, targeting patients to predictive characteristics at baseline or, on the other hand, redesigning BSCT to better accommodate current population characteristics and needs.

Acknowledgments

Both the Research Group Technology, Health & Care at Saxion University of Applied Sciences, and the Centre for eHealth and Well-being Research at University of Twente funded this research. We would like to thank the directors and the personnel of both Medisch Spectrum Twente and Tactus Addiction Treatment.

Authors' Contributions

LS, MEP, MGJBK, MGP, and AP identified the study questions and designed the study and its measuring instruments. LS is the principal investigator and wrote the first draft of this manuscript. LS, MEP, MGJBK, MGP, AP, RS, and SBA edited this manuscript. LS, MEP, AP, and RS revised the manuscript. All authors approved the final version of this manuscript for publication.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1

Predictors of adherence or nonadherence to blended smoking cessation treatment.

[PDF File (Adobe PDF File), 71KB - jmir_v20i8e246_app1.pdf]

References

1. Stead LF, Koilpillai P, Fanshawe TR, Lancaster T. Combined pharmacotherapy and behavioural interventions for smoking cessation. Cochrane Database Syst Rev 2016 Mar 24;3:CD008286. [doi: 10.1002/14651858.CD008286.pub3] [Medline: 27009521]
2. Lancaster T, Stead LF. Individual behavioural counselling for smoking cessation. Cochrane Database Syst Rev 2017 Mar 31;3:CD001292. [doi: 10.1002/14651858.CD001292.pub3] [Medline: 28361496]
3. Civljak M, Stead LF, Hartmann-Boyce J, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2013 Jul 10(7):CD007078. [doi: 10.1002/14651858.CD007078.pub4] [Medline: 23839868]
4. Taylor GMJ, Dalili MN, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2017 Sep 04;9:CD007078. [doi: 10.1002/14651858.CD007078.pub5] [Medline: 28869775]
5. Whittaker R, Borland R, Bullen C, Lin RB, McRobbie H, Rodgers A. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2009 Oct 07(4):CD006611. [doi: 10.1002/14651858.CD006611.pub2] [Medline: 19821377]
6. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016 Apr 10;4:CD006611. [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
7. Wentzel J, van der Vaart R, Bohlmeijer ET, van Gemert-Pijnen JE. Mixing Online and Face-to-Face Therapy: How to Benefit From Blended Care in Mental Health Care. JMIR Ment Health 2016 Feb 09;3(1):e9 [FREE Full text] [doi: 10.2196/mental.4534] [Medline: 26860537]
8. Barak A, Hen L, Boniel-Nissim M, Shapira N. A Comprehensive Review and a Meta-Analysis of the Effectiveness of Internet-Based Psychotherapeutic Interventions. Journal of Technology in Human Services 2008 Jul 03;26(2-4):109-160. [doi: 10.1080/15228830802094429]
9. Erbe D, Eichert HC, Riper H, Ebert DD. Blending Face-to-Face and Internet-Based Interventions for the Treatment of Mental Disorders in Adults: Systematic Review. J Med Internet Res 2017 Sep 15;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]
10. Postel MG, Witting M, van Gemert-Pijnen JE. Blended behandeling in de geestelijke gezondheidszorg. Kwartaaltijdschrift voor directieve therapie en hypnose 2013;33(3):210-221.
11. Siemer L, Pieterse ME, Brusse-Keizer MGJ, Postel MG, Ben Allouch S, Sanderman R. Study protocol for a non-inferiority trial of a blended smoking cessation treatment versus face-to-face treatment (LiveSmokefree-Study). BMC Public Health 2016 Dec 24;16(1):1187 [FREE Full text] [doi: 10.1186/s12889-016-3851-x] [Medline: 27881108]
12. van der Vaart R, Witting M, Riper H, Kooistra L, Bohlmeijer ET, van Gemert-Pijnen LJ. Blending online therapy into regular face-to-face therapy for depression: content, ratio and preconditions according to patients and therapists using a Delphi study. BMC Psychiatry 2014 Dec 14;14:355 [FREE Full text] [doi: 10.1186/s12888-014-0355-z] [Medline: 25496393]
13. Kloek C, Bossen D, de Bakker DH, Veenhof C, Dekker J. Blended Interventions to Change Behavior in Patients With Chronic Somatic Disorders: Systematic Review. J Med Internet Res 2017 Dec 21;19(12):e418 [FREE Full text] [doi: 10.2196/jmir.8108] [Medline: 29269338]
14. Schuster R, Leitner I, Carlbring P, Laireiter A. Exploring blended group interventions for depression: Randomised controlled feasibility study of a blended computer- and multimedia-supported psychoeducational group intervention for adults with depressive symptoms. Internet Interventions 2017 Jun;8:63-71 [FREE Full text] [doi: 10.1016/j.invent.2017.04.001]
15. Kooistra LC, Wiersma JE, Ruwaard J, van Oppen P, Smit F, Lokkerbol J, et al. Blended vs. face-to-face cognitive behavioural treatment for major depression in specialized mental health care: study protocol of a randomized controlled cost-effectiveness trial. BMC Psychiatry 2014 Oct 18;14:290 [FREE Full text] [doi: 10.1186/s12888-014-0290-z] [Medline: 25326035]
16. Bruinsma A, Kampman M, Exterkate CC, Hendriks GJ. [An exploratory study of 'blended' cognitive behavioural therapy (CBT) for patients with a panic disorder: results and patients' experiences]. Tijdschr Psychiatr 2016;58(5):361-370 [FREE Full text] [Medline: 27213635]
17. Massoudi B, Blanker MH, van Valen E, Wouters H, Bockting CLH, Burger H. Blended care vs. usual care in the treatment of depressive symptoms and disorders in general practice [BLENDING]: study protocol of a non-inferiority randomized trial. BMC Psychiatry 2017 Dec 13;17(1):218 [FREE Full text] [doi: 10.1186/s12888-017-1376-1] [Medline: 28610561]
18. Harrington KF, McDougal JA, Pisu M, Zhang B, Sadasivam RS, Houston TK, CHART Collaborative Group. Web-based smoking cessation intervention that transitions from inpatient to outpatient: study protocol for a randomized controlled trial. Trials 2012 Aug 01;13:123 [FREE Full text] [doi: 10.1186/1745-6215-13-123] [Medline: 22852802]
19. Månsson KNT, Skagius RE, Gervind E, Dahlin M, Andersson G. Development and initial evaluation of an Internet-based support system for face-to-face cognitive behavior therapy: a proof of concept study. J Med Internet Res 2013 Dec 10;15(12):e280 [FREE Full text] [doi: 10.2196/jmir.3031] [Medline: 24326065]

XSL•FO
RenderX

20. Kooistra LC, Ruwaard J, Wiersma JE, van Oppen P, van der Vaart R, van Gemert-Pijnen JE, et al. Development and initial evaluation of blended cognitive behavioural treatment for major depression in routine specialized mental health care. Internet Interventions 2016 May;4:61-71. [doi: 10.1016/j.invent.2016.01.003]
21. Kleiboer A, Smit J, Bosmans J, Ruwaard J, Andersson G, Topooco N, et al. European COMPARative Effectiveness research on blended Depression treatment versus treatment-as-usual (E-COMPARED): study protocol for a randomized controlled, non-inferiority trial in eight European countries. Trials 2016 Aug 03;17(1):387 [FREE Full text] [doi: 10.1186/s13063-016-1511-1] [Medline: 27488181]
22. Kemmeren LL, van Schaik DJ, Riper H, Kleiboer AM, Bosmans JE, Smit JH. Effectiveness of blended depression treatment for adults in specialised mental healthcare: study protocol for a randomised controlled trial. BMC Psychiatry 2016 Apr 21;16:113 [FREE Full text] [doi: 10.1186/s12888-016-0818-5] [Medline: 27102812]
23. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52. [doi: 10.2196/jmir.1772] [Medline: 21821503]
24. Sabate E, Staff W. Adherence to Long-Term Therapies : Evidence for Action. Geneva: World Health Organization; 2003.
25. Bosworth H, Hayden. Improving Patient Treatment Adherence: A Clinician's Guide. New York: Springer-Verlag; 2018.
26. Alterman AI, Gariti P, Cook TG, Cnaan A. Nicodermal patch adherence and its correlates. Drug Alcohol Depend 1999 Jan 07;53(2):159-165. [Medline: 10080041]
27. Westman EC, Behm FM, Simel DL, Rose JE. Smoking behavior on the first day of a quit attempt predicts long-term abstinence. Arch Intern Med 1997 Feb 10;157(3):335-340. [Medline: 9040302]
28. Fish LJ, Peterson BL, Namenek BRJ, Lyna P, Oncken CA, Swamy GK, et al. Adherence to nicotine replacement therapy among pregnant smokers. Nicotine Tob Res 2009 May;11(5):514-518 [FREE Full text] [doi: 10.1093/ntr/ntp032] [Medline: 19351783]
29. Kay-Lambkin F, Baker A, Lewin T, Carr V. Acceptability of a clinician-assisted computerized psychological intervention for comorbid mental health and substance use problems: treatment adherence data from a randomized controlled trial. J Med Internet Res 2011 Jan 27;13(1):e11 [FREE Full text] [doi: 10.2196/jmir.1522] [Medline: 21273184]
30. Wright JH, Wright AS, Albano AM, Basco MR, Goldsmith LJ, Raffield T, et al. Computer-assisted cognitive therapy for depression: maintaining efficacy while reducing therapist time. Am J Psychiatry 2005 Jun;162(6):1158-1164. [doi: 10.1176/appi.ajp.162.6.1158] [Medline: 15930065]
31. Hantsoo L, Epperson CN, Thase ME, Kim DR. Antepartum depression: treatment with computer-assisted cognitive-behavioral therapy. Am J Psychiatry 2013 Aug;170(8):929-930 [FREE Full text] [doi: 10.1176/appi.ajp.2013.13020254] [Medline: 23903340]
32. Sabate E, Staff W. Behavioural mechanisms explaining adherence: What every health professional should know. In: Adherence to Long-Term Therapies : Evidence for Action. Geneva: World Health Organization; 2003:135-149.
33. Hays JT, Leischow SJ, Lawrence D, Lee TC. Adherence to treatment for tobacco dependence: association with smoking abstinence and predictors of adherence. Nicotine Tob Res 2010 Jun;12(6):574-581. [doi: 10.1093/ntr/ntq047] [Medline: 20457644]
34. Taleb ZB, Ward KD, Asfar T, Bahelah R, Maziak W. Predictors of adherence to pharmacological and behavioral treatment in a cessation trial among smokers in Aleppo, Syria. Drug Alcohol Depend 2015 Aug 01;153:167-172 [FREE Full text] [doi: 10.1016/j.drugalcdep.2015.05.038] [Medline: 26077603]
35. Mathew M, Morrow JR, Frierson GM, Bain TM. Assessing digital literacy in web-based physical activity surveillance: the WIN study. Am J Health Promot 2011;26(2):90-95. [doi: 10.4278/ajhp.091001-QUAN-320] [Medline: 22040389]
36. Gerhards SAH, Abma TA, Arntz A, de Graaf LE, Evers SM, Huibers MJ, et al. Improving adherence and effectiveness of computerised cognitive behavioural therapy without support for depression: a qualitative study on patient experiences. J Affect Disord 2011 Mar;129(1-3):117-125. [doi: 10.1016/j.jad.2010.09.012] [Medline: 20889214]
37. Asfar T, Al Ali R, Rastam S, Maziak W, Ward KD. Behavioral cessation treatment of waterpipe smoking: The first pilot randomized controlled trial. Addict Behav 2014 Jun;39(6):1066-1074 [FREE Full text] [doi: 10.1016/j.addbeh.2014.02.012] [Medline: 24629480]
38. Heffner JL, Lewis DF, Winhusen TM. Preliminary evidence that adherence to counseling mediates the effects of pretreatment self-efficacy and motivation on outcome of a cessation attempt in smokers with ADHD. Nicotine Tob Res 2013 Feb;15(2):393-400 [FREE Full text] [doi: 10.1093/ntr/nts135] [Medline: 22949577]
39. Busnello RG, Melchior R, Faccin C, Vettori D, Petter J, Moreira LB, et al. Characteristics associated with the dropout of hypertensive patients followed up in an outpatient referral clinic. Arq Bras Cardiol 2001 May;76(5):349-354 [FREE Full text] [Medline: 11359183]
40. López-Torrecillas F, Rueda MM, López-Quirantes EM, Santiago JM, Tapioles RR. Adherence to treatment to help quit smoking: effects of task performance and coping with withdrawal symptoms. BMC Public Health 2014 Nov 25;14:1217 [FREE Full text] [doi: 10.1186/1471-2458-14-1217] [Medline: 25424314]
41. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor GS, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008 Nov 06;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]

XSL•FO
RenderX

42. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the Concept of Adherence to eHealth Technology: Systematic Review on When Usage Becomes Adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
43. Carolan S, Harris PR, Greenwood K, Cavanagh K. Increasing engagement with, and effectiveness of, an online CBT-based stress management intervention for employees through the use of an online facilitated bulletin board: study protocol for a pilot randomised controlled trial. Trials 2016 Dec 15;17(1):598 [FREE Full text] [doi: 10.1186/s13063-016-1733-2] [Medline: 27978858]
44. West R, Walia A, Hyder N, Shahab L, Michie S. Behavior change techniques used by the English Stop Smoking Services and their associations with short-term quit outcomes. Nicotine Tob Res 2010 Jul;12(7):742-747. [doi: 10.1093/ntr/ntq074] [Medline: 20478957]
45. CBO. Richtlijn. Utrecht: Trimbos-instituut - Netherlands Institute of Mental Health and Addiction; 2016. Behandeling van tabaksverslaving en stoppen met roken ondersteuning, Herziening 2016 URL: https://assets.trimbos.nl/docs/7435d8f9-c8f4-4302-aec6-f4fddade81f2.pdf [accessed 2018-07-18] [WebCite Cache ID 710HyJuwE]
46. Partnership SMR. Zorgmodule. Den Haag: Partnership Stop met Roken; 2009. Stoppen met Roken URL: https://www.zorginzicht.nl/bibliotheek/Tabaksverslaving/RegisterKwaliteitsstandaardenDocumenten/ZorgmoduleStoppenmetRoken.pdf [accessed 2018-07-18] [WebCite Cache ID 710I5ORG9]
47. Kelders SM, Kok RN, Ossebaard HC, van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
48. Lehto T, Oinas-Kukkonen H. Persuasive features in web-based alcohol and smoking interventions: a systematic review of the literature. J Med Internet Res 2011 Jul 22;13(3):e46 [FREE Full text] [doi: 10.2196/jmir.1559] [Medline: 21795238]
49. DiMatteo MR. Social support and patient adherence to medical treatment: a meta-analysis. Health Psychol 2004 Mar;23(2):207-218. [doi: 10.1037/0278-6133.23.2.207] [Medline: 15008666]
50. Shiffman S. Relapse following smoking cessation: a situational analysis. J Consult Clin Psychol 1982 Feb;50(1):71-86. [Medline: 7056922]
51. Marlatt GA, Gordon JR. Determinants of relapse : implications for the maintenance of behavior change. In: Technical report (University of Washington. Alcoholism & Drug Abuse Institute).: Alcoholism & Drug Abuse Institute; 1978 Presented at: Tenth International Conference on Behavior Modification; March, 1978; Banff, Alberta, Canada.
52. Cummings KM, Jaén CR, Giovino G. Circumstances surrounding relapse in a group of recent exsmokers. Prev Med 1985 Mar;14(2):195-202. [Medline: 4048082]
53. Moncher F, Prinz R. Treatment fidelity in outcome studies. Clinical Psychology Review 1991 Jan;11(3):247-266 [FREE Full text] [doi: 10.1016/0272-7358(91)90103-2]
54. Waller G. Evidence-based treatment and therapist drift. Behav Res Ther 2009 Feb;47(2):119-127. [doi: 10.1016/j.brat.2008.10.018] [Medline: 19036354]

Abbreviations

BSCT: blended smoking cessation treatment
CO: carbon monoxide
f2f: face-to-face
IQR: interquartile range
MREC: Medical Research Ethics Committee
OR: odds ratio
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 29.01.18; peer-reviewed by B Bonevski, S Barbieri; comments to author 18.04.18; accepted 08.05.18; published 01.08.18.

Please cite as:
Siemer L, Brusse-Keizer MGJ, Postel MG, Ben Allouch S, Patrinopoulos Bougioukas A, Sanderman R, Pieterse ME
Blended Smoking Cessation Treatment: Exploring Measurement, Levels, and Predictors of Adherence
J Med Internet Res 2018;20(8):e246
URL: http://www.jmir.org/2018/8/e246/
doi:10.2196/jmir.9969
PMID:30068503

XSL•FO
RenderX

©Lutz Siemer, Marjolein GJ Brusse-Keizer, Marloes G Postel, Somaya Ben Allouch, Angelos Patrinopoulos Bougioukas, Robbert Sanderman, Marcel E Pieterse. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 01.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Differences in the Effect of Internet-Based Cognitive Behavioral Therapy for Improving Nonclinical Depressive Symptoms Among Workers by Time Preference: Randomized Controlled Trial

Kotaro Imamura[1], PhD, MPH; Toshi A Furukawa[2], MD, PhD; Yutaka Matsuyama[3], PhD; Akihito Shimazu[4], PhD; Kazuto Kuribayashi[5], MHSc; Kiyoto Kasai[6], MD, PhD; Norito Kawakami[1], MD, DMSc

[1]Department of Mental Health, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[2]Departments of Health Promotion and Human Behavior and of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto University, Kyoto, Japan

[3]Department of Biostatistics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[4]Center for Human and Social Sciences, Kitasato University College of Liberal Arts and Sciences, Sagamihara, Japan

[5]Department of Psychiatric Nursing, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[6]Department of Neuropsychiatry, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

Corresponding Author:
Norito Kawakami, MD, DMSc
Department of Mental Health
Graduate School of Medicine
The University of Tokyo
7-3-1, Hongo
Bunkyo-ku
Tokyo, 113-0033
Japan
Phone: 81 3 5841 3522
Fax: 81 3 5841 3392
Email: nkawakami@m.u-tokyo.ac.jp

Abstract

Background: Previous randomized controlled trials (RCTs) have shown a significant intervention effect of internet-based computerized cognitive behavioral therapy (iCBT) on improving nonclinical depressive symptoms among healthy workers and community residents in a primary prevention setting. Time preference is one's relative valuation for having a reward (eg, money) at present than at a later date. Time preference may affect the effectiveness of cognitive behavioral therapy.

Objective: This RCT aimed to test the difference of intervention effect of an iCBT program on improving nonclinical depressive symptoms between two subgroups classified post-hoc on the basis of time preference among workers in Japan.

Methods: All workers in one corporate group (approximate n=20,000) were recruited. Participants who fulfilled the inclusion criteria were randomly allocated to either intervention or control groups. Participants in the intervention group completed 6 weekly lessons and homework assignments within the iCBT program. The Beck Depression Inventory-II (BDI-II) and Kessler's Psychological Distress Scale (K6) measures were obtained at baseline and 3-, 6-, and 12-month follow-ups. Two subgroups were defined by the median of time preference score at baseline.

Results: Only few (835/20,000, 4.2%) workers completed the baseline survey. Of the 835 participants, 706 who fulfilled the inclusion criteria were randomly allocated to the intervention or control group. Participants who selected irrational time preference options were excluded (21 and 18 participants in the intervention and control groups, respectively). A three-way interaction (group [intervention/control] × time [baseline/follow-up] × time preference [higher/lower]) effect of iCBT was significant for BDI-II ($t_{1147.42}$=2.33, P=.02) and K6 ($t_{1254.04}$=2.51, P=.01) at the 3-month follow-up, with a greater effect of the iCBT in the group with higher time preference. No significant three-way interaction was found at the 6- and 12-month follow-ups.

Conclusions: The effects of the iCBT were greater for the group with higher time preference at the shorter follow-up, but it was leveled off later. Workers with higher time preference may change their cognition or behavior more quickly, but these changes may not persist.

XSL•FO
RenderX

Trial Registration: UMIN Clinical Trials Registry UMIN000014146; https://upload.umin.ac.jp/cgi-open-bin/ctr_e/ctr_view.cgi?recptno=R000016466 (Archived by WebCite at http://www.webcitation.org/70o2rNk2V)

(J Med Internet Res 2018;20(8):e10231) doi:10.2196/10231

KEYWORDS

internet-based computerized cognitive behavioral therapy; time preference; nonclinical depressive symptoms; workers

Introduction

Depressive disorder is one of the most prevalent psychiatric disorders, affecting around 340 million people worldwide [1] and is associated with a substantial deterioration in quality of life and economic loss in the community and the workplace [2,3]. The primary prevention of depressive disorder is an important strategy for global mental health. The presence of nonclinical depressive symptoms (ie, subthreshold depressive symptoms) is associated with high prospective risk of developing major depressive disorder (MDD) [4,5], and a previous meta-analysis reported that it was possible to prevent the onset of MDD using psychological interventions by targeting individuals with no diagnosed depression at baseline survey [6].

One of the most effective psychological interventions for depression is cognitive behavioral therapy (CBT) [7], and internet-based computerized CBT (iCBT) has received attention in recent years because it is less expensive, more easily administered, and potentially more accessible than conventional CBT. Previous randomized controlled trials (RCTs) have shown a significant intervention effect of iCBT for improving nonclinical depressive symptoms [8,9] and preventing the onset of new major depressive episodes (MDEs) [10] among healthy workers and community residents.

Recently, variables that might predict treatment response to CBT for depression have been investigated. Previous studies have reported that the severity of depressive symptoms at baseline and the rate of change in depressive symptom severity within 5 treatment sessions significantly predicted treatment response to CBT [11,12]. As a predictor of treatment response to iCBT, there was a significant association with pretreatment severity of depression, gender, marital status, and education [13-15]. In addition, a recent study reported that individual differences in reward processing, measured by reward positivity, contribute to the effectiveness of CBT for depression [16]. This result implies that CBT may decrease depressive symptoms by enhancing the brain's reward function. Sensitivity for reward may be an important predictor of the effectiveness of CBT.

Time preference (or time discounting) has attracted interest in the field of behavioral economics and behavioral medicine as a potentially common factor of multiple behaviors that pose risks for health [17-19]. Time preference is one's relative reward valuation (eg, money) at present than at a later date [20]. Frederick et al (2002) stated that time discounting means caring less about a future consequence, including factors that diminish the expected utility generated by a future consequence, and time preference refers to the preference for immediate over delayed utility [20]. Time preference can be shortly defined as the degree to which people prefer present to future satisfaction [21]. Individuals that have a high rate of time preference or tend to prefer utility in the present are often designated as present-oriented and labeled as impatient. On the other hand, individuals with a low rate of time preference or those who tend to prefer future utilities are often designated as future-oriented and are said to be patient [22]. Time preference may affect the effectiveness of CBT for various reasons. First, people with higher time preference may be less eager to participate in a health education program for preventing future psychological distress or mental disorders, such as CBT. Second, people with higher time preference may be less willing to change their behaviors in order to improve their future health [17-19]. Third, for the same reason, the effect of CBT may not be persistent among people with higher time preference. However, no previous study has investigated the effect of time preference on the intervention effect of CBT for improving nonclinical depression. Investigating the impact of time preference on the effectiveness of CBT would contribute to the development of a theory for behavioral determinants of the effectiveness of CBT. In practice, it would also lead to identifying a subgroup for which CBT is less effective and allow us to improve interventions and treatment effects for this population.

This RCT aimed to examine whether an iCBT program was effective in improving nonclinical depressive symptoms among healthy workers in Japan, at 3-, 6-, and 12-month follow-ups, particularly to test the difference of the intervention effect between two subgroups classified post-hoc on the basis of time preference: a lower time preference subgroup and a higher time preference subgroup.

Methods

Trial Design

This study was a randomized controlled trial. The allocation ratio of the intervention group to the control group was 1:1. The Research Ethics Review Board of the Graduate School of Medicine and Faculty of Medicine, the University of Tokyo approved the study procedures (no. 3083-2). The study protocol was registered at the University Hospital Medical Information Network Clinical Trials Registry (UMIN000014146). The protocol article for this trial is available [23]. This study focused on the first-year recruitment for the planned larger study. The original protocol of this RCT aims to investigate whether an iCBT program could prevent the onset of MDE as a primary outcome. Outcomes in this study (ie, depressive symptoms and psychological distress) were collected as secondary outcomes. This manuscript was reported according to the Consolidated Standards of Reporting Trials guidelines.

XSL•FO
RenderX

Participants

All workers in one corporate group (the total employee population, approximately 20,000) were recruited from one of the major telecom carrier companies in Japan by an invitation email from their internal employee assistance program staff in March 2015. Those who were interested in participating in the study were asked to go to a research website to obtain a full explanation of the study's aim. Consent from a respondent was obtained when he or she completed a baseline questionnaire. Before the Web-based baseline survey, participants were invited to read the explanation on the research website and asked to click on an "agree" button to show their consent to participate in the study; then they proceeded to the baseline questionnaire page. Written consent was not required by the National Ethical Guidelines for Epidemiologic Research, Japan; the Research Ethics Review Board of Graduate School of Medicine and Faculty of Medicine, the University of Tokyo, approved this procedure for obtaining participants' consents.

The inclusion criteria at the baseline survey were as follows: (1) age 20-60 years at the study entry, (2) currently employed full-time by the company, and (3) being able to access the internet via a PC at home or at their workplace. The exclusion criteria were as follows: (1) nonregular or part-time employees, (2) having an MDE in the past month, based on the diagnostic criteria on the web version of World Health Organization Composite International Diagnostic Interview 3.0 [24], (3) having lifetime history of bipolar disorder (World Health Organization Composite International Diagnostic Interview 3.0), (4) on sick leave for 15 or more days for a physical or mental condition in the past 3 months, and (5) undergoing current treatment for a mental health problem.

Intervention

Participants assigned to an intervention group participated in the iCBT program called *the Internet CBT program; useful mental health solutions series for business*. Please refer for the details of this program elsewhere [23]. Briefly, the program was a 6-week, 6-lesson, Web-based training course to provide CBT-based stress management skills via one 30-minute lesson per week. The CBT components of the program included self-case formulation, cognitive restructuring, assertiveness, problem-solving, and relaxation. At the end of each lesson, the participants were asked to submit homework to facilitate their understanding, but on voluntary basis. Participants who submitted their homework received feedback from trained clinical psychologists.

Intervention Group

Participants in the intervention group completed 6 weekly lessons and homework within the iCBT program. They were allowed to complete the 6 lessons and submit their homework within 10 weeks after the baseline survey. The participants were reminded by email to complete each lesson and to submit their homework if they had not already done so. Reminders were sent from the research office to the participants every Monday.

Control Group

Participants in the control group were able to use an internal employee assistance program service, such as consulting with a physician or a psychologist, and group or Web-based education/training programs for promoting mental health as a treatment as usual. These programs contained few descriptions of CBT knowledge and skills.

Outcome

All outcomes were measured using a Web-based self-report questionnaire at baseline and 3-, 6-, and 12-month follow-ups.

Depressive Symptoms

The Beck Depression Inventory-II (BDI-II) is a 21-item self-report inventory that measures depressive symptoms such as sadness, pessimism, suicidal thoughts or wishes, tiredness or fatigue, loss of energy, and loss of pleasure, among others [25,26]. Each item was scored on a scale ranging from 0 to 3, with a higher score indicating more serious depressive symptoms.

Psychological Distress

Kessler's Psychological Distress Scale (K6) consists of 6 items assessing the frequency with which respondents experienced symptoms of psychological distress during the past 30 days [27,28]. The response options range from 0 (*none of the time*) to 4 (*all of the time*). The internal reliability and validity found in previous studies were acceptable [27].

Time Preference

In this study, time preference was assessed by the following procedure [29,30]. The respondents were asked to choose between two options, A or B. The respondent would receive 1 million yen (approximately US $12,000) in 1 month upon choosing option A, or a different amount to be received in 13 months upon choosing option B. This question comprised 9 choices with each annual interest rate ranging from −5% to ≥10% (Multimedia Appendix 1). For instance, individuals who tended to choose option B, despite lower annual interest in 13 months, were considered more future-oriented (ie, lower time preference). On the other hand, individuals who tended to choose option A, despite higher annual interest in 13 months, were considered more present-oriented (ie, higher time preference).

In this study, we defined two subgroups according to the median time preference score at baseline because the concept of time preference has no clear cutoff point. One was the *lower time preference* subgroup (ie, the participants who had low levels of time preference and selected the 0.1%-6% annual interest rate), and the other was the *higher time preference* subgroup (ie, the participants who had high levels of time preference and selected the 10% annual interest rate or more). Participants who selected irrational options (interest rate, −5% or 0%) were excluded.

Demographic Characteristics

Demographic data such as age, gender, marital status, occupation, education, and chronic disease were also collected.

Sample Size

We determined that to detect an effect size, a minimum sample size of 4136 in each group was necessary. This calculation considered an incidence ratio of 0.62 or greater for the onset of

XSL•FO
RenderX

an MDE, at an alpha error rate of 0.05 (two-tailed) and a beta error rate of 0.20, with an expected dropout rate of 25%.

No previous study reported an effect size for a difference of intervention effect between lower and higher time preference groups. The estimated post-hoc power (1-beta) was 0.54 if the effect size was 0.2, assuming that the alpha was less than 0.05 (two-tailed), and 70% (314/448) of the initial 448 respondents in the lower time preference subgroup and 219 respondents in the higher time preference subgroup respondents completed the follow-up using the G*Power 3 program [31,32].

Randomization

Participants who fulfilled the inclusion criteria were randomly assigned to an intervention or control group. Stratified permuted-block randomization was conducted as well. Participants were stratified into two strata according to K6 score (5 or greater or less than 5) on the baseline survey. A stratified permuted-block random table was generated by an independent biostatistician. Enrollment was conducted by a clinical research coordinator, and assignment was conducted by an independent research assistant. The stratified permuted-block random table was password protected and kept blind to the researcher. Only the research assistant was able to access it for random allocation. A prestratification for randomization by time preference was not conducted.

Statistical Methods

Primary analyses were conducted for the whole sample. For main analysis, a mixed model for repeated measures conditional growth model analysis was conducted to estimate the fixed effect of a three-way interaction as an indicator of intervention effect: group (intervention and control) × time (baseline and 1-, 6-, and 12-month follow-ups) × subgroup (lower time preference and higher time preference). For sensitivity analysis, a mixed model for repeated measures analysis of variance was conducted to estimate the fixed effect of three-way interaction as an indicator of intervention effect at each follow-up: group (intervention and control) × time (baseline and 1-, 6-, or 12-month follow-up) × subgroup (lower time preference and higher time preference). In these analyses, two models were applied. Model 1 was crude (not adjusted). Model 2 was adjusted by the potential confounders: gender, education, and occupation. All analyses were conducted according to the intention-to-treat principles. The MIXED procedure in SPSS Statistics 21.0 (IBM Corp, Armonk, NY, USA) was used.

Secondary analyses were conducted for all respondents as well as separately for each subgroup. A mixed model for repeated measures conditional growth model analysis was conducted to estimate the fixed effect of a group (intervention and control) × time (baseline, 1-, 6-, and 12-month follow-ups) interaction as an indicator of intervention effect. As a sensitivity analysis, a mixed model for repeated measures analysis of variance was conducted to estimate the fixed effect of a group (intervention and control) × time (baseline and 1-, 6-, or 12-month follow-up) interaction as an indicator of the intervention effect at each follow-up.

In addition, the effect sizes were calculated using estimated means based on the MIXED procedure among all respondents in each subgroup. First, estimated mean differences between baseline and follow-ups of each intervention and control group were calculated. Next, the effect sizes (ESs) were calculated by dividing between differences of the intervention and control groups by pooled SDs, which were calculated using respondents who completed the questionnaire at baseline and at follow-ups. The values of 0.2, 0.5, and 0.8 were interpreted as small, medium, and large ESs, respectively [33].

As a process evaluation, the rate (percentage) of completers of lessons and submitters of iCBT program homework were calculated among participants in the intervention group, for each lower and higher time preference subgroup.

Results

Recruitment

Recruitment and the baseline survey were conducted in March 2015. The intervention and control groups were assessed at approximately 3 months (June 2015), 6 months (September 2015), and 12 months (March 2016) after the baseline survey.

The participant flowchart is shown in Figure 1. In total, 4.2% of workers (835/20,000) participated in a baseline survey. Out of those workers, 706 met the eligibility criteria of this study and 129 were excluded (39 cases fulfilled exclusion criteria 1; 9 cases fulfilled exclusion criteria 2; 16 cases fulfilled exclusion criteria 3; 87 cases fulfilled exclusion criteria 4). Out of those excluded workers, a total of 13 cases fulfilled exclusion criteria 1 and 4; 2 cases fulfilled exclusion criteria 2 and 4; 5 cases fulfilled exclusion criteria 3 and 4; and 1 case fulfilled exclusion criteria 2, 3, and 4. Seven hundred and six participants were randomly allocated to the intervention or control group (n=353 for each). Figure 1 also shows excluded participants (21 in the intervention group and 18 in the control group) after randomization because they selected irrational time preference options (interest rate, −5% or 0%). At each follow-up, the response rate of the control group was higher than that of the intervention group. The reasons for dropping out were not assessed in this study.

Baseline Characteristics

Demographic characteristics are presented in Table 1. Compared with the lower time preference subgroup, there were more males, managers, participants with a graduate school education, and those with chronic diseases in the higher time preference subgroup. In the whole sample, most participants were married, held clerical positions, received a university or higher education, and did not report having chronic diseases.

After excluding participants who selected irrational options (interest rate, −5% or 0%), we divided the total sample into two groups of participants: one with low levels of time preference (6% or lower annual interest rate) and one with high levels of time preference (10% or higher annual interest rate). The details of the number of respondents in each group are shown in Multimedia Appendix 1.

Changing Outcomes by Groups During the Follow-Up

Tables 2 and 3 show the means and SDs of the outcome variables at baseline and 3-, 6-, and 12-month follow-ups in the

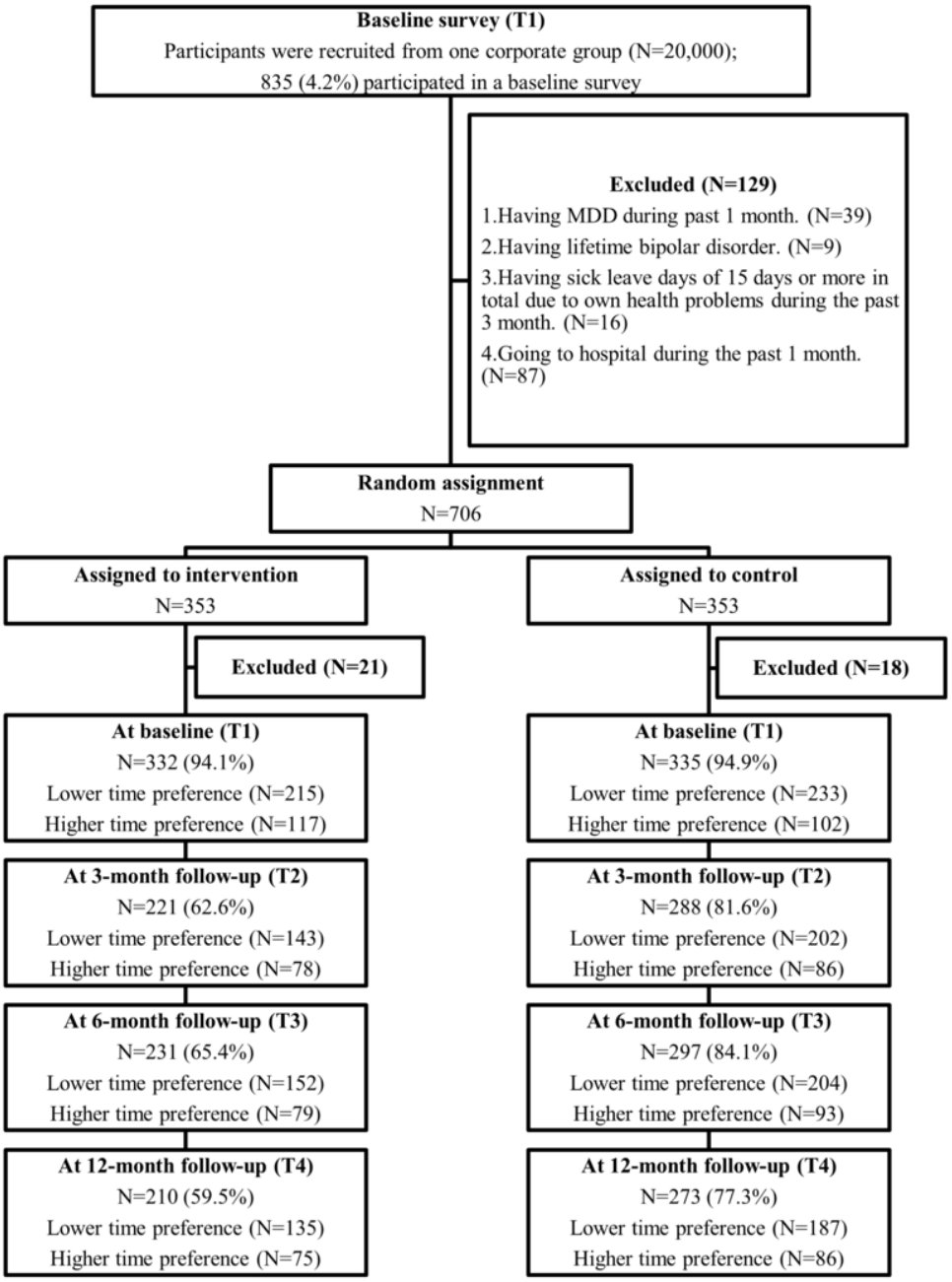
XSL•FO
RenderX

intervention and control groups. In addition, we report estimated mean differences between the intervention and control groups, pooled SDs, and ESs in each lower and higher time preference subgroup. In the lower time preference subgroup, the estimated mean differences were −1.21 (ES=−0.30) on K6 at 12-month and −1.68 (ES=−0.23) and −2.63 (ES=−0.33) on BDI-II at 6- and 12-month follow-ups, respectively. In the higher time preference subgroup, the estimated mean differences were −1.13 (ES=−0.30) on K6 at 3-month follow-up and −2.55 (ES=−0.37) and −2.59 (ES=−0.41) on BDI-II at 3- and 6-month follow-ups, respectively.

Interaction Effects of Internet-Based Computerized Cognitive Behavioral Therapy and Time Preference

Table 4 shows the estimated three-way interaction effects of iCBT on the outcome variables on the basis of the mixed model analyses. iCBT showed a significant effect on BDI-II and K6 at 3-month follow-up, and only a marginally significant effect on BDI-II at 6-month follow-up. These results were consistent with the results after adjusting for gender, occupational status, and education.

Figure 1. Participant flowchart. MDD: major depressive disorder.



XSL•FO
RenderX

Table 1. Baseline characteristics of participants in the intervention and control groups, in each of the two subgroups.

Characteristic	Lower time preference[a], mean (SD)		Higher time preference[b], mean (SD)	
	Intervention (n=215)	Control (n=233)	Intervention (n=117)	Control (n=102)
Age (years)	38.7 (8.1)	39.0 (7.7)	39.6 (9.1)	40.3 (9.3)
Gender				
Male	106 (49.3)	126 (54.1)	75 (64.1)	69 (67.6)
Female	109 (50.7)	107 (45.9)	42 (35.9)	33 (32.4)
Marital status				
Never married	91 (42.3)	80 (34.3)	46 (39.3)	30 (29.4)
Married	115 (53.5)	146 (62.7)	67 (57.3)	67 (65.7)
Divorced or bereaved	9 (4.2)	7 (3.0)	4 (3.4)	5 (4.9)
Occupation				
Manager	42 (19.5)	45 (19.3)	35 (29.9)	39 (38.2)
Professional	49 (22.8)	65 (27.9)	25 (21.4)	20 (19.6)
Clerical	101 (47.0)	96 (41.2)	41 (35.0)	37 (36.3)
Production	1 (0.5)	1 (0.4)	0 (0.0)	0 (0.0)
Sales	18 (8.4)	20 (8.6)	12 (10.3)	5 (4.9)
Others	4 (1.9)	6 (2.6)	4 (3.4)	1 (1.0)
Education				
High school	9 (4.2)	13 (5.6)	5 (4.3)	5 (4.9)
Some college	38 (17.7)	38 (16.3)	23 (19.7)	14 (13.7)
University	151 (70.2)	165 (70.8)	72 (61.5)	65 (63.7)
Graduate school	17 (7.9)	17 (7.3)	17 (14.5)	18 (17.6)
Chronic disease				
Yes	19 (8.8)	25 (10.7)	19 (16.2)	16 (15.7)
No	196 (91.2)	208 (89.3)	98 (83.8)	86 (84.3)

[a]0.1%-6% annual percentage yield.

[b]≥10% annual percentage yield.

Effects of Internet-Based Computerized Cognitive Behavioral Therapy by Time Preference Subgroups

In the whole sample (n=667), the iCBT program showed a significant pooled intervention effect on BDI-II ($t_{548.79}$=−3.36, P<.01) and K6 ($t_{551.08}$=−2.70, P=.01) at 12-month follow-up. For each follow-up, iCBT showed a significant effect on BDI-II ($t_{1238.9}$=−2.61, P=.01) at 3-months, on BDI-II ($t_{1232.1}$=−3.18, P<.01) and K6 ($t_{1253.8}$=−2.36, P=.02) at 6-months, and on BDI-II ($t_{862.64}$=−3.19, P<.01) and K6 ($t_{875.92}$=−2.37, P=.02) at 12-month follow-up.

In the lower time preference subgroup (n=448), iCBT program showed a significant pooled effect on BDI-II ($t_{374.06}$=−3.31, P<.01) and K6 ($t_{368.30}$=−3.09, P<.01) at 12-month follow-up. For each follow-up, iCBT showed a significant effect on BDI-II ($t_{842.44}$=−2.25, P=.02) at 6-month and on BDI-II ($t_{583.45}$=−3.32, P<.01) and K6 ($t_{652.08}$=−2.81, P=.01) at 12-month follow-up. The other combinations were not statistically significant (data available upon request).

In the higher time preference subgroup (n=219), the pooled effects were not significant for both BDI-II ($t_{172.52}$=−1.10, P=.27) and K6 ($t_{180.18}$=−0.39, P=.70). For each follow-up, iCBT showed a significant effect on BDI-II ($t_{384.05}$=−2.54, P=.01) and K6 ($t_{426.28}$=−2.04, P=.04) only at 3-months and on BDI-II ($t_{385.61}$=−2.44, P=.02) at 6-month follow-up. The other combinations were not statistically significant (data available upon request).

XSL•FO
RenderX

Table 2. Average scores of depressive symptoms (Beck Depression Inventory-II [BDI-II] and Kessler's Psychological Distress Scale 6 [K6]) at baseline and 1-, 6-, and 12-month follow-up.

Subgroup and follow-up	Intervention			Control		
	n	K6, mean (SD)	BDI-II, mean (SD)	n	K6, mean (SD)	BDI-II, mean (SD)
Lower time preference group[a]						
T1[b]	215	6.3 (4.5)	12.7 (8.6)	233	6.0 (4.3)	12.2 (8.7)
T2[c]	143	5.9 (4.8)	10.5 (8.6)	202	6.0 (4.3)	11.7 (8.2)
T3[d]	152	5.7 (4.6)	10.1 (8.2)	204	6.3 (5.0)	11.6 (9.0)
T4[e]	135	5.6 (5.0)	10.6 (9.5)	187	6.7 (4.8)	12.7 (9.9)
Higher time preference group[f]						
T1	117	5.9 (5.2)	12.2 (8.4)	102	6.5 (4.9)	13.7 (10.2)
T2	78	4.8 (4.5)	9.5 (7.9)	86	6.7 (5.1)	13.9 (10.9)
T3	79	4.9 (4.5)	9.4 (7.6)	93	6.9 (5.5)	13.9 (11.1)
T4	75	5.9 (5.0)	10.8 (9.9)	86	6.3 (4.6)	12.5 (10.5)

[a]0.1%-6% annual percentage yield.

[b]T1: baseline.

[c]T2: 3-month follow-up.

[d]T3: 6-month follow-up.

[e]T4: 12-month follow-up.

[f]≥10% annual percentage yield.

Table 3. Estimated mean difference[a], pooled SD[b], and effect size[c] between groups.

	K6[d]			BDI-II[e]		
Subgroup and follow-up	Estimated Δ[a]	Pooled SD	Effect size	Estimated Δ[a]	Pooled SD	Effect size
Lower time preference group[f]						
T2[g]-T1[h]	−0.02	3.69	−0.01	−0.98	6.69	−0.15
T3[i]-T1	−0.78	4.16	−0.19	−1.68	7.21	−0.23
T4[j]-T1	−1.21	4.01	−0.30	−2.63	8.02	−0.33
Higher time preference group[k]						
T2-T1	−1.13	3.77	−0.30	−2.55	6.85	−0.37
T3-T1	−0.88	4.38	−0.20	−2.59	6.36	−0.41
T4-T1	−0.17	4.40	−0.04	−0.91	7.40	−0.12

[a]Estimated means were calculated using a MIXED procedure.

[b]Pooled SDs were calculated using respondents those who completed the questionnaire at baseline and at follow-ups.

[c]Effect sizes were calculated by dividing estimated mean difference by pooled SD.

[d]K6: Kessler's Psychological Distress Scale.

[e]BDI-II: Beck Depression Inventory-II.

[f]0.1%-6% annual percentage yield.

[g]T2, 3-month follow-up.

[h]T1, baseline.

[i]T3, 6-month follow-up.

[j]T4, 12-month follow-up.

[k]≥10% annual percentage yield.

XSL•FO
RenderX

Table 4. Three-way interaction effects of the internet-based cognitive behavioral therapy, time, and time preference on Beck Depression Inventory-II (BDI-II) and Kessler's Psychological Distress Scale (K6).

Scale and follow-up	Crude					Gender, occupational status, and education adjusted				
	Effect	95% CI	SE	*t*	*P*	Effect	95% CI	SE	*t*	*P*
K6										
3 months[a]	1.96	0.31 to 3.61	0.84	2.33	.02	2.12	0.46 to 3.77	0.84	2.51	.01
6 months[a]	0.98	−0.67 to 2.63	0.84	1.16	.25	1.14	−0.51 to 2.79	0.84	1.35	.18
12 months[a]	−0.19	−1.88 to 1.50	0.86	−0.22	.83	−0.01	−1.70 to 1.68	0.86	−0.01	.99
Pooled[b]	−0.06	−0.45 to 0.34	0.20	−0.27	.79	−0.02	−0.42 to 0.38	0.20	−0.11	.91
BDI-II										
3 months[a]	3.57	0.43 to 6.72	1.60	2.23	.03	3.75	0.60 to 6.91	1.61	2.33	.02
6 months[a]	2.87	−0.28 to 6.01	1.60	1.79	.07	3.07	−0.08 to 6.22	1.61	1.91	.06
12 months[a]	0.23	−2.97 to 3.44	1.64	0.14	.89	0.44	−2.78 to 3.65	1.64	0.27	.79
Pooled[b]	0.01	−0.75 to 0.78	0.39	0.04	.97	0.05	−0.72 to 0.82	0.39	0.13	.90

[a]A mixed model for repeated measures analysis of variance model analyses was conducted to estimate a three-way interaction effect among intervention, time, and time preference.

[b]A mixed model for repeated measures conditional growth model analyses was conducted to estimate a three-way interaction effect.

Table 5. Progress of learning in the internet-based cognitive behavioral therapy program in the two subgroups.

Contents	Lower time preference (n=215), n (%)		Higher time preference (n=117), n (%)	
	Completers of lessons	Submitters of homework	Completers of lessons	Submitters of homework
Lesson (L)1: *Learning about stress*	186 (86.5)	124 (57.7)	103 (88.0)	70 (59.8)
L2: *Knack for self-case formulation based oncognitive behavioral model*	181 (84.2)	85 (39.5)	97 (82.9)	51 (43.6)
L3: *Try cognitive restructuring part 1*	167 (77.7)	84 (39.1)	87 (74.4)	48 (41.0)
L4: *Try cognitive restructuring part 2*	153 (71.2)	68 (31.6)	80 (68.4)	37 (31.6)
L5: *Knack for communication*	142 (66.0)	54 (25.1)	76 (65.0)	34 (29.1)
L6: *How to solve your problem effectively*	138 (64.2)	55 (25.6)	69 (59.0)	28 (23.9)
All 6 lessons	136 (63.3)	37 (17.2)	68 (58.1)	20 (17.1)

Process Evaluation

Table 5 shows the process evaluation indicators of iCBT programs for the lower and higher time preference subgroups. Most participants in the intervention group completed Lesson 1 (186/215, 86.5% in the lower time preference group and 103/117, 88.0% in the higher time preference group), and about 60% in both subgroups (124/215 in the lower time preference group and 70/117 in the higher time preference group) submitted their homework after completing this lesson. The proportion of those who completed lessons and submitted homework gradually decreased during the later lessons. About 60% in both subgroups (136/215 in the lower time preference group and 68/117 in the higher time preference group) completed all 6 lessons, while only about 17% of them (37/215 in the lower time preference group and 20/117 in the higher time preference group) submitted all 6 homework assignments. In the lower time preference group, the average number of lessons that the respondents received was 4.5 and the average number of homework assignments submitted was 2.2. Of all participants, 77.7% (177/215) completed at least 3 lessons, and 38.1% (82/215) submitted at least 3 homework assignments. In the higher time preference group, the average number of lessons that the respondents received was 4.4 and the average number of homework assignments submitted was 2.3. In total, 75.2% (88/117) participants completed at least 3 lessons, and 40.2% (47/117) participants submitted at least 3 homework assignments. There were no differences of completers of lessons or submitters of homework of the iCBT program in both subgroups.

Discussion

Principal Findings

This RCT examined the effects of iCBT on improving nonclinical depressive symptoms at 3-, 6-, and 12-month follow-ups among healthy workers by lower and higher time preference subgroups in Japan. As a result, the three-way interaction effect of iCBT was significant for nonclinical depressive symptoms at 3-month follow-up, after adjusting for

XSL•FO
RenderX

gender, occupational status, and education. In the higher time preference subgroup, iCBT showed a significant intervention effect on nonclinical depressive symptoms at 3- and 6-month follow-ups, while the pooled effect was not significant. On the other hand, in the lower time preference subgroup, iCBT showed significant ESs on nonclinical depressive symptoms at 6- and 12-month follow-ups. The iCBT program showed a significant pooled effect on nonclinical depressive symptoms at 12-month follow-up.

To our knowledge, this is the first RCT that has demonstrated the effect of an iCBT on improving nonclinical depressive symptoms, specifically targeting workers with lower or higher time preference. iCBT showed a significantly higher effect for improving nonclinical depressive symptoms in the higher time preference subgroup than the lower time preference subgroup at 3-month follow-up. Workers with higher time preference may more easily change their cognition or behavior, but these changes persisted for only a short period. The pooled effect of iCBT was significant only in the lower time preference subgroup. Workers with lower time preference may be more likely to keep their cognitive or behavioral changes for a longer period.

Comparison with Prior Work

This study showed a difference in the intervention effect of iCBT between the higher time preference subgroup and the lower time preference subgroup. However, in the process evaluation, there were no differences between completers of lessons and submitters of homework of the iCBT program in both subgroups. Our findings caused us to reject the hypothesis that participants with higher time preference were less likely to follow the program.

Previous systematic reviews suggested that higher time preference was associated with poor responses to health promotion interventions such as dietary and weight loss programs [17,18]. However, this study showed that the higher time preference subgroup experienced a faster improvement in depressive symptoms than the lower time preference subgroup. Our study did not support the hypothesis that participants with higher time preferences were less likely to react to the program. Rather these participants exhibited significant mood improvements within a short (3-month) period. Workers with higher time preference may be more likely to change their behavior following engagement with an intervention that is immediately useful for treating their problems such as CBT, rather than an intervention that leads to long-term benefits such as healthy behaviors for preventing lifestyle-related diseases. Learning during the early period enhanced the intervention effects for the lower time preference subgroups.

The intervention effects of iCBT were less persistent among workers with higher long-term time preferences (eg, over 6 months). These findings support the hypothesis that the effect of CBT is not persistent among people with higher time preferences. Workers with higher time preferences may experience difficulty in maintaining their cognitive and behavioral changes. Workers with higher time preferences may stop using their new CBT-related perspectives or behaviors when their problems are solved (ie, improvement of nonclinical depressive symptoms). They may underestimate the future risk for a recurrence of the problems and not keep practicing a preventive effort. A follow-up program providing incentives (eg, allocating points or giving a prize as a reward) may reinforce continuing activities, making the iCBT program more effective even after 6 months for workers with higher time preferences.

These findings may contribute to further understanding of behavioral characteristics of people based on their (higher or lower) time preference. In this study, workers with higher time preferences were less likely to maintain the effects of a CBT-based program over the long-term, compared with those with lower time preference, while both groups engaged in learning to a similar extent. This pattern was consistent with previous reports on the impact of time preferences on health-related behaviors such as obesity and smoking [17-19]. This study expanded on already observed behavioral characteristics of individuals with higher time preferences, indicating that behavior patterns associated with higher time preference can be applied to the CBT-based programs. In addition, this study observed a very interesting pattern associated with higher time preference: the intervention effect was temporarily boosted among workers with higher time preferences, which was not seen for those with lower time preferences. This behavior pattern may be observed for other health-related behaviors such as diet, weight loss, and smoking cessation. Further research is warranted to examine the generalizability to other behaviors and the nature of this short-term boost effect. By utilizing the temporary boost of behavior changes to form sustainable changes it could be possible to develop an effective health promotion program especially targeting people with higher time preferences.

Limitations

There are several limitations of this study that should be considered. First, we did not conduct a prestratification for randomization by time preference. The sample may be biased between the intervention and control groups in each subgroup. Second, participants were recruited from one corporate group in Japan. The participation rate was very low (835/20,000, 4.2%). Most participants were married, working in clerical positions, and university graduates. They had their own PCs or tablet computers in their offices or homes. The participants were also supposed to have experience using a PC and studying through Web-based programs. Higher education level may also help participants learn from the iCBT program. The generalization of these findings to the general working population is limited. Third, while we excluded those who had MDE before, the scores of depressive symptoms and psychological distress of the participants at baseline were relatively high. These findings may be more applicable to respondents with mild depression. Fourth, the dropout rates in this study were 27.9% (197/706), 25.2% (178/706), and 31.6% (223/706) at the 3-, 6-, and 12-month follow-ups, respectively. The dropout rates were higher in the intervention group than in the control group during the entire follow-up period. The dropouts may have caused a selection bias, particularly if the intervention group participants with higher levels of depression were more likely to quit the program. Fifth, it is possible that

XSL•FO
RenderX

participants in the control group acquired information about the iCBT program from participants in the intervention group at the same workplace. This contamination could weaken the intervention effect. Sixth, all outcomes in this study were measured by self-report, which might have been affected by the perception of participants or by situational factors at work.

Conclusions

The iCBT program was significantly better at improving nonclinical depressive symptoms in the higher time preference subgroup compared with the lower time preference subgroup at the 3-month follow-up. Workers with higher time preferences may easily change their cognition or behavior, but the change may persist for only a short period. On the other hand, the pooled effect of iCBT during the entire follow-up period was significant only in the lower time preference subgroup. Workers with lower time preferences may be likely to keep their cognitive or behavioral changes for a longer period. A further RCT with a precise design, such as stratified permuted-block randomization, should be conducted to test the potential different intervention effects of the iCBT program on nonclinical depressive symptoms between lower and higher time preference subgroups.

Acknowledgments

The authors appreciate the help of the following persons in completing this project: Takayuki Narumi, Jun Naoi, Keisuke Kito, Chinatsu Narumi, Hayato Mori, Mitsuyasu Mizusaki, Chihiro Hoshino, and Aya Matsumoto.

Authors' Contributions

KI and NK conceived of the study, developed study design, conducted literature search, collected, analyzed, and interpreted data, and prepared the first draft. TAF, YM, AS, KKuribayashi, and KKasai developed study design and interpreted data. All authors reviewed the manuscript.

Conflicts of Interest

TAF has received lecture fees from Janssen, Meiji, Mitsubishi-Tanabe, MSD KK, and Pfizer. He has received research support from Mitsubishi-Tanabe. NK reports grants from Infocom Corp, Fujitsu Ltd, Fujitsu Software Technologies, and TAK Ltd, personal fees from Occupational Health Foundation, Japan Dental Association, Sekisui Chemicals, Junpukai Health Care Center, Osaka Chamber of Commerce and Industry, outside the submitted work. The other authors declare that they have no competing interests.

Multimedia Appendix 1

The time preference questionnaire used in the study and the number of respondents in each category.

[PDF File (Adobe PDF File), 77KB - jmir_v20i8e10231_app1.pdf]

Multimedia Appendix 2

CONSORT-EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 175KB - jmir_v20i8e10231_app2.pdf]

References

1. Hosman C, Jané-Llopis E, Saxena S. Prevention of mental disorders: Effective interventions and policy options: Summary report. Geneva: World Health Organization; 2004.
2. Kessler RC. The costs of depression. Psychiatr Clin North Am 2012 Mar;35(1):1-14 [FREE Full text] [doi: 10.1016/j.psc.2011.11.005] [Medline: 22370487]
3. Saarni SI, Suvisaari J, Sintonen H, Pirkola S, Koskinen S, Aromaa A, et al. Impact of psychiatric disorders on health-related quality of life: general population survey. Br J Psychiatry 2007 Apr;190:326-332 [FREE Full text] [doi: 10.1192/bjp.bp.106.025106] [Medline: 17401039]
4. Karsten J, Hartman CA, Smit JH, Zitman FG, Beekman ATF, Cuijpers P, et al. Psychiatric history and subthreshold symptoms as predictors of the occurrence of depressive or anxiety disorder within 2 years. Br J Psychiatry 2011 Mar;198(3):206-212 [FREE Full text] [doi: 10.1192/bjp.bp.110.080572] [Medline: 21357879]
5. Cuijpers P, Smit F, Willemse G. Predicting the onset of major depression in subjects with subthreshold depression in primary care: a prospective study. Acta Psychiatr Scand 2005 Feb;111(2):133-138. [doi: 10.1111/j.1600-0447.2004.00416.x] [Medline: 15667432]
6. van ZK, Buntrock C, Ebert DD, Smit F, Reynolds CF, Beekman ATF, et al. Preventing the onset of major depressive disorder: a meta-analytic review of psychological interventions. Int J Epidemiol 2014 Apr;43(2):318-329 [FREE Full text] [doi: 10.1093/ije/dyt175] [Medline: 24760873]

XSL•FO
RenderX

7. Cuijpers P, Muñoz RF, Clarke GN, Lewinsohn PM. Psychoeducational treatment and prevention of depression: the "Coping with Depression" course thirty years later. Clin Psychol Rev 2009 Jul;29(5):449-458. [doi: 10.1016/j.cpr.2009.04.005] [Medline: 19450912]
8. Christensen H, Griffiths KM, Jorm AF. Delivering interventions for depression by using the internet: randomised controlled trial. BMJ 2004 Jan 31;328(7434):265-7 concl [FREE Full text] [doi: 10.1136/bmj.37945.566632.EE] [Medline: 14742346]
9. Imamura K, Kawakami N, Furukawa TA, Matsuyama Y, Shimazu A, Umanodan R, et al. Effects of an Internet-based cognitive behavioral therapy (iCBT) program in Manga format on improving subthreshold depressive symptoms among healthy workers: a randomized controlled trial. PLoS One 2014 May;9(5):e97167 [FREE Full text] [doi: 10.1371/journal.pone.0097167] [Medline: 24844530]
10. Imamura K, Kawakami N, Furukawa T, Matsuyama Y, Shimazu A, Umanodan R, et al. Does Internet-based cognitive behavioral therapy (iCBT) prevent major depressive episode for workers? A 12-month follow-up of a randomized controlled trial. Psychol Med 2015 Jul;45(9):1907-1917. [doi: 10.1017/S0033291714003006] [Medline: 25562115]
11. Haby MM, Donnelly M, Corry J, Vos T. Cognitive behavioural therapy for depression, panic disorder and generalized anxiety disorder: a meta-regression of factors that may predict outcome. Aust N Z J Psychiatry 2006 Jan 21;40(1):9-19. [doi: 10.1080/j.1440-1614.2006.01736.x] [Medline: 16403033]
12. Lewis CC, Simons AD, Kim HK. The role of early symptom trajectories and pretreatment variables in predicting treatment response to cognitive behavioral therapy. J Consult Clin Psychol 2012 Aug;80(4):525-534. [doi: 10.1037/a0029131] [Medline: 22730951]
13. 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]
14. Warmerdam L, Van SA, Twisk J, Cuijpers P. Predicting outcome of Internet-based treatment for depressive symptoms. Psychother Res 2013;23(5):559-567. [doi: 10.1080/10503307.2013.807377] [Medline: 23848944]
15. 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]
16. Burkhouse KL, Kujawa A, Kennedy AE, Shankman SA, Langenecker SA, Phan KL, et al. NEURAL REACTIVITY TO REWARD AS A PREDICTOR OF COGNITIVE BEHAVIORAL THERAPY RESPONSE IN ANXIETY AND DEPRESSION. Depress Anxiety 2016 Apr;33(4):281-288 [FREE Full text] [doi: 10.1002/da.22482] [Medline: 27038409]
17. Barlow P, Reeves A, McKee M, Galea G, Stuckler D. Unhealthy diets, obesity and time discounting: a systematic literature review and network analysis. Obes Rev 2016 Dec;17(9):810-819 [FREE Full text] [doi: 10.1111/obr.12431] [Medline: 27256685]
18. Kang M, Ikeda S. Time discounting, present biases, and health-related behaviors: Evidence from Japan. Econ Hum Biol 2016 Dec;21:122-136. [doi: 10.1016/j.ehb.2015.09.005] [Medline: 26851396]
19. Barlow P, McKee M, Reeves A, Galea G, Stuckler D. Time-discounting and tobacco smoking: a systematic review and network analysis. Int J Epidemiol 2017 Jun 01;46(3):860-869 [FREE Full text] [doi: 10.1093/ije/dyw233] [Medline: 27818375]
20. Frederick S, Loewenstein G, O'donoghue T. Time Discounting and Time Preference: A Critical Review. Journal of Economic Literature 2002 Jun 21;40(2):351-401. [doi: 10.1257/002205102320161311] [Medline: 5637250]
21. Rothbard M. America's great depression. -: Ludwig von Mises Institute; 1972.
22. Zumwalt A. The role of time preference on wealth. -: University of Missouri-Columbia; 2008.
23. Imamura K, Kawakami N, Furukawa TA, Matsuyama Y, Shimazu A, Kasai K. Effects of an internet-based cognitive behavioural therapy intervention on preventing major depressive episodes among workers: a protocol for a randomised controlled trial. BMJ Open 2015 May 12;5(5):e007590 [FREE Full text] [doi: 10.1136/bmjopen-2015-007590] [Medline: 25968004]
24. Shimoda H, Inoue A, Tsuno K, Kawakami N. One-year test-retest reliability of a Japanese web-based version of the WHO Composite International Diagnostic Interview (CIDI) for major depression in a working population. Int J Methods Psychiatr Res 2015 Sep;24(3):204-212. [doi: 10.1002/mpr.1437] [Medline: 24677699]
25. Beck A, Steer R, Brown G. BDI-II, Beck Depression Inventory: Manual. San Antonio, Texas: Psychological Corp; 1996.
26. Hiroe T, Kojima M, Yamamoto I, Nojima S, Kinoshita Y, Hashimoto N, et al. Gradations of clinical severity and sensitivity to change assessed with the Beck Depression Inventory-II in Japanese patients with depression. Psychiatry Res 2005 Jun 30;135(3):229-235. [doi: 10.1016/j.psychres.2004.03.014] [Medline: 15996749]
27. Furukawa TA, Kawakami N, Saitoh M, Ono Y, Nakane Y, Nakamura Y, et al. The performance of the Japanese version of the K6 and K10 in the World Mental Health Survey Japan. Int J Methods Psychiatr Res 2008;17(3):152-158. [doi: 10.1002/mpr.257] [Medline: 18763695]
28. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SLT, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002 Aug;32(6):959-976. [Medline: 12214795]

XSL•FO
RenderX

29. Takada M, Kondo N, Hashimoto H, J-SHINE Data Management Committee. Japanese study on stratification, health, income, and neighborhood: study protocol and profiles of participants. J Epidemiol 2014;24(4):334-344 [FREE Full text] [Medline: 24814507]
30. Kang M, Ikeda S. Time discounting and smoking behavior: evidence from a panel survey(*). Health Econ 2014 Dec;23(12):1443-1464. [doi: 10.1002/hec.2998] [Medline: 24136867]
31. Faul F, Erdfelder E, Buchner A, Lang A. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods 2009 Nov 21;41(4):1149-1160. [doi: 10.3758/Brm.41.4.1149] [Medline: 5637250]
32. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods 2007 May 21;39(2):175-191. [doi: 10.3758/Bf03193146] [Medline: 5637250]
33. Cohen J. A power primer. Psychological Bulletin 1992 Mar 21;112(1):155-159. [doi: 10.1037/0033-2909.112.1.155] [Medline: 5637250]

Abbreviations

BDI-II: Beck Depression Inventory-II
CBT: cognitive behavioral therapy
ES: effect size
iCBT: internet-based computerized cognitive behavioral therapy
K6: Kessler's Psychological Distress Scale
MDD: major depressive disorder
MDE: major depressive episode
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 01.03.18; peer-reviewed by K Mathiasen, J Wang; comments to author 05.04.18; revised version received 05.05.18; accepted 15.05.18; published 10.08.18.

Please cite as:
Imamura K, Furukawa TA, Matsuyama Y, Shimazu A, Kuribayashi K, Kasai K, Kawakami N
Differences in the Effect of Internet-Based Cognitive Behavioral Therapy for Improving Nonclinical Depressive Symptoms Among Workers by Time Preference: Randomized Controlled Trial
J Med Internet Res 2018;20(8):e10231
URL: http://www.jmir.org/2018/8/e10231/
doi:10.2196/10231
PMID:30097419

©Kotaro Imamura, Toshi A Furukawa, Yutaka Matsuyama, Akihito Shimazu, Kazuto Kuribayashi, Kiyoto Kasai, Norito Kawakami. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 10.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Web-Based Acceptance-Facilitating Intervention for Identifying Patients' Acceptance, Uptake, and Adherence of Internet- and Mobile-Based Pain Interventions: Randomized Controlled Trial

Jiaxi Lin[1,2], MSc (Psych), PhD; Bianca Faust[3], MSc; David Daniel Ebert[4], PhD; Lena Krämer[5], Dipl-Psych, PhD; Harald Baumeister[6], Dipl-Psych, PhD

[1]Institute of Psychiatry, Psychology & Neuroscience, Health Psychology Department, King's College London, London, United Kingdom
[2]Institute of Sports and Sport Science, Department of Sport Psychology, University of Freiburg, Freiburg, Germany
[3]Rehabilitation-Center Todtmoos, Clinic Wehrawald, Todtmoos, Germany
[4]Clinical Psychology and Psychotherapy, Friedrich-Alexander University of Erlangen-Nürnberg, Erlangen, Germany
[5]Institute of Psychology, Department of Rehabilitation Psychology and Psychotherapy, University of Freiburg, Freiburg, Germany
[6]Institute of Psychology, Department of Clinical Psychology and Psychotherapy, University of Ulm, Ulm, Germany

Corresponding Author:
Harald Baumeister, Dipl-Psych, PhD
Institute of Psychology
Department of Clinical Psychology and Psychotherapy
University of Ulm
Albert-Einstein-Allee 47, D Ulm
Ulm, 89069
Germany
Phone: 49 07315032800
Fax: 49 07315032809
Email: harald.baumeister@uni-ulm.de

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2019/2/e12015/

Abstract

Background: Internet- and mobile-based interventions are effective for the treatment of chronic pain. However, little is known about patients' willingness to engage with these types of interventions and how the uptake of such interventions can be improved.

Objective: The aim of this study was to identify people's acceptance, uptake, and adherence (primary outcomes) with regard to an internet- and mobile-based intervention for chronic pain and the influence of an information video as an acceptance-facilitating intervention (AFI).

Methods: In this randomized controlled trial with a parallel design, we invited 489 individuals with chronic pain to participate in a Web-based survey assessing the acceptance of internet- and mobile-based interventions with the offer to receive an unguided internet- and mobile-based intervention for chronic pain after completion. Two versions of the Web-based survey (with and without AFI) were randomly sent to two groups: one with AFI (n=245) and one without AFI (n=244). Participants who completed the Web-based survey with or without AFI entered the intervention group or the control group, respectively. In the survey, the individuals' acceptance of pain interventions, measured with a 4-item scale (sum score ranging from 4 to 20), predictors of acceptance, sociodemographic and pain-related variables, and physical and emotional functioning were assessed. Uptake rates (log in to the intervention) and adherence (number of completed modules) to the intervention was assessed 4 months after intervention access. To examine which factors influence acceptance, uptake rate, and adherence in the internet- and mobile-based interventions, we conducted additional exploratory subgroup analyses.

Results: In total, 57 (intervention group) and 58 (control group) participants in each group completed the survey and were included in the analyses. The groups did not differ with regard to acceptance, uptake rate, or adherence (*P*=.64, *P*=.56, *P*=.75, respectively). Most participants reported moderate (68/115, 59.1%) to high (36/115, 31.3%) acceptance, with 9.6% (11/115)

XSL•FO
RenderX

showing low acceptance (intervention group: mean 13.91, SD 3.47; control group: mean 13.61, SD 3.50). Further, 67% (38/57, intervention group) and 62% (36/58, control group) had logged into the intervention. In both groups, an average of 1.04 (SD 1.51) and 1.14 (SD 1.90) modules were completed, respectively.

Conclusions: The informational video was not effective with regard to acceptance, uptake rate, or adherence. Despite the high acceptance, the uptake rate was only moderate and adherence was remarkably low. This study shows that acceptance can be much higher in a sample participating in an internet- and mobile-based intervention efficacy trial than in the target population in routine health care settings. Thus, future research should focus not only on acceptance and uptake facilitating interventions but also on ways to influence adherence. Further research should be conducted within routine health care settings with more representative samples of the target population.

Trial Registration: German Clinical Trial Registration DRKS00006183; http://www.drks.de/drks_web/navigate.do?navigationId=trial.HTML&TRIAL_ID=DRKS00006183 (Archived by WebCite at http://www.webcitation.org/70ebHDhne)

(J Med Internet Res 2018;20(8):e244) doi:10.2196/jmir.9925

KEYWORDS

uptake; acceptance; adherence; eHealth; chronic pain; randomized controlled trial

Introduction

Chronic pain as a disease in its own right is a major global health problem [1-3]. In the Global Burden of Disease Study of 2013 [4], low back pain, neck pain, and migraine, which often take a chronic course, were found among the top 10 causes of years lived with disability in every country under investigation. This not only reflects the high prevalence of chronic pain affecting 1 in 5 adults [2,5] but also the urgent need to improve health care. A large-scale survey of chronic pain in Europe found that 40% of the participants reported that their pain was inadequately controlled and only 2% were treated by pain specialists [2].

Therefore, innovative, effective, and cost-effective health care models for chronic pain are needed. This should include a multimodal, biopsychosocial approach, considered as the gold standard in the treatment of pain [6-8], with self-management where possible. In this context, internet- and mobile-based interventions (IMIs) might be a feasible means to provide psychological interventions such as cognitive behavioral interventions [9-14]. As most IMIs provide evidence-based strategies as interactive self-help lessons on a Web-based platform, they can reach large numbers of people simultaneously, anytime and anywhere [15-17]. A rising number of studies indicate the efficacy of IMIs for a wide range of mental and physical health conditions including chronic pain, depression, and anxiety [16-21]. A recent meta-analysis by Buhrman et al [9] on IMIs for chronic pain found small but significant positive effects for interference or disability, pain intensity, catastrophizing, and depression at Hedge's g=−0.39, g=−0.33, g=−0.49, and g=−0.26, respectively. A recent randomized controlled trial (RCT) on an IMI based on Acceptance and Commitment Therapy (ACT) showed guided, but not unguided, IMIs being effective in improving pain interference (Cohen *d*=0.58 at posttreatment and follow-up, respectively [22]) in individuals with chronic, nonmalignant pain for 6 months or longer. Moreover, changes in psychological flexibility mediated all outcomes of ACT-based online treatment for chronic pain (ACTonPain) [23], and cost-effectiveness analyses revealed that ACTonPain is potentially cost-effective, depending on the amount of society's willingness to pay [24]. In this trial, the uptake rate was 97% in both ACTonPain groups, and guided participants completed more modules (0-8) than those who were unguided (mean 5.94, SD 2.80 vs mean 4.74, SD 2.89, $F_{1,199}$=8.92; P<.01). The overall effect sizes in pain IMIs are in line with the effects of psychological interventions in face-to-face settings [25]. Hence, IMIs have the potential to improve chronic pain health care by providing evidence-based, possibly cost-effective psychological interventions [9,16-25] with high accessibility and scalability [15-17].

Two main barriers have repeatedly been discussed to limit the full potential of IMIs when implemented in routine health care settings: low uptake rates (logging into the intervention) and low levels of adherence (completing modules of the intervention) [26,27]. Evidence from multiple pragmatic studies examining depression IMIs implemented in real-life settings under less-structured and monitored conditions indicates that uptake rates vary between 3% and 25% [28-31]. Low intervention adherence in IMIs is a frequently reported problem as it can ultimately limit the effectiveness of IMIs [26,32-35]. In an RCT on the effectiveness of an ACT-based IMI for chronic pain, Trompetter et al [36] found that participants in the intervention and waitlist control group differed in pain interference only in the analysis with treatment completers. In routine health care settings, in contrast to developer-led efficacy trials on the same IMI [37-39], the issue of adherence seems to be particularly important when IMIs are offered.

A repeatedly suggested reason for low uptake and adherence is the low level of patients' acceptance of IMIs, conceptualized as the intention to use the intervention [40-42]. Other factors, such as internet usage and anxiety [41,43], uncertainty concerning data security, discomfort with use of IMIs and psychological interventions in general, and social influence by friends, family, and health professionals as well as a lack of trust in the effectiveness of IMIs are often reported to influence the acceptance and uptake of IMIs [40,42,44-47].

Aiming at these aspects of acceptance, acceptance-facilitating interventions (AFIs) are suggested to reduce patients´ apprehensions and misconceptions about IMIs. They provide trustworthy information on, as well as an introduction to IMIs [40,48-51]. To date, 3 RCTs have investigated the influence of a video-based [42,47] or personal [46] AFI in the clinical

XSL•FO
RenderX

population of pain [47], diabetes [46], and primary care patients with depressive symptoms [42]. All studies consistently reported low baseline acceptance and an increase in acceptance following AFI [42,46,47]. However, all three studies only examined patients' acceptance and lack more important information on whether AFI effectively increased intervention uptake.

Only two studies have reported on the relationship between IMI acceptance and IMI usage [27,52]. In both studies, a significant association was found between IMI acceptance and usage (log-in and adherence). This finding suggests that AFIs might also influence IMI usage. However, research on the influence of an AFI on intervention uptake and adherence is missing.

Therefore, in this study we examined whether an informational video (AFI) can increase patients' (1) acceptance of an IMI for chronic pain, (2) uptake of an IMI for chronic pain, and (3) adherence in an IMI for chronic pain.

We expected that AFI would positively increase patients' acceptance as well as the uptake rate and adherence. In addition, we expected that AFI would increase the predictors of acceptance and have a reducing effect on internet anxiety. To examine which factors influence acceptance, uptake rate, and adherence in IMIs, we conducted additional exploratory subgroup analyses.

Methods

Study Design

This study is linked to an outcome evaluation study with the German Clinical Trial Registration (DRKS): DRKS00006183, which is approved by the Ethics Committee of the Albert-Ludwigs-University of Freiburg. This trial is reported in accordance with the Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth checklist [53]. This was a two-arm pragmatic study using a parallel-group design with balanced (1:1) randomization. The intervention group (IG) received AFI with a subsequent Web-based survey (homepage provided by the University of Freiburg, Germany); the control group (CG) filled out the same Web-based survey without receiving AFI. In this RCT, randomization took place before the assessment of eligibility and inclusion of participants. We chose this procedure as it allowed us to send an invitation email providing a link to the survey in either the IG or CG condition. This is a case of randomization before data are available to confirm the individuals' eligibility without risking bias in the analysis [54]. Therefore, postrandomization exclusions of all noneligible participants can be regarded as acceptable [54].

Reading and providing online informed consent and answering the survey took about 20-30 minutes. After completing the survey, the participants could choose to receive the unguided version of ACTonPain [22,55] by providing their email address in order to access ACTonPain.

Sample

The recruitment took place in September 2015. We sent email invitations to all individuals to participate in this study who had earlier expressed interest in participating in an evaluation study on ACTonPain [22,55]. Individuals in the following categories could not be included in the evaluation study on ACTonPain for the following reasons: (1) screening or baseline assessment not completed or no informed consent for main trial (n=332) or (2) expressed their interest in participating after the target sample size of the main trial was reached (n=157). Applicants for participation in the main trial indicating an elevated risk of suicide were not invited. We assessed the following inclusion criteria based on the Web-based self-report: (1) ≥18 years of age, (2) pain duration ≥3 months, (3) sufficient German language skills, and (4) sufficient computer and internet skills to proceed with the Web-based questionnaire. We excluded all participants with an incomplete informed consent form and those not fulfilling the inclusion criteria. The intervention ACTonPain was conceptualized for chronic pain as a disease in its own right and not as a symptom of any specific disease (eg, chronic low back pain, migraine, or fibromyalgia) [1,55]. Moreover, ACT is applicable as a general therapeutic model [56], and therefore, no further specification concerning any specific disease related to chronic pain was made. All participants had full access to treatment as usual.

This study aimed at a minimal sample size of 102 participants to detect a clinically relevant medium effect size (Cohen d=0.50) with a power of 80% and a two-sided 5% significance level.

Randomization and Allocation

For allocation to IG or CG, a computer-generated list of random numbers with randomly varying block sizes of 4, 6, and 8 was used by BF (the sealedenvelope website). IG watched an AFI video before answering the Web-based questionnaire. CG filled out the questionnaire immediately. Out of 489 potential participants, 115 provided informed consent and fulfilled the inclusion criteria (Figure 1).

Sociodemographic Data, Clinical Characteristics, and Internet Usage

The questionnaire comprised sociodemographic items concerning age, sex, relationship status, education, and employment status. Moreover, we asked participants about current or past psychological pain treatment (yes or no) and how satisfied they were with it.

Pain

Participants evaluated their actual pain as well as the worst, least, and average pain during the last week on a scale from 0 to 10. Additionally, pain duration was assessed with 5 categories ranging from "1 month to 6 months" to ">5 years."

Physical Functioning

The Interference Scale of the Multidimensional Pain Inventory (MPI [57,58]) was used to measure the degree of pain interference with regard to all-day activities. This questionnaire is a valid measure of the interference of pain with physical functioning [59]. The Cronbach alpha in this study was at .91.

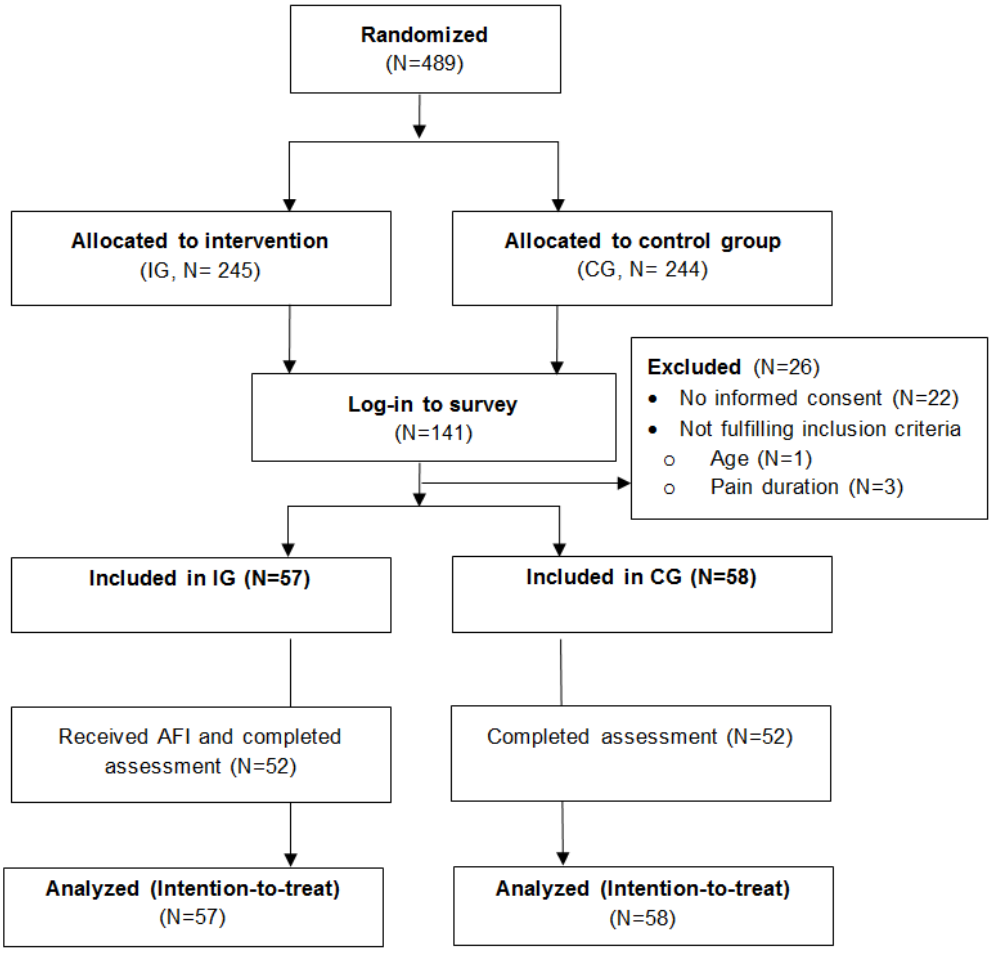
Emotional Functioning

We used the Patient Health Questionnaire depression scale (PHQ-8 [60-63]) for depressive symptomatology and the Generalized Anxiety Disorder Screener 7-item (GAD-7 [64])

XSL•FO
RenderX

for symptomatology associated with generalized anxiety disorder.

Figure 1. Flow chart. AFI: acceptance-facilitating intervention; CG: control group; IG: intervention group.



PHQ-8 assesses all Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-V) symptoms of major depression with the exception of suicidal or self-injurious thoughts during the last 2 weeks. Ratings are given on a 4-point Likert scale ranging from 0 “not at all” to 3 “nearly every day.” The scores for each item are summed up to produce a total score between 0 and 24 points. A cutoff score of 5-9 represents mild depressive symptoms; 10-14, moderate; 15-19, moderately severe; and 20-24, severe [61]. The Cronbach alpha of PHQ-8 was at .81 in this study.

GAD-7 consists of 7 core symptoms of the DSM-V diagnostic criteria A, B, and C for generalized anxiety [64]. The items are scored from 0 “not at all” to 3 “more than half the days” regarding the last 2 weeks. Scores range from 0 to 21; the cutoff points of 5, 10, and 15 represent the thresholds for mild, moderate, and severe anxiety symptom levels, respectively [64]. The Cronbach alpha in this study was at .87.

Internet Usage

We measured internet usage using the question “How often do you surf the internet?” Answers are rated on a 5-point Likert scale ranging from 1 “seldom or never” to 5 “multiple times per day.”

Acceptance-Facilitating Intervention

AFI consisted of a 3-minute introductory and information video to ACTonPain with screenshots of the program in order to improve patients’ acceptance. Figure 2 provides screenshots of AFI. We designed the content of the intervention to address the aforementioned barriers and drivers of acceptance. We conceptualized the video based on our previous AFIs that showed to be effective in increasing acceptance [42,47,65]. Our AFI is an adopted version of AFI used in a former study with individuals with chronic pain [47] with a specific introduction to ACTonPain. The content of the video comprised information on (1) the effectiveness of IMIs, (2) data security and anonymity in IMIs, (3) various advantages of IMIs (eg, ease and comfort of use, flexible time management), (4) the possibility of receiving technical support, and (5) assistance during the program. Furthermore, the video presented the process for using ACTonPain, encompassing the log-in or log-off processes and an overview of the modules and different features (audio files, video clips, and homework assignments).

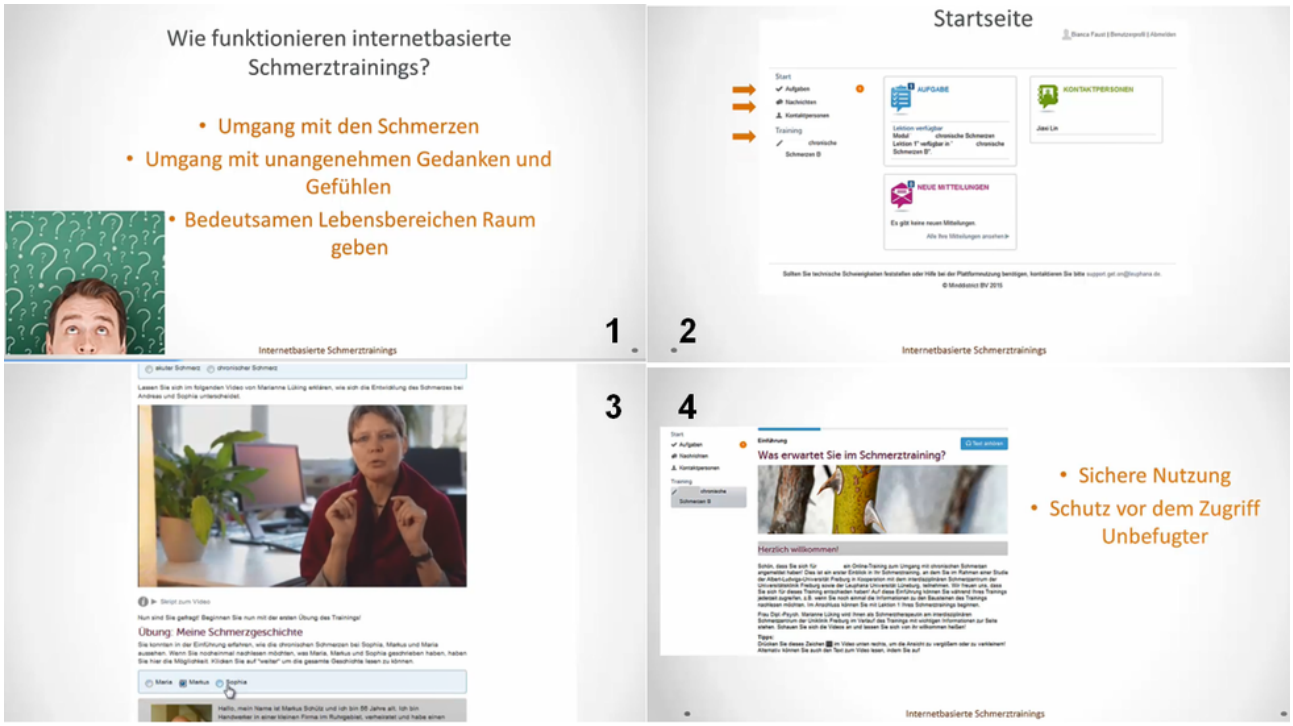
Acceptance and Commitment-Based Online Treatment for Chronic Pain

After completing the questionnaire, participants were invited to receive ACTonPain treatment in an unguided version and without short message service (SMS) text messages (SMS

XSL•FO
RenderX

Coach). This version of ACTonPain was provided without any human support and should be therefore of special interest for public health services due to its high scalability and low costs.

Figure 2. Screenshots of the acceptance-facilitating intervention. (1) content of an online pain intervention; (2) introduction to Acceptance and Commitment–based online treatment for chronic pain (ACTonPain) log-in page; (3) introduction to ACTonPain features; and (4) information concerning data security.



ACTonPain consists of an introduction and 7 consecutive modules. The intervention targets core change processes proposed by Hayes et al [56] and is described in more detail by Lin et al [55]. Participants were advised to complete one session per week with a completion time of approximately 60 minutes. Participants could access ACTonPain via the Web or on their mobile phone with an adapted mobile view. ACTonPain was not delivered as a mobile phone app, and AFI demonstrated the use of ACTonPain via the Web.

The effectiveness of ACTonPain was investigated in a three-armed RCT with a total of 302 participants who were randomly assigned to either ACTonPain guided, ACTonPain unguided, or waitlist control. Guidance was given by trained eCoaches (psychologists) who provided individualized standardized feedback after each module and reminded the participants to keep to the schedule of the treatment and set up deadlines. Additionally, participants could receive supportive SMS text messages (SMS Coach).

Measures

Primary Outcomes

The primary outcomes were acceptance, uptake, and adherence.

Acceptance

We operationalized acceptance on the basis of the well-established unified theory of acceptance and use of technology (UTAUT [66,67]). This framework provides a reliable theoretical basis of drivers and barriers for users' acceptance of information technology [66-68] and has been used in numerous IMIs studies [27,47,65,69-72]. The UTAUT model postulates acceptance as the intention to use technology and the proximal predictor for actual use [73].

The items of the UTAUT acceptance were developed based on previous studies [46,47]. The sum score of the scale ranges from 4 to 20, and the 3 levels of acceptance can be categorized: low (sum score: 4-9), moderate (sum score: 10-15), and high (sum score: 16-20). The Cronbach alpha in this study was relatively low at .71. Table 1 provides an overview of the items for acceptance and predictors of acceptance (see secondary outcomes) in this study, including their scales.

Uptake

We operationalized uptake as log-in (yes or no) to IMI assessed 4 months after intervention access. The period of 4 months was chosen, as this should have been enough time for the participants to start with the intervention and work through all 8 modules. We assumed that 4 months after intervention access is a reasonable time to assess uptake and adherence.

Adherence

We operationalized adherence as the number of completed modules of the intervention assessed 4 months after intervention access.

Secondary Outcomes

The secondary outcomes were the predictors of acceptance according to UTAUT as well as internet anxiety.

Predictors of Acceptance

According to the UTAUT model, there are 4 key predictors of either the behavioral intention or usage behavior of IT:

XSL•FO
RenderX

performance expectancy, effort expectancy, social influence, and facilitating conditions [67]. The items measuring the construct's performance expectancy and effort expectancy were drawn from Wilson and Lankton [74]. The items for social influence and facilitating conditions were adapted from Venkatesh et al [67].

Table 1. Items of acceptance and predictors of acceptance according to the unified theory of acceptance and use of technology model.

Outcomes	Items	Rating scale	Reliability
Acceptance	• If offered, I intend to try out an internet-based psychological pain intervention • If offered, I intend to use an internet-based psychological pain intervention regularly • I would recommend an internet-based psychological pain intervention to a friend • I am willing to pay for an internet-based psychological pain intervention	5-point scale (1 "does not apply at all" to 5 "applies completely")	.71
Predictors of acceptance			
Performance expectancy	• Using an internet-based psychological pain intervention would increase the effectiveness of my pain treatment • Using an internet-based psychological pain intervention would be beneficial for my health care • Overall, an internet-based psychological pain intervention would support me in coping with my chronic pain	5-point scale (1 "does not apply at all" to 5 "applies completely")	.86
Effort expectancy	• Using an internet-based psychological pain intervention would be simple • Using an internet-based psychological pain intervention would be an easy task for me • An internet-based psychological pain intervention would be clear and easily comprehensible to me	5-point scale (1 "does not apply at all" to 5 "applies completely")	.79
Social influence	• People close to me would recommend me to use an internet-based psychological pain intervention • My general practitioner would recommend me to use an internet-based psychological pain intervention	5-point scale (1 "does not apply at all" to 5 "applies completely")	.69
Facilitating conditions	• I do have all necessary technical preconditions for using an internet-based psychological pain intervention • In case of technical problems with an internet-based psychological pain intervention, I would receive technical support	5-point scale (1 "does not apply at all" to 5 "applies completely")	Two separate items, not a uniform scale

Internet Anxiety

Two items for internet anxiety were adapted from Venkatesh et al [67] (1) "The internet is something threatening to me" and (2) "I am afraid of making an irrevocable mistake while using the internet"). The items were rated on a 5-point Likert scale ranging from 1 "does not apply at all" to 5 "applies completely." The Cronbach alpha in this study was at .69.

Statistical Analyses

We conducted data analysis using SPSS Statistics 22 (IBM Corporation, Armonk, NY, USA). Descriptive statistics are provided for sociodemographic data and functioning to describe the sample. To test for randomization imbalance between IG and CG, we employed chi-square tests and *t*-tests for independent samples. The descriptive statistics were based on nonimputed data, while all following analyses were conducted after multiple imputations with 20 imputations using the imputation algorithm implemented in SPSS (intention-to-treat analysis).

To detect differences between IG and CG regarding acceptance, uptake, adherence, and the predictors of acceptance as well as internet anxiety, we conducted *t*-tests for independent samples and chi-square tests. In case of significant group differences, standardized mean differences (Cohen *d*) with a 95% CI were computed to quantify the effect. As this study includes multiple primary outcomes, we used a Bonferroni adjustment for the *P* values of .02 (3 tests at an alpha level of .05). This procedure resulted in sufficient statistical power with the sample to detect differences between the two conditions that were larger than Cohen *d*=0.65.

To examine potential subgroup differences (age, gender, education, pain duration and intensity, prior or present psychological intervention, internet usage and anxiety, and physical and emotional functioning) regarding acceptance, uptake, and adherence, exploratory analyses are provided (mean, SD, *t*-tests, and chi-square test). For this purpose, variables were dichotomized using defined cutoffs (gender, pain duration, education, and psychological intervention) or a median split (age, pain intensity, internet usage and anxiety, physical and emotional functioning, and level of acceptance regarding uptake and adherence). Note that the results of the subgroup analyses and analysis on secondary outcomes are exploratory and

XSL•FO
RenderX

underpowered; adjusting for multiple testing would not be meaningful [75].

Results

Participants

Of 489 persons, 141 (28.8%) responded to the invitation. After we excluded those who did not provide informed consent (n=22) or did not fulfill the inclusion criteria (n=4), we included 57 and 58 participants in IG and CG, respectively. The missing value was between 0% and 5.7% per variable, and Little's Missing Completely at Random test indicated that the data were missing at random (χ^2_{41}=45.31, *P*=.30).

The majority (82/115, 71.3%) of the participants were female. Ages ranged from 18 to 76 years with a mean age of 50.42 (SD 13.67) years. The majority of the sample (96/115, 83.5%) reported a pain duration of longer than 2 years, with 57.4% (66/115) suffering for more than 5 years. In addition, 86.1% (99/115) and 75.7% (87/115) of the participants reported at least mild symptomatology of depression and anxiety, respectively. Table 2 shows sociodemographic and clinical characteristics as well as internet usage in the sample. No significant differences for demographic and pain- and function-related variables were found between the two groups except in regard to employment, as more participants in CG were employed. Differences between IG and CG in all outcomes are summarized in Table 3.

Primary Outcomes

There was no significant difference between IG and CG with regard to acceptance, uptake, or adherence (*P*=.64, *P*=.56, *P*=.75, respectively). Among the total sample, 8.7% (10/115) showed a low, 59.1% (68/115) a moderate, and 31.3% (36/115) a high level of acceptance, with an average sum score of 13.76 (SD 3.54). Figure 3 displays the levels of acceptance in both groups. The participants who applied for access to ACTonPain numbered 48 in IG and 50 in CG.

Note that 9% (5/57) and 10% (6/58) of participants in IG and CG, respectively, did not complete the survey and therefore did not indicate whether they wanted to receive the intervention. Then, 7% (4/57) and 3% (2/58) of participants in IG and CG, respectively, did not want to receive the intervention, and 84% (48/57) and 86% (50/58) of participants in IG and CG, respectively, signed up to receive the intervention. Four months after receiving access to ACTonPain, 65.2% (75/115) of the sample had logged in. This represents an uptake rate of 68% (38/57, IG) and 62% (36/58, CG). With regard to adherence, the participants completed 1.09 (SD 1.72) modules on average. That is, the average participant only completed the introduction module. The results showed that 5.2% (6/115) participants did not complete any modules after log-in and 3.5% (4/115) completed all the modules in the study. Hence, the treatment dropout rate was at 96.5% (111/115). Figure 4 presents the number of log-ins and completed modules in each group.

Secondary Outcomes

There was no significant difference between IG and CG with regard to performance expectancy, effort expectancy, social influence, facilitating conditions, or internet anxiety (*P*=.88, *P*=.16, *P*=.96, *P*=.69, *P*=.68, respectively; Table 2).

Subgroup Analyses

Since there were no group effects, we conducted the subgroup analyses with no group consideration in order to increase the power of the analyses. Participants with lower internet anxiety and higher anxiety symptoms showed significantly higher acceptance than their equivalent counterparts (Table 4). With regard to uptake rates, more participants with higher depressive symptoms (45/60, 75%) and acceptance (47/59, 80%) logged into the platform than those with lower depressive symptoms (30/55, 55%) and acceptance (28/56, 50%). We also found that participants with a higher level of acceptance completed more modules compared with participants with a lower level of acceptance (1.43 vs 0.72 modules).

XSL•FO
RenderX

Table 2. Sociodemographic and clinical characteristics and internet usage.

Characteristics	Total (N=115)	Intervention group (n=57)	Control group (n=58)	*P* value[a]
Sociodemographic characteristics				
Age (years), mean (SD)	50.42 (13.32)	51.65 (14.02)	49.21 (12.60)	.33
Sex (female), n (%)	82 (71.3)	33 (57.9)	39 (67.2)	.53
Married or in a relationship, n (%)	77 (66.9)	37 (64.9)	40 (69.0)	.69
Educational level, n (%)[b]				
No school-leaving qualification	25 (21.7)	13 (22.8)	12 (20.7)	.82
Lower secondary	10 (8.7)	7 (12.3)	3 (5.2)	.20
Middle secondary	20 (17.4)	8 (14.0)	12 (20.7)	.46
Higher secondary	6 (5.2)	2 (3.5)	4 (6.9)	.68
Highest secondary	12 (10.4)	6 (10.5)	6 (10.3)	.99
Vocational training	24 (20.9)	14 (24.6)	8 (13.8)	.16
University degree	18 (15.6)	5 (8.8)	11 (19.0)	.18
Employment				
(Self-) Employed, n (%)	63 (54.8)	24 (42.1)	39 (67.2)	.01
Pain				
Intensity, mean (SD)	4.62 (1.72)	4.83 (1.35)	4.55 (1.88)	.46
Duration, n (%)				
3-6 months	2 (1.7)	1 (1.7)	1 (1.7)	.99
1-2 years	17 (14.8)	6 (10.5)	11 (19.0)	.29
2-5 years	30 (26.1)	17 (29.8)	13 (22.4)	.40
Over 5 years	66 (57.4)	33 (57.9)	33 (56.9)	.99
Prior psychological pain treatment[c] (n=111), n (%)	49 (42.6)	22 (38.6)	27 (46.5)	.41
Current psychological pain treatment[c] (n=111), n (%)	24 (20.9)	13 (22.8)	11 (19.0)	.65
Physical functioning				
Multidimensional Pain Inventory, mean (SD)	3.79 (1.09)	3.81 (1.04)	3.77 (1.34)	.84
Emotional functioning (n=108)				
PHQ-8[d], mean (SD)[c]	10.67 (4.86)	10.79 (4.79)	10.55 (4.97)	.79
GAD-7[e], mean (SD)[c]	8.23 (4.86)	8.42 (5.24)	8.04 (4.46)	.69
Internet usage, mean (SD)	3.43 (1.21)	3.21 (1.18)	3.64 (1.22)	.06

[a]The *P* value refers to the significance level of the test on differences between the intervention and control groups on sociodemographic and clinical characteristics and internet usage.

[b]Secondary education according to the German classification: “Hauptschule” (“lower,” 9 years, until age 15/16), “Realschule” (“middle,” 10 years, until age 16/17), “Fachhochschulreife” (“higher,” 12 years, until age 17/18), “Abitur” (“highest,” 12 or 13 years, until age 17-19).

[c]Incomplete data.

[d]PHQ-8: Patient Health Questionnaire depression scale.

[e]GAD-7: Generalized Anxiety Disorder Screener 7-item.

XSL•FO
RenderX

Table 3. Differences between the intervention and control groups in all outcomes (intention-to-treat analysis dataset).

Outcomes	Total (N=115)	Intervention group (n=57)	Control group (n=58)	*P* value
Primary outcomes				
Acceptance, mean (SD)	13.76 (3.54)	13.91 (3.47)	13.61 (3.50)	.64
Uptake, n (%)	75 (65.2)	39 (68.4)	36 (62.1)	.56
Adherence, mean (SD)	1.09 (1.72)	1.04 (1.51)	1.14 (1.90)	.75
Secondary outcomes, mean (SD)				
Performance expectancy	9.82 (2.79)	9.78 (3.10)	9.86 (2.51)	.88
Effort expectancy	10.85 (2.90)	10.47 (3.02)	11.23 (2.82)	.16
Social influence	5.88 (2.36)	5.89 (2.42)	5.87 (2.36)	.96
Facilitating conditions	7.45 (2.14)	7.53 (2.11)	7.37 (2.21)	.69
Internet anxiety	3.15 (1.61)	3.21 (1.51)	3.09 (1.60)	.68

Figure 3. Level of acceptance. CG: control group; IG: intervention group.

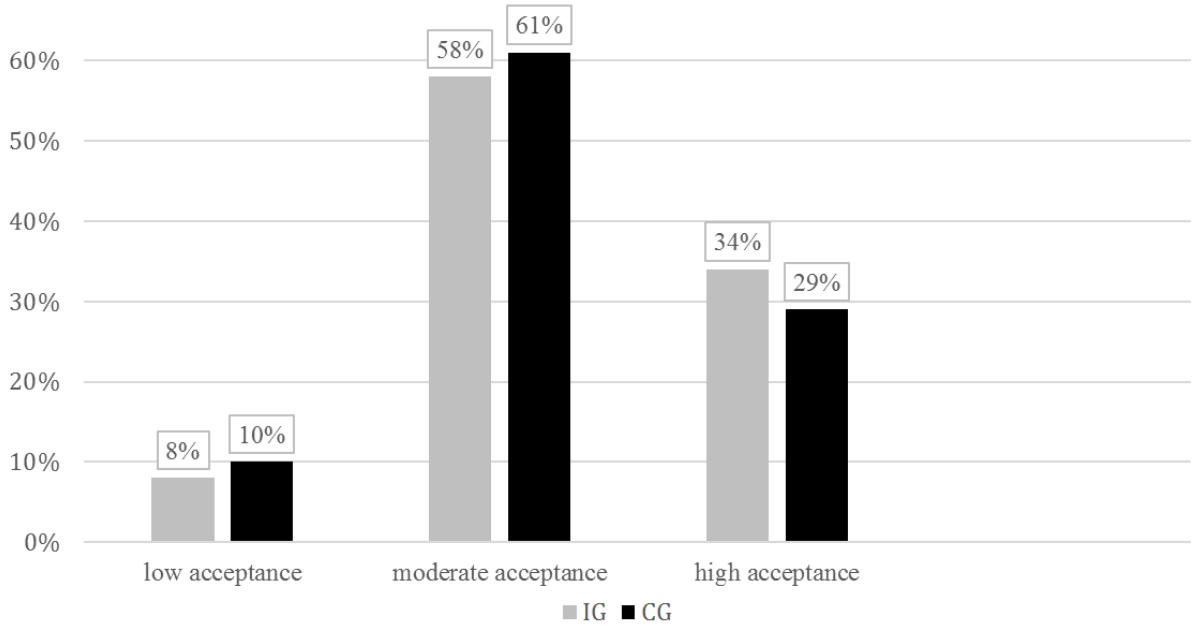
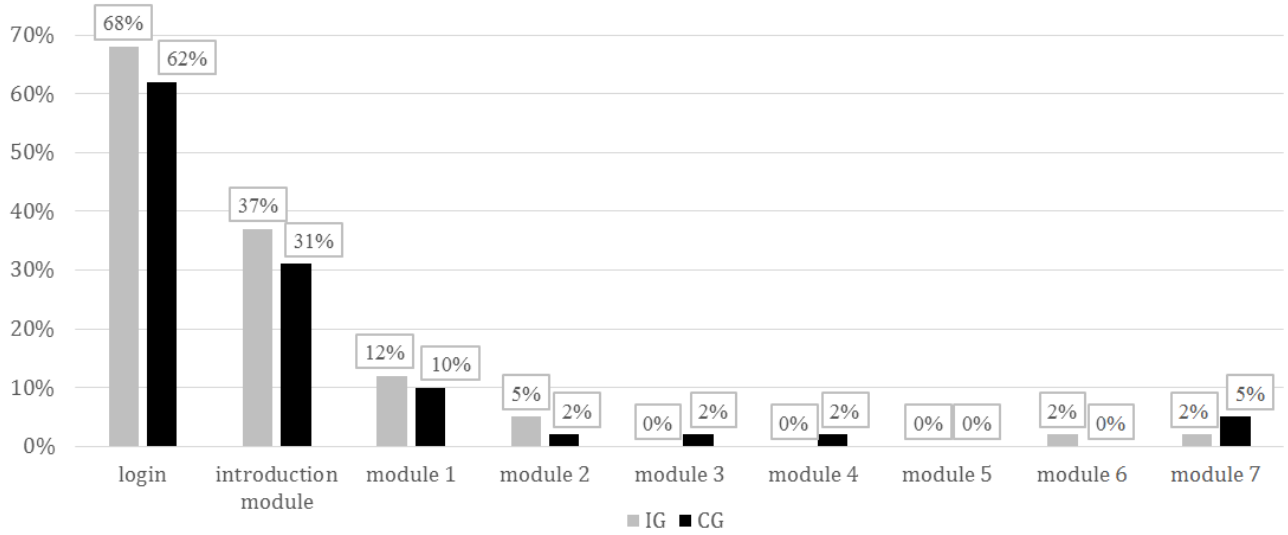


Figure 4. Number of log-ins and completed modules. CG: control group; IG: intervention group.



XSL•FO
RenderX

Table 4. Subgroup-specific effects on acceptance, uptake, and adherence, intention-to-treat analysis dataset.

Subgroups	Acceptance		Uptake rate		Adherence	
	Mean (SD)	*P* value	n (%)	*P* value	Mean (SD)	*P* value
Age		.64		.33		.85
<51 (n=54)	13.92 (2.87)		38 (70)		1.06 (1.62)	
≥51 (n=61)	13.62 (3.91)		37 (61)		1.11 (1.81)	
Sex		.44		.99		.89
Female (n=8)	13.93 (3.26)		53 (64)		1.07 (1.69)	
Male (n=33)	13.35 (4.05)		22 (67)		1.12 (1.82)	
Educational level		.63		.44		.50
Low[a] (n=60)	13.61 (2.98)		37 (62)		0.98 (1.71)	
High[b] (n=55)	13.93 (2.80)		38 (69)		1.20 (1.87)	
Pain intensity		.75		.17		.78
<4.50 (n=52)	13.65 (3.60)		30 (58)		1.04 (1.73)	
≥4.50 (n=63)	13.85 (3.49)		45 (71)		1.13 (1.86)	
Pain duration		.52		.32		.37
<5 years (n=49)	13.52 (3.57)		29 (59)		0.92 (1.60)	
≥5 years (n=66)	13.94 (3.41)		46 (70)		1.21 (1.80)	
Former psychological intervention		.43		.99		.77
Ever (n=52)	13.59 (3.97)		34 (65)		1.18 (1.80)	
Never (n=63)	14.11 (2.86)		41 (65)		1.08 (1.75)	
Internet usage		.19		.08		.76
<4.00 (n=56)	13.31 (3.67)		32 (57)		1.04 (1.75)	
≥4.00 (n=59)	14.19 (3.30)		43 (73)		1.14 (1.71)	
Internet anxiety		<.001		.19		.17
<3.00 (n=58)	15.04 (2.89)		40 (73)		1.32 (2.06)	
≥3.00 (n=57)	12.58 (3.58)		35 (58)		0.87 (1.39)	
Physical functioning (Multidimensional Pain Inventory)		.05		.58		.45
<3.90 (n=57)	13.12 (3.40)		36 (63)		1.21 (1.85)	
≥3.90 (n=58)	14.39 (3.47)		39 (67)		0.97 (1.58)	
Emotional functioning						
PHQ-8[c]		.06		.04		.58
<10.00 (n=55)	13.12 (3.32)		30 (55)		0.99 (1.83)	
≥10.00 (n=60)	14.35 (3.54)		45 (75)		1.17 (1.63)	
GAD-7[d]		.02		.47		.60
<8.00 (n=57)	13.03 (3.37)		35 (61)		1.00 (1.55)	
≥8.00 (n=58)	14.49 (3.50)		40 (69)		1.17 (1.90)	
Acceptance		—		<.001		.03
<14.00 (n=56)	—		28 (50)		0.72 (1.30)	
≥14.00 (n=59)	—		47 (80)		1.43 (2.02)	

[a]Low: no school-leaving qualification-higher secondary.

[b]High: highest secondary-university degree.

XSL•FO
RenderX

[c]PHQ-8: Patient Health Questionnaire depression scale.

[d]GAD-7: Generalized Anxiety Disorder Screener 7-item.

Discussion

Principal Findings

To the best of our knowledge, this study is the first to examine the impact of AFI on patients' acceptance, actual uptake, and adherence of an IMI. AFI consisted of a short informational video.

In this study, the average level of acceptance indicated a moderate to high acceptance in the sample (mean 13.76, SD 3.54) with no group differences between IG and CG. This acceptance level is higher than the levels examined in equivalent previous studies [42,46,47]. In these studies, acceptance levels in the intervention group after receiving AFI were at a mean of 11.42 (SD 4.28), 12.17 (SD 4.22), and 10.55 (SD 4.69) in samples of patients with depression [42], pain [47], and diabetes [46], respectively, in routine health care settings. The control groups in these studies displayed substantially lower levels of acceptance below the sum score of 10, indicating a low acceptance level on average. Contrary to previous studies, AFI in our study did not influence acceptance and its predictors, performance expectancy, effort expectancy, social influence, facilitating conditions, or internet anxiety. This might be due to the high baseline level of acceptance in the sample.

The comparatively high acceptance in both groups of this study is potentially due to selective sampling. We recruited the participants from a pool of persons who had already expressed interest in participating in a previous study on ACTonPain. After the end of recruitment for the main study, we invited all persons who were not randomized in the study to participate in this study and to receive ACTonPain as an incentive after completion of the survey. Hence, the participants in this study expressed their interest for ACTonPain twice. Therefore, the level of acceptance most likely reflects the acceptance and uptake in many IMI efficacy studies consisting of a population that is considerably more interested and open to IMIs than the general population [76]. Therefore, our previous work on acceptance in the general population [42,46,47] might give us a more realistic estimate of acceptance. By comparing the acceptance rates throughout the studies, this study quantifies how acceptance and uptake rates can differ between populations in efficacy studies and routine health care settings.

Despite the high level of acceptance, the uptake rate was only moderate at 65.2%. In comparison, the uptake rate in the main evaluation of ACTonPain [22,55] was at 97% in the guided and unguided group, respectively. Furthermore, adherence was considerably low in both groups, again without any influence of AFI. Similar to the results on acceptance, there was no difference between the two groups with regard to uptake and adherence rates, indicating that AFI did not influence uptake rates either. An explanation of why AFI did not influence intervention uptake and adherence is that it targeted acceptance rather than intervention use.

In conclusion, AFI had no effects in a sample with high initial acceptance. This is in line with the assumptions of the Health Action Process Approach (HAPA [77]), which disentangles the processes of intention formation and intention implementation. According to this model, a behavioral intention (ie, acceptance in the UTAUT model) is a necessary but not sufficient precondition of behavior change [78]. In HAPA, three groups of persons are differentiated: nonintenders, intenders, and actors. Each group needs specific behavioral interventions. While nonintenders profit from self-efficacy interventions and information about pros and cons of the behavior change in order to increase behavioral intention, intenders and actors must be provided with concrete help on how to implement their intentions in actual behavior [77]. This includes concrete action planning (when, where, and how to act), coping planning (how to deal with barriers), social support, and action control. By considering these postintentional tasks, HAPA extends the scope of UTAUT and explains the whole range of behavior change, along with the process of intention formation.

Applying the assumptions of HAPA to our sample, an AFI might be the wrong means to increase intervention uptake and adherence. The participants showed high intentions to use an IMI, which means they are classified as intenders in the sense of HAPA. Instead of an AFI, which provides information on efficacy, data security, etc, our participants might have profited more from concrete action and coping planning, action control, and social support. This assumption is supported by a recent study of Zarski et al [79], where planning, out of all the investigated HAPA variables, was the strongest predictor of treatment adherence in highly motivated participants in an IMI.

With most participants only completing the introduction module, adherence is substantially lower in this study than in the evaluation study of the exact same intervention [22,55]. There is little empirical evidence yet on what constitutes an optimal dose of an intervention, for either face-to-face, individual, group, or Web-based interventions. According to the framework of psychological flexibility as the theoretical basis of ACT, the 6 underlying subprocesses are hypothesized to promote higher psychological flexibility as the main goal in ACT [56]. Regarding ACTonPain, this would implicate that participants should have worked on modules 1-6 in order to benefit from ACTonPain. In comparison to this study, participants in the ACTonPain evaluation study completed 5.94 (SD 2.80) and 4.74 (SD 2.89) modules in the guided and unguided groups, respectively. Only 3.5% (4/115) participants completed all 8 modules in this study. In the ACTonPain evaluation study, 60 and 40 participants completed all modules in the guided and unguided groups, respectively. These differences in completion rates are another indicator that participants profit from support in implementing their behavioral intentions. In the ACTonPain evaluation study, participants were enrolled in a broad study procedure, received support from the study team on how to create an account, and were asked to fill out all assessments, including two after randomization. The whole trial procedure might have supported intention implementation via strategies such as reminding prompts or social support [80].

XSL•FO
RenderX

The different findings concerning adherence and dropout not only highlight the influence of guidance and SMS but also the importance of the setting in which participants receive IMIs. Guidance and prompts (eg, through SMS text messages) are two of the most investigated adherence facilitating factors in the research on IMIs, ultimately increasing the effect of the respective IMI [35,81-86]. However, the absence of guidance and SMS Coach alone cannot fully explain the difference in adherence and dropout between the two studies. As there were no following assessments and further administrational contact in this study like in efficacy evaluation trials on IMIs, this study likely imitates a real-life setting. Hence, this finding is consistent with a number of effectiveness trials indicating that the actual use of IMIs is substantially lower when IMIs are implemented in real-life settings [28-31]. Should such findings be replicated in future studies, this could indicate that effect sizes for unguided interventions found in clinical RCTs might be substantially overestimated for what can be expected when embedded in routine health care settings [38]. In conclusion, the findings on adherence and dropout in this study provide an estimate on the use of IMIs when offered in routine health care settings.

In addition to the abovementioned high baseline acceptance, the rather general content of AFI might also explain why AFI was not effective in this study. As discussed in a previous study on acceptance of IMIs in patients with diabetes [65], AFIs tailored to the specific concerns and needs of the respective population might be more effective. The exploratory subgroup analyses in this study showed a trend wherein less anxious (GAD-7<8.00) participants with higher internet anxiety had lower acceptance. Therefore, an AFI with information that is more specific to the characteristics of individuals with lower acceptance might have been more effective. However, as this study was not designed and sufficiently powered to reliably detect heterogeneity in various subgroups, these findings need to be interpreted with caution.

Limitations

Several limitations in this study are noteworthy. First, the recruiting strategy might have influenced the way the participants filled out the survey, and their answers might have been more socially desirable. Consequently, the results on acceptance and uptake might not be representative for the population of patients with chronic pain, but they are likely to be representative for the population of patients with chronic pain in previous efficacy trials on IMIs for chronic pain. Hence, this study provides information on participants' acceptance in efficacy studies that can be useful for the interpretation of their respective results. This is especially the case regarding their generalizability to routine clinical practice given that most of these studies are conducted under ideal circumstances with highly specified inclusion and exclusion criteria [76].

In connection with the abovementioned lack of implementation facilitating factors in our AFI, a further limitation of this study is that it is only based on the UTAUT model. The UTAUT model and other equivalent models on the acceptance of IMIs as evaluated in a previous study [74], as well as in some empirical studies [27,52], suggest a relationship between acceptance and IMI use but might not consider sustained use, which is required in IMIs. Therefore, the findings of our study indicate that adherence facilitating factors are crucial even when acceptance is high. Hence, future research is needed to test interventions aimed at increasing adherence. HAPA can serve as an intervention model.

Finally, the reliability of the acceptance scale was relatively low at .71 compared with previous studies (Cronbach alpha ranged from .84 [42] to .87 [46,47]). However, the Cronbach alpha in this study is still in an acceptable range, especially as the scale consists of only 4 items [87].

Conclusions

Overall, this study yields evidence that patients' uptake and adherence to an IMI for chronic pain is low, despite high acceptance. The first main contribution of this study is that it shows how acceptance rates can differ between a sample participating in an IMI efficacy trial as represented in this study and a sample collected from a routine health care setting, represented in our previous studies [42,47,65]. This discrepancy should be kept in mind when efficacy trials are interpreted and also when IMIs should be implemented in routine health care settings. In the context of routine health care settings, educational level and motivation are likely to differ from IMI efficacy trial settings [76]. Therefore, effectiveness studies aimed at the clinical target groups in the respective health care settings are needed. As an example, in two studies on an IMI for the treatment [88] and prevention [89] of depression in patients with back pain, recruitment took place following orthopedic rehabilitation. These studies were designed to reach the entire potential target population within a naturalistic setting where the aftercare IMI was implemented. The results of these studies can therefore provide more generalizable results on the effectiveness of IMIs than most of the efficacy trials.

This study also indicates that high acceptance does not guarantee sustainable use of IMIs. Further models, such as the HAPA model, need to be used in order to develop strategies to increase adherence in IMIs. Equivalent to the discussion on acceptance rates, the different settings of efficacy trials and routine health care settings appear to play a crucial role for adherence in IMIs. This might explain the high discrepancy between adherence in this study and the evaluation study of the same intervention [22,55], as well as in a recent meta-analysis on adherence in IMIs for depression [90]. In this review, Van Ballegooijen et al concluded that adherence to guided IMIs (81% of IMI was completed on average) appears to be equal to adherence to face-to-face therapies (84%). Similarly, Christensen et al [91] found dropout rates in IMIs for anxiety and depression to be similar to dropout rates reported in the context of noninternet-based treatments. The findings of this study, however, indicate that when IMIs are offered in routine health care settings, attrition rates might be higher and be a problem specific for IMIs. This is especially the case when guidance as the most important adherence facilitating factor [82] is not provided. Apart from guidance and prompts, it is unclear what specific technological features improve adherence and outcome. Therefore, investigations on attrition and adherence and their underlying mechanisms are needed. In addition to AFIs, engagement facilitating interventions to increase the continuous

XSL•FO
RenderX

use of IMIs need to be developed and evaluated. This should comprise constant support systems during the beginning and throughout the use of IMIs, such as continuous monitoring of patients' health care providers [42]. At this point, it becomes evident that not only IMI users but also their developers and providers need to become involved in order to maximize the acceptance, adherence, and eventually the effectiveness of evidence-based IMIs.

In conclusion, this study shows that acceptance can be much higher in a sample participating in an IMI efficacy trial than in the target population in routine health care settings. Therefore, future research should be conducted within naturalistic settings with more representative samples. Further, strategies to increase adherence in IMIs need to be developed involving IMI users, developers, and providers.

Acknowledgments

The authors would like to thank the participants who made it possible to conduct the study. This work was supported by a postdoctoral fellowship of the German Academic Exchange Service.

Authors' Contributions

JL, HB, and BF designed the study, and DDE and LVK contributed to the design. BF, DDE, HB, and JL developed AFI content and the assessment. BF performed the data collection. JL and BF analyzed the data. JL was a major contributor in writing the manuscript. All authors contributed to the interpretation of the data and the further writing of the manuscript, and read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT - EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 4MB - jmir_v20i8e244_app1.pdf]

References

1. Goldberg DS, McGee SJ. Pain as a global public health priority. BMC Public Health 2011 Oct 06;11:770 [FREE Full text] [doi: 10.1186/1471-2458-11-770] [Medline: 21978149]
2. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. Eur J Pain 2006 May;10(4):287-333. [doi: 10.1016/j.ejpain.2005.06.009] [Medline: 16095934]
3. Breivik H, Eisenberg E, O'Brien T. The individual and societal burden of chronic pain in Europe: the case for strategic prioritisation and action to improve knowledge and availability of appropriate care. BMC Public Health 2013 Dec 24;13:1229 [FREE Full text] [doi: 10.1186/1471-2458-13-1229] [Medline: 24365383]
4. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015 Aug 22;386(9995):743-800 [FREE Full text] [doi: 10.1016/S0140-6736(15)60692-4] [Medline: 26063472]
5. Gureje O, Von KM, Simon GE, Gater R. Persistent pain and well-being: a World Health Organization Study in Primary Care. JAMA 1998 Jul 8;280(2):147-151. [Medline: 9669787]
6. Gatchel RJ, Peng YB, Peters ML, Fuchs PN, Turk DC. The biopsychosocial approach to chronic pain: scientific advances and future directions. Psychol Bull 2007 Jul;133(4):581-624. [doi: 10.1037/0033-2909.133.4.581] [Medline: 17592957]
7. Turk DC, Wilson HD, Cahana A. Treatment of chronic non-cancer pain. Lancet 2011 Jun 25;377(9784):2226-2235. [doi: 10.1016/S0140-6736(11)60402-9] [Medline: 21704872]
8. Kerns RD, Sellinger J, Goodin BR. Psychological treatment of chronic pain. Annu Rev Clin Psychol 2011;7:411-434. [doi: 10.1146/annurev-clinpsy-090310-120430] [Medline: 21128783]
9. Buhrman M, Gordh T, Andersson G. Internet interventions for chronic pain including headache: A systematic review. Internet Interventions 2016 May;4:17-34. [doi: 10.1016/j.invent.2015.12.001]
10. Bender JL, Radhakrishnan A, Diorio C, Englesakis M, Jadad AR. Can pain be managed through the Internet? A systematic review of randomized controlled trials. Pain 2011 Aug;152(8):1740-1750. [doi: 10.1016/j.pain.2011.02.012] [Medline: 21565446]
11. Rini C, Williams DA, Broderick JE, Keefe FJ. Meeting them where they are: Using the Internet to deliver behavioral medicine interventions for pain. Transl Behav Med 2012 Mar;2(1):82-92 [FREE Full text] [doi: 10.1007/s13142-011-0107-2] [Medline: 22924084]

XSL•FO
RenderX

12. Eccleston C, Fisher E, Craig L, Duggan GB, Rosser BA, Keogh E. Psychological therapies (Internet-delivered) for the management of chronic pain in adults. Cochrane Database Syst Rev 2014;2:CD010152. [doi: 10.1002/14651858.CD010152.pub2] [Medline: 24574082]
13. Macea DD, Gajos K, Daglia CYA, Fregni F. The efficacy of Web-based cognitive behavioral interventions for chronic pain: a systematic review and meta-analysis. J Pain 2010 Oct;11(10):917-929. [doi: 10.1016/j.jpain.2010.06.005] [Medline: 20650691]
14. Williams DA. Web-based behavioral interventions for the management of chronic pain. Curr Rheumatol Rep 2011 Dec;13(6):543-549. [doi: 10.1007/s11926-011-0212-8] [Medline: 21913061]
15. Lin J, Ebert DD, Lehr D, Berking M, Baumeister H. [Internet based cognitive behavioral interventions: state of the art and implementation possibilities in rehabilitation]. Rehabilitation (Stuttg) 2013 Jun;52(3):155-163. [doi: 10.1055/s-0033-1343491] [Medline: 23761204]
16. Baumeister H, Lin J, Ebert DD. Internetbasierte Gesundheitsinterventionen. In: Koch U, Bengel J, editors. Enzyklopädie der Psychologie: Medizinische Psychologie, Band 2. Göttingen: Hogrefe; 2017:231-258.
17. Ebert DD, Cuijpers P, Muñoz RF, Baumeister H. Prevention of Mental Health Disorders Using Internet- and Mobile-Based Interventions: A Narrative Review and Recommendations for Future Research. Front Psychiatry 2017;8:116 [FREE Full text] [doi: 10.3389/fpsyt.2017.00116] [Medline: 28848454]
18. Andersson G, Carlbring P, Ljótsson B, Hedman E. Guided Internet-Based CBT for Common Mental Disorders. J Contemp Psychother 2013 May 21;43(4):223-233. [doi: 10.1007/s10879-013-9237-9]
19. Andersson G, Ljótsson B, Weise C. Internet-delivered treatment to promote health. Curr Opin Psychiatry 2011 Mar;24(2):168-172. [doi: 10.1097/YCO.0b013e3283438028] [Medline: 21285706]
20. Cuijpers P, van Straten A, Andersson G. Internet-administered cognitive behavior therapy for health problems: a systematic review. J Behav Med 2008 Apr;31(2):169-177 [FREE Full text] [doi: 10.1007/s10865-007-9144-1] [Medline: 18165893]
21. Ebert DD, Zarski A, Christensen H, Stikkelbroek Y, Cuijpers P, Berking M, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. PLoS One 2015;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
22. Lin J, Paganini S, Sander L, Lüking M, Ebert DD, Buhrman M, et al. An Internet-Based Intervention for Chronic Pain. Dtsch Arztebl Int 2017 Oct 13;114(41):681-688 [FREE Full text] [doi: 10.3238/arztebl.2017.0681] [Medline: 29082858]
23. Lin J, Klatt LI, McCracken LM, Baumeister H. Psychological flexibility mediates the effect of an online-based acceptance and commitment therapy for chronic pain: an investigation of change processes. Pain 2018 Apr;159(4):663-672. [doi: 10.1097/j.pain.0000000000001134] [Medline: 29320375]
24. Paganini S, Lin J, Kahlke F, Buntrock C, Leiding D, Ebert DD. A guided and unguided internet- and mobile-based intervention for chronic pahealth economic evaluation alongside a randomized controlled trial. (forthcoming).
25. Eccleston C, Morley SJ, Williams ACDC. Psychological approaches to chronic pain management: evidence and challenges. Br J Anaesth 2013 Jul;111(1):59-63 [FREE Full text] [doi: 10.1093/bja/aet207] [Medline: 23794646]
26. 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]
27. Chiu TML, Eysenbach G. Stages of use: consideration, initiation, utilization, and outcomes of an internet-mediated intervention. BMC Med Inform Decis Mak 2010;10:73 [FREE Full text] [doi: 10.1186/1472-6947-10-73] [Medline: 21092275]
28. Kaltenthaler E, Sutcliffe P, Parry G, Beverley C, Rees A, Ferriter M. The acceptability to patients of computerized cognitive behaviour therapy for depression: a systematic review. Psychol Med 2008 Nov;38(11):1521-1530. [doi: 10.1017/S0033291707002607] [Medline: 18205964]
29. Lillevoll KR, Vangberg HCB, Griffiths KM, Waterloo K, Eisemann MR. Uptake and adherence of a self-directed internet-based mental health intervention with tailored e-mail reminders in senior high schools in Norway. BMC Psychiatry 2014;14:14 [FREE Full text] [doi: 10.1186/1471-244X-14-14] [Medline: 24443820]
30. Whiteside U, Richards J, Steinfeld B, Simon G, Caka S, Tachibana C, et al. Online cognitive behavioral therapy for depressed primary care patients: a pilot feasibility project. Perm J 2014;18(2):21-27 [FREE Full text] [doi: 10.7812/TPP/13-155] [Medline: 24867546]
31. Woodford J, Farrand P, Bessant M, Williams C. Recruitment into a guided internet based CBT (iCBT) intervention for depression: lesson learnt from the failure of a prevalence recruitment strategy. Contemp Clin Trials 2011 Sep;32(5):641-648. [doi: 10.1016/j.cct.2011.04.013] [Medline: 21570485]
32. 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]
33. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen1 JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
34. Cuijpers P, van SA, van SA, Andersson G. Psychological treatment of depression in primary care: a meta-analysis. Br J Gen Pract 2009 Feb;59(559):e51-e60 [FREE Full text] [doi: 10.3399/bjgp09X395139] [Medline: 19192368]

XSL•FO
RenderX

35. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
36. Trompetter HR, Bohlmeijer ET, Veehof MM, Schreurs KMG. Internet-based guided self-help intervention for chronic pain based on Acceptance and Commitment Therapy: a randomized controlled trial. J Behav Med 2015 Feb;38(1):66-80. [doi: 10.1007/s10865-014-9579-0] [Medline: 24923259]
37. Kok R, Beekman A, Cuijpers P, van Straten A. Internet Interventions. 2017 Sep. Adherence to a web-based pre-treatment for phobias in outpatient clinics URL: https://www.sciencedirect.com/science/article/pii/S2214782916300409
38. Ebert DD, Baumeister H. Internet-Based Self-help Interventions for Depression in Routine Care. JAMA Psychiatry 2017 Aug 01;74(8):852-853. [doi: 10.1001/jamapsychiatry.2017.1394] [Medline: 28636717]
39. Spek V, Nyklícek I, Smits N, Cuijpers P, Riper H, Keyzer J, et al. Internet-based cognitive behavioural therapy for subthreshold depression in people over 50 years old: a randomized controlled clinical trial. Psychol Med 2007 Dec;37(12):1797-1806. [doi: 10.1017/S0033291707000542] [Medline: 17466110]
40. Mitchell N, Gordon P. Attitudes Towards Computerized CBT for Depression Amongst a Student Population. Behav. Cognit. Psychother 2007 May 14;35(04):421. [doi: 10.1017/S1352465807003700]
41. Jimison H, Gorman P, Woods S, Nygren P, Walker M, Norris S, et al. Barriers and drivers of health information technology use for the elderly, chronically ill, and underserved. Evid Rep Technol Assess (Full Rep) 2008 Nov(175):1-1422. [Medline: 19408968]
42. Ebert DD, Berking M, Cuijpers P, Lehr D, Pörtner M, Baumeister H. Increasing the acceptance of internet-based mental health interventions in primary care patients with depressive symptoms. A randomized controlled trial. J Affect Disord 2015 May 1;176:9-17. [doi: 10.1016/j.jad.2015.01.056] [Medline: 25682378]
43. Celik V, Yesilyurt E. Attitudes to technology, perceived computer self-efficacy and computer anxiety as predictors of computer supported education. Computers & Education 2013 Jan;60(1):148-158. [doi: 10.1016/j.compedu.2012.06.008]
44. Gun SY, Titov N, Andrews G. Acceptability of Internet treatment of anxiety and depression. Australas Psychiatry 2011 Jun;19(3):259-264. [doi: 10.3109/10398562.2011.562295] [Medline: 21682626]
45. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. Psychol Med 2014 Apr;44(6):1303-1317 [FREE Full text] [doi: 10.1017/S0033291713001943] [Medline: 23931656]
46. Baumeister H, Nowoczin L, Lin J, Seifferth H, Seufert J, Laubner K, et al. Impact of an acceptance facilitating intervention on diabetes patients' acceptance of Internet-based interventions for depression: a randomized controlled trial. Diabetes Res Clin Pract 2014 Jul;105(1):30-39. [doi: 10.1016/j.diabres.2014.04.031] [Medline: 24862240]
47. Baumeister H, Seifferth H, Lin J, Nowoczin L, Lüking M, Ebert D. Impact of an Acceptance Facilitating Intervention on Patients' Acceptance of Internet-based Pain Interventions: A Randomized Controlled Trial. Clin J Pain 2015 Jun;31(6):528-535. [doi: 10.1097/AJP.0000000000000118] [Medline: 24866854]
48. Cranen K, Veld RHI, Ijzerman M, Vollenbroek-Hutten M. Change of patients' perceptions of telemedicine after brief use. Telemed J E Health 2011 Sep;17(7):530-535. [doi: 10.1089/tmj.2010.0208] [Medline: 21767150]
49. Crutzen R, de Nooijer J, Brouwer W, Oenema A, Brug J, de Vries NK. Strategies to facilitate exposure to internet-delivered health behavior change interventions aimed at adolescents or young adults: a systematic review. Health Educ Behav 2011 Feb;38(1):49-62. [doi: 10.1177/1090198110372878] [Medline: 21189422]
50. 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]
51. Brouwer W, Oenema A, Crutzen R, de Nooijer J, 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]
52. Ma J, Xiao L, Blonstein AC. Measurement of self-monitoring web technology acceptance and use in an e-health weight-loss trial. Telemed J E Health 2013 Oct;19(10):739-745 [FREE Full text] [doi: 10.1089/tmj.2013.0009] [Medline: 23952787]
53. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
54. Fergusson D, Aaron SD, Guyatt G, Hébert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. BMJ 2002 Sep 21;325(7365):652-654 [FREE Full text] [Medline: 12242181]
55. Lin MJ, Lüking M, Ebert DD, Buhrman M, Andersson G, Baumeister H. Effectiveness and cost-effectiveness of a guided and unguided internet-based Acceptance and Commitment Therapy for chronic pain: Study protocol for a three-armed randomised controlled trial. Internet Interventions 2015 Mar;2(1):7-16. [doi: 10.1016/j.invent.2014.11.005]
56. Hayes S, Strosahl K. A Practical Guide to Acceptance and Commitment Therapy. New York: Springer; 2004.
57. Flor H, Rudy T, Birbaumer N, Streit B, Schugens M. [The applicability of the West Haven-Yale multidimensional pain inventory in German-speaking countries. Data on the reliability and validity of the MPI-D.]. Schmerz 1990 Jun;4(2):82-87. [doi: 10.1007/BF02527839] [Medline: 18415223]

XSL•FO
RenderX

58. Kerns RD, Turk DC, Rudy TE. The West Haven-Yale Multidimensional Pain Inventory (WHYMPI). Pain 1985 Dec;23(4):345-356. [Medline: 4088697]
59. Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. Pain 2003 Dec;106(3):337-345. [Medline: 14659516]
60. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
61. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
62. Löwe B, Kroenke K, Herzog W, Gräfe K. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). J Affect Disord 2004 Jul;81(1):61-66. [doi: 10.1016/S0165-0327(03)00198-8] [Medline: 15183601]
63. Erbe D, Eichert H, Rietz C, Ebert D. Interformat reliability of the patient health questionnaire: Validation of the computerized version of the PHQ-9. Internet Interventions 2016 Sep;5:1-4. [doi: 10.1016/j.invent.2016.06.006]
64. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
65. Baumeister H, Nowoczin L, Lin J, Seifferth H, Seufert J, Laubner K, et al. Impact of an acceptance facilitating intervention on diabetes patients' acceptance of Internet-based interventions for depression: a randomized controlled trial. Diabetes Res Clin Pract 2014 Jul;105(1):30-39. [doi: 10.1016/j.diabres.2014.04.031] [Medline: 24862240]
66. Taiwo A, Downe A. The theory of user acceptance and use of technology (UTAUT): A meta-analytic review of empirical findings. J Theor Appl Inf Technol. ? 2013;49(1):48-58.
67. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 2003;27(3):425-481. [doi: 10.2307/30036540] [Medline: 10758835]
68. Khechine H, Lakhal S, Ndjambou P. A meta-analysis of the UTAUT model: Eleven years later. Can J Adm Sci 2016 Jun 06;33(2):138-152. [doi: 10.1002/cjas.1381]
69. Thiart H, Lehr D, Ebert DD, Berking M, Riper H. Log in and breathe out: internet-based recovery training for sleepless employees with work-related strain - results of a randomized controlled trial. Scand J Work Environ Health 2015 Mar;41(2):164-174. [doi: 10.5271/sjweh.3478] [Medline: 25590336]
70. Chiu TML, Eysenbach G. Theorizing the health service usage behavior of family caregivers: a qualitative study of an internet-based intervention. Int J Med Inform 2011 Nov;80(11):754-764. [doi: 10.1016/j.ijmedinf.2011.08.010] [Medline: 21958550]
71. van der Vaart R, Atema V, Evers AWM. Guided online self-management interventions in primary care: a survey on use, facilitators, and barriers. BMC Fam Pract 2016 Mar 09;17:27 [FREE Full text] [doi: 10.1186/s12875-016-0424-0] [Medline: 26961547]
72. Hennemann S, Beutel ME, Zwerenz R. Drivers and Barriers to Acceptance of Web-Based Aftercare of Patients in Inpatient Routine Care: A Cross-Sectional Survey. J Med Internet Res 2016 Dec 23;18(12):e337 [FREE Full text] [doi: 10.2196/jmir.6003] [Medline: 28011445]
73. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
74. Wilson EV, Lankton NK. Modeling patients' acceptance of provider-delivered e-health. J Am Med Inform Assoc 2004;11(4):241-248 [FREE Full text] [doi: 10.1197/jamia.M1475] [Medline: 15064290]
75. Bender R, Lange S. Adjusting for multiple testing--when and how? J Clin Epidemiol 2001 Apr;54(4):343-349. [Medline: 11297884]
76. Andersson G, Carlbring P, Cuijpers P. Internet interventions: Moving from efficacy to effectiveness. EJAP 2010 Jan 03;5(2):24. [doi: 10.7790/ejap.v5i2.169]
77. Schwarzer R, Lippke S, Luszczynska A. Mechanisms of health behavior change in persons with chronic illness or disability: the Health Action Process Approach (HAPA). Rehabil Psychol 2011 Aug;56(3):161-170. [doi: 10.1037/a0024509] [Medline: 21767036]
78. Webb TL, Sheeran P. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychol Bull 2006 Mar;132(2):249-268. [doi: 10.1037/0033-2909.132.2.249] [Medline: 16536643]
79. Zarski A, Berking M, Reis D, Lehr D, Buntrock C, Schwarzer R, et al. Turning Good Intentions Into Actions by Using the Health Action Process Approach to Predict Adherence to Internet-Based Depression Prevention: Secondary Analysis of a Randomized Controlled Trial. J Med Internet Res 2018 Jan 11;20(1):e9 [FREE Full text] [doi: 10.2196/jmir.8814] [Medline: 29326097]
80. Krämer L, Fuchs R. Barrieren und Barrierenmanagement im Prozess der Sportteilnahme. Zeitschrift für Gesundheitspsychologie 2010 Oct;18(4):170-182. [doi: 10.1026/0943-8149/a000026]
81. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]

XSL•FO
RenderX

82. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — A systematic review. Internet Interventions 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
83. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
84. Childs JD, Teyhen DS, Van Wyngaarden JJ, Dougherty BF, Ladislas BJ, Helton GL, et al. Predictors of web-based follow-up response in the Prevention Of Low Back Pain In The Military Trial (POLM). BMC Musculoskelet Disord 2011 Jun 13;12:132 [FREE Full text] [doi: 10.1186/1471-2474-12-132] [Medline: 21668961]
85. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]
86. 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]
87. Field A, Andy. Discovering Statistics Using IBM SPSS Statistics. California: SAGE PUBN; 2018.
88. Lin J, Sander L, Paganini S, Schlicker S, Ebert D, Berking M, et al. Effectiveness and cost-effectiveness of a guided internet- and mobile-based depression intervention for individuals with chronic back pain: protocol of a multi-centre randomised controlled trial. BMJ Open 2017 Dec 28;7(12):e015226 [FREE Full text] [doi: 10.1136/bmjopen-2016-015226] [Medline: 29288172]
89. Sander L, Paganini S, Lin J, Schlicker S, Ebert DD, Buntrock C, et al. Effectiveness and cost-effectiveness of a guided Internet- and mobile-based intervention for the indicated prevention of major depression in patients with chronic back pain-study protocol of the PROD-BP multicenter pragmatic RCT. BMC Psychiatry 2017 Jan 21;17(1):36 [FREE Full text] [doi: 10.1186/s12888-017-1193-6] [Medline: 28109247]
90. van Ballegooijen W, Cuijpers P, van Straten A, Karyotaki E, Andersson G, Smit JH, et al. Adherence to Internet-based and face-to-face cognitive behavioural therapy for depression: a meta-analysis. PLoS One 2014;9(7):e100674 [FREE Full text] [doi: 10.1371/journal.pone.0100674] [Medline: 25029507]
91. 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]

Abbreviations

AFI: acceptance-facilitating intervention
ACT: Acceptance and Commitment Therapy
ACTonPain: Acceptance and Commitment-based online treatment for chronic pain
CG: control group
DSM-V: Diagnostic and Statistical Manual of Mental Disorders, fifth edition
GAD-7: Generalized Anxiety Disorder Screener 7-item
HAPA: Health Action Process Approach
IG: intervention group
IMI: internet- and mobile-based intervention
PHQ-8: Patient Health Questionnaire depression scale
RCT: randomized controlled trial
SMS: short message service
UTAUT: unified theory of acceptance and use of technology

Edited by G Eysenbach; submitted 13.02.18; peer-reviewed by K Rosen, M Alshehri; comments to author 08.03.18; revised version received 26.04.18; accepted 12.05.18; published 21.08.18.

Please cite as:
Lin J, Faust B, Ebert DD, Krämer L, Baumeister H
A Web-Based Acceptance-Facilitating Intervention for Identifying Patients' Acceptance, Uptake, and Adherence of Internet- and Mobile-Based Pain Interventions: Randomized Controlled Trial
J Med Internet Res 2018;20(8):e244
URL: http://www.jmir.org/2018/8/e244/
doi:10.2196/jmir.9925
PMID:30131313

XSL•FO
RenderX

©Jiaxi Lin, Bianca Faust, David Daniel Ebert, Lena Krämer, Harald Baumeister. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 21.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Feasibility and Effects of Digital Interventions to Support People in Recovery From Substance Use Disorders: Systematic Review

Sverre Nesvåg[1], PhD; James R McKay[1,2,3], PhD

[1]Centre for Alcohol and Drug Research, Stavanger University Hospital, Stavanger, Norway

[2]Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[3]Philadelphia VA Medical Center, Philadelphia, PA, United States

Corresponding Author:
Sverre Nesvåg, PhD
Centre for Alcohol and Drug Research
Stavanger University Hospital
PO Box 8100
Stavanger,
Norway
Phone: 47 908 37 431
Email: ness@sus.no

Abstract

Background: The development and evaluation of digital interventions aimed at preventing or treating substance use–related problems and disorders is a rapidly growing field. Previous reviews of such interventions reveal a large and complex picture with regard to targeted users, use, and efficacy.

Objective: The objective of this review was to investigate the feasibility and effects of interventions developed specifically for digital platforms. These interventions are focused on supporting people in recovery from substance use disorders by helping them achieve their substance use goals and develop a more satisfying life situation.

Methods: The review is based on a systematic search in MEDLINE, Embase, PsycInfo, and Cochrane Library databases. Of the 1149 identified articles, 722 were excluded as obviously not relevant. Of the remaining articles, 21 were found to be previous reviews, 269 were on interventions aimed at reducing hazardous alcohol or cannabis use, and 94 were on digitized versions of standard treatment methods. The remaining 43 articles were all read in full and systematically scored by both authors.

Results: The 43 articles cover 28 unique interventions, of which 33 have been published after 2013. The interventions are aimed at different target groups (defined by age, substance, or comorbidity). Based on the number of features or modules, the interventions can be categorized as simple or complex. Fourteen of the 18 simple interventions and 9 of the 10 complex interventions have been studied with quantitative controlled methodologies. Thirteen of the 18 simple interventions are integrated in other treatment or support systems, mainly delivered as mobile phone apps, while 6 of the 10 complex interventions are designed as stand-alone interventions, most often delivered on a platform combining desktop/Web and mobile phone technologies. The interventions were generally easy to implement, but in most cases the implementation of the complex interventions was found to be dependent on sustained organizational support. Between 70% and 90% of the participants found the interventions to be useful and easy to use. The rates of sustained use were also generally high, except for simple interventions with an open internet-based recruitment and some information and education modules of the complex interventions. Across all interventions, slightly more than half (55%) of the studies with control groups generated positive findings on 1 or more substance use outcomes, with 57% of the interventions also found to be efficacious in 1 or more studies. In the positive studies, effects were typically in the small to moderate range, with a few studies yielding larger effects. Largely due to the inclusion of stronger control conditions, studies of simple interventions were less likely to produce positive effects.

Conclusions: The digital interventions included in this review are in general feasible but are not consistently effective in helping people in recovery from substance use disorder reduce their substance use or achieving other recovery goals.

(J Med Internet Res 2018;20(8):e255) doi:10.2196/jmir.9873

KEYWORDS

digital interventions; substance use disorders; recovery support; feasibility; effects

XSL•FO
RenderX

Introduction

Treatment for substance use disorders (SUD) can be effective. However, individuals who enter treatment often struggle with factors that are slow to change or do not change at all, placing them at heightened risk for relapse for considerable lengths of time [1]. These include genetic factors, interpersonal problems, co-occurring psychiatric disorders, employment problems, and various neurocognitive conditions [2-6]. Moreover, most positive factors associated with recovery, such as the development of supportive social networks, interests and passions that reinforce abstinence, improved coping responses, employment, and other activities that provide a sense of worth and self-esteem, are slow to change and require ongoing support to prevent deterioration [7-9].

These findings may explain why treatments derived from an acute care model are of limited effectiveness in the long-term management of SUD. Specifically, vulnerability to relapse remains relatively high for significant periods after standard treatment protocols of 3 to 6 months have ended [10,11]. Better management requires longer periods of continued contact with the patient [9,12-14] to address flagging motivation, increased craving, diminished participation in self/mutual help, limitations in neurocognitive function, continued biological vulnerability to stress, and various other problems that arise. Therefore, extended treatment, otherwise known as continuing care, is often recommended to patients.

In addition to continuing care interventions focused on substance use, there are additional sources of long-term recovery support including mutual help programs such as Alcoholics Anonymous. Individuals who attend these programs often have good substance use outcomes, but only a minority of people who might benefit actually attend any meetings and very few continue to participate at a high level over long periods of time [15,16].

There is some evidence that interventions to improve housing and employment status produce improved substance use outcomes. For example, work by Silverman and colleagues [17] has shown that a therapeutic workplace intervention improves substance use outcomes and employment status over periods as long as 5 years for homeless individuals with SUD. Milby and colleagues [18] found that adding abstinence-contingent housing and work therapy to standard care improved short-term substance use and housing outcomes for homeless cocaine-dependent individuals but not 12-month outcomes [19]. In a second study, Milby et al [20] found that housing, whether contingent on abstinence or not, produced better substance use outcomes than no housing out to 6 months. However, providing housing did not improve housing or employment outcomes over 12 months relative to the no-housing condition.

Evidence from well-done randomized studies also supports the efficacy of recovery check-ups and case management in the longer term management of SUD. Brief quarterly check-ups designed to identify individuals with out-of-control drug use following treatment and quickly re-engage them in SUD care have improved substance use outcomes over 4 years relative to standard care, although the magnitude of the effects was small [21]. Intensive case management provided over 12 months has been shown to improve substance use outcomes and employment in welfare recipients [22,23].

Although extended treatment for SUD is effective [24], the magnitude of the effects is often not large and tends to decrease over time [9]. There are several reasons for this:

- Information on relapse risk is only obtained during treatment sessions. Some relapse vulnerability factors can change rapidly—over periods as short as a few hours—often with little or no warning. A continuing care intervention in which data on relapse risks are obtained only during treatment sessions cannot be responsive to sudden shifts in risk level between sessions.
- Counselor availability is limited. Patients are urged to contact their counselors if they experience increases in relapse risk in between regularly scheduled sessions. However, such increases often come during evenings and weekends or when therapists are not available for other reasons.
- Procedures for marshaling other recovery supports are slow and cumbersome. Patients are urged to call peers in recovery and other supports when they feel at risk for relapse. However, patients may not have the necessary information when they most need it. They may also hesitate to reach out due to embarrassment or shame.

In the search for solutions to these challenges, digital interventions have become increasingly popular. Previous reviews of such interventions [25,26] reveal a large and complex picture with regard to targeted users, use, and efficacy. For example, these interventions have been developed for 3 distinct groups: those with hazardous alcohol or drug use, those currently in treatment for SUD, and those in recovery from SUD after undergoing treatment.

The concept “digital” also covers a variety of methodological strategies and elements. These interventions may contain a single element or a more complex collection of elements that build on digitized methods previously used in face-to-face interventions or methods uniquely developed for the current digital intervention. They may be meant to function as a stand-alone intervention or as an element in a larger intervention/support program. They may contain no interactive elements or different interactive elements in the form of automated responses or online real-time communication.

The technological solutions and platforms also vary. Some interventions are delivered on computer-based platforms (desktop or Web-based applications), while others are based on mobile phone technology platforms. Interventions feature a range of technological platforms, such as information websites or apps, assessment and monitoring technologies, automated or interactive voice response, text messaging, and chat rooms.

Papers on digital interventions for hazardous or risky alcohol use have been reviewed several times in the last 3 years, and papers on interventions for cannabis use were reviewed by Tait et al [27] in 2013. Dedert et al [26] describe these kinds of interventions as brief normative feedback on self-reported alcohol consumption, much in line with what is known as

XSL•FO
RenderX

Screening and Brief Intervention (SBI). Dedert et al [26] found that the results of these kinds of interventions are much the same as found in nondigital SBI; a small short-term effect (consumption at 6 months follow-up) but no long-term effects. These results are confirmed in a newly published Cochrane review by Kaner et al [28]. In a meta-analytical comparison of digital and in-person delivered interventions, Cadigan et al [29] found no difference between the 2 modalities on short-term effect (less than 4 months), while the in-person interventions had stronger long-term effects. The same conclusion was drawn in the reviews conducted by Dotson et al [30] and Leeman et al [31], while Huh et al [32] even questioned the short-term effect of digital SBI. Tait et al [27], in their review of interventions for cannabis use, describe both the content of the interventions and the results regarding reduction in cannabis use, arriving at the same results as found in reviews of interventions for alcohol use.

In a review of 7 recent studies using more technologically developed program elements, Berman et al [33] were not able to find more positive results than in the traditional interventions. Interactive voice response interventions showed some short-term positive results on consumption, while text messaging and mobile phone apps showed no significant effects. A study by Cunningham et al [34] compared a brief with an extended intervention (AlcoholHelpCentre.net) and found that the extended intervention did not increase the effect. In recent studies, however, promising results have been obtained from adding new elements to standard interventions, such as gamification [35], booster email sessions [36,37], skills training via mobile phone apps [38], individually tailored text messaging [39], and Facebook delivery of personalized normative feedbacks [40].

Studies have also been done to evaluate digitized versions of existing interventions such as cognitive behavioral therapy (CBT) [41], motivational interviewing (MI) [42], and cognitive enhancement therapies such as cognitive training [43] or mentalization-based therapy [44]. None of the previous reviews on digital interventions report on results from these kinds of interventions, and it is outside the scope of this paper to do so. With the large number of interventions of this kind now developed and researched, such a review would be welcomed.

In this review, our prime interest is in new interventions developed specifically for digital platforms focused on supporting people in recovery from SUD by helping them achieve their substance use goals (eg, stay abstinent, using less or having a less damaging use pattern) and develop a more satisfying life situation. The review will cover the whole variety of interventions with regard to methodological strategies and technological solutions. In the review, we are interested in the feasibility of the interventions and effects on substance use and other aspects of recovery.

Methods

Search and Evaluation Criteria

In this review, our aim is to synthesize results on feasibility and effects found in primary study reports on digital interventions focused on supporting people in recovery from SUD. The concept “digital interventions” covers a range of intervention methods delivered through digital channels, and therefore a traditional meta-analytic review of the results of a specific method is not possible. However, a systematic approach was taken to identify relevant publications for our review of feasibility and effects. First, this review was based on a systematic search for studies in several databases using specific criteria on target groups, types of interventions, and outcomes. Second, the studies found in the search were systematically evaluated based on criteria relevant to the aims of the review. For example, in addition to reviewing quantifiable effects on substance use outcomes, we address feasibility features including patterns of use and user satisfaction. Third, all studies included in the review were systematically scored on specific features and effects of the interventions. The presentation and discussion of results is based on these scorings.

Search and Exclusion History

Previous reviews on digital alcohol and drug interventions have shown that this is a large and complex research field, often with an imprecise use of concepts and descriptions regarding target groups, methods, and outcomes. In this review, we focus on digital interventions aimed at supporting a specific target: people with SUD who are working to achieve their long-term goals regarding substance use and the achievement of a more satisfying life through recovery.

To be sure that we did not exclude any relevant papers, we started with as wide a search strategy as possible. To this end, we adopted the strategy used by Dedert et al [26] as our starting point. These authors also conducted a comprehensive review, and their strategy is presented in detail in an appendix in their paper. Our first step was to implement the same search strategy as Dedert et al [26] and expand the search period to our current search date (November 2017). As they only searched for alcohol interventions, our next step was to conduct the same search, replacing the search term “alcohol” and its National Library of Medicine Medical Subject Headings terms with “drugs” and its Medical Subject Headings terms. The search was done in the following databases: MEDLINE, Embase, PsycInfo, and Cochrane Library. Together these searches (after excluding doublets) came up with a list of 1147 papers. The final step was to add additional papers from reference lists in the most recent of these papers. This resulted in adding 2 more papers.

We then started the process of excluding papers with no relevance to the aim of our review. This process was conducted in stages.

After a quick reading of titles and abstracts, the first author excluded 457 articles that were not about substance use/disorders and/or not about digital interventions. Based on a more thorough reading of the abstracts, the first author then excluded 265 articles found to be protocols, short intervention descriptions, editorials, notes, or comments.

Based on the title and a thorough reading of the abstracts, both authors cooperated in dividing the remaining 427 articles into 4 categories: (1) reviews on different kinds of digital interventions (n=21), (2) interventions aimed at reducing

XSL•FO
RenderX

hazardous alcohol use (or, in a few cases, cannabis or other drug use) (n=269), (3) digitized versions of standard treatments such as CBT (n=94), and (4) unique digital interventions aimed at helping or supporting persons in recovery from SUD with or without co-occurring physical or psychiatric disorders (n=43).

The 43 articles on unique digital intervention were all read in full and scored by both authors. The scoring categories are shown in Multimedia Appendix 1 [45-86] and 2 [45-86]. Due to the wide range of retention, feasibility, and efficacy variables employed across the studies, it was not possible to score studies using the same set of variables. Therefore, outcomes were described in the appendices as reported in the articles reviewed. The lack of consistency across studies is most apparent in the data presented in the Retention/Feasibility column in the appendices. In the Effects column, findings from the main outcome variables of the studies are presented. Studies were categorized as positive when there was a statistically positive effect favoring the experimental intervention on at least one of the primary outcomes with no significant findings in the other direction on any other primary outcomes. Agreement between the 2 authors was very high with the few disagreements easily resolved through discussion. It is these 43 papers that form the basis for this review.

Results

Overview

The tables in Multimedia Appendix 1 present our review of 43 articles on 28 unique digital interventions meant to support people in recovery from SUD. Articles presenting different studies of the same unique intervention are grouped together for ease of interpretation of the full set of findings pertaining to each intervention.

The interventions are varied when it comes to methodological strategies and technological platforms and solutions. One of the primary distinctions pertains to their degree of complexity. We have chosen to divide the interventions into 2 categories based on complexity. Eighteen interventions are defined as simple in the sense that they consist only of 1 or 2 elements. They may contain only text messaging or only online counseling, or they may contain some sort of self-monitoring and brief feedback (text messaging or online counseling). Ten interventions are defined as complex in the sense that they consist of more than 2 elements. These digital support programs typically contain several functions.

In Multimedia Appendix 1, information is provided on the 3 other criteria we used to describe and categorize studies: age of participants, gender, and substance used. With regard to age, 65% (28/43) of articles focused on adults, 23% (10/43) on adolescents, and 12% (5/43) on both age groups. With regard to gender, 91% (39/43) studies included both men and women, 2% (1/43) included men only, and 7% (3/43) were unclear on the gender of participants. With regard to substance(s) used, 40% (17/43) studies focused on alcohol, 33% (14/43) on mixed substances, 9% (4/43) on alcohol and cannabis, 5% (2/43) on opioids, 7% (3/43) on stimulants, and 2% (1/43) on cannabis.

One article [45] reports on a survey to assess patient preferences on content in text messaging interventions but addresses no specific intervention. Another article [46] describes results from a survey about the general acceptance of different kinds of digital aftercare interventions among inpatients. These articles are therefore not included in the review of outcome effects.

Publication Year

The rapid increase in development and research on the kind of digital interventions of interest to this review is clearly demonstrated when one looks at the publication year of the included articles. Even with the search spanning the 17 years, from 2000 to 2016, only 2 of the included articles were published before 2010, and 32 of the 43 included articles were published in the last 3 years, 2014 to 2017.

Country

About half of the studied interventions are from the United States (16/28, 57%), 8 are from Europe (3 from Germany, 3 from Switzerland, 1 from Ireland, and 1 from Norway), 3 are from Australia, 1 is from Canada, and 1 is from Brazil.

Types of Interventions

Only 6 of the 18 simple interventions were presented with a brand name. The most common element in these interventions was a 1-way or interactive text message service. One-way solutions typically consisted of a series of text messages with informative or supportive content delivered each day or less often for a fixed period of time. Interactive solutions typically contained standardized self-assessments of substance use, life situations, relapse risk factors, or medication compliance that were delivered as text messages and triggered automated responses or text messages from a counselor. Some simple interventions contained an online counseling service or online counseling in addition to monitoring.

Only 5 of the 18 simple interventions were meant to be stand-alone interventions (ie, without any other contact with the professional support system). All others were integrated in a larger support system that most often offered other kinds of face-to-face counseling or support services.

A clear majority of the simple interventions (12/18, 67%) were delivered on a mobile platform, and mobile phones are now the dominant device used in these kinds of interventions. Four interventions were delivered on a desktop or Web-based platform while 2 interventions used both platforms.

Nine of the 10 complex interventions were presented with a brand name. Six of the interventions are meant to be stand-alone interventions, while 4 are integrated with other treatment or aftercare services. Three of the interventions used a mobile platform such as a mobile phone, while 3 used a desktop or Web-based platform. Six interventions used a combination of desktop/website and mobile technologies.

The complex interventions contained a number of elements in different combinations. The most common features were systems for monitoring or check-ups, and some also included a Global Positioning System–based warning system. Other features included information and education modules; exercise modules

XSL•FO
RenderX

for better concentration and relaxation; and modules to foster more effective coping strategies for harm avoidance, relapse prevention, and dealing with stress. Most of these interventions also contain interactivity modules such as feedback on monitoring results, delivery of supportive messages, online counseling and contact with peers, and chat rooms or digital self-help groups. Several programs have some kind of panic button, making it possible to reach counselors or peers in situations of urgent need for support.

Some of the interventions, such as Addiction Comprehensive Health Enhancement Support System (A-CHESS), have a clear theoretical foundation. A-CHESS is based on self-determination theory (SDT) and the relapse prevention model developed by Marlatt et al [87]. Consistent with SDT, the intervention program is designed to meet 3 fundamental needs: developing perceived competence, relatedness, and internal motivation. Consistent with the relapse prevention model, the program is meant to address and offer support in high-risk situations where relapse vulnerability is high.

Other interventions build on established nondigital support programs or borrow elements from treatment methods. Overcoming Addictions [70] builds on a support program developed and implemented by the Smart Recovery organization. Location-Based Monitoring and Intervention System for Alcohol Use Disorders (LBMI-A) is an example of an intervention borrowing principles or elements from different treatment methods, in this case CBT. Principles and elements from CBT and MI were also the basis for development of the Snow Control [83] and Can Reduce [84] interventions and the intervention (no brand name) studied by Tait [85,88]. However, these interventions are not strictly digitized versions of existing CBT or MI protocols; rather, they incorporate some of these features or elements within their own unique frameworks.

Target Groups

Twelve of the simple interventions had adults as their target group, while 4 were intended for adolescents, and 2 did not discriminate their target group by age. None of the simple interventions discriminated their target group by gender, but it seems like 1 of the interventions only targeted men and it is uncertain if 1 other did the same. Nine of the simple interventions targeted people who had used or were using alcohol as their main substance, while 1 targeted stimulant users, 1 targeted cannabis users, 1 targeted opioid users (in maintenance treatment), 1 targeted alcohol and cannabis users, and 5 targeted those with mixed substance use patterns.

With regard to the complex interventions, 4 of 10 had adults as their target group while 2 had adolescents and 2 targeted both. Two of these interventions were not discriminating their target groups by age. None of these interventions discriminated their target group by gender, but 2 may have only reached men. Two of the interventions had people who had used or were using alcohol as their target group, while 2 were focused on cannabis users and 2 targeted both groups. Two interventions targeted stimulants users, and 1 addressed opioid users in maintenance treatment.

Types of Studies

Fourteen of the 18 simple interventions had been studied with quantitative controlled methodologies while 3 had been studied quantitatively without control groups. In addition, 3 interventions had been studied with qualitative methodologies while 1 had been studied both quantitatively with a control group and qualitatively. One paper presented results from a survey on preferred text message content without referring to a specific intervention. The number of participants in the quantitative studies varied from 54 to 408, while the qualitative studies were smaller (eg, from 16 to 80 participants). The follow-up periods varied from 1 to 12 months.

Nine complex interventions had been studied with quantitative controlled methodologies and 1 quantitatively without a control group. The number of participants in these studies varied from 50 to 84, and the follow-ups varied from 2 to 12 months in duration. In addition, 2 complex interventions had been studied with mixed methods research designs and 1 intervention (A-CHESS) has been studied with several designs: controlled, uncontrolled, and mixed methods. These studies had from 29 to 349 participants and 2 to 12 months follow-up.

The last paper in Multimedia Appendix 1 was based on a survey of 374 inpatients about the feasibility of digitized aftercare interventions without referring to a specific intervention.

Implementation

Overall, it appeared that the interventions were implemented successfully (without technical difficulties), although there was not a lot of information on this. The interventions were made available to eligible participants in the studies by forwarding links to internet sites or via mobile phone apps. The interventions in our review are generally not made accessible through an open app store download. A new commercial version of A-CHESS is, however, made available through the app stores. On the other hand, many recovery apps, not supported by research, are openly available in app stores. It is outside the scope of this review to make any evaluation of such interventions.

There was only 1 study focusing explicitly on prerequisites for a successful implementation of digital interventions. In a study on the implementation of A-CHESS, Ford [74] found that the following factors were important for a successful and sustained implementation of the intervention: strong leadership support, a staff that is passionate about the intervention, interpreting user feedback to re-engage users who had dropped out, including the intervention in meetings with staff and users, developing internal guidelines for using the intervention, and developing sustainable strategies for financing the intervention.

In 2 of the studies of complex interventions, participants were offered free phones and offered replacement phones if the first ones were lost, broken, or stolen. In the Check-In Program [79], 44% returned their first phone after the end of the 3-month study while 44% needed a replacement phone. In the first study of A-CHESS [76], 170 participants needed 116 replacement phones during the 8-month trial. In the other papers, there is no information about whether the participants were offered free phones or used their own phones.

XSL•FO
RenderX

Rates of Sustained Use of the Interventions

The information in the papers on the simple interventions suggests that rates of use of the app or intervention were fairly uneven across the studies. If recruitment to the intervention (and study follow-up) took place through open websites, there was a large drop in the number of participants from those who accessed the site to those who registered in the intervention and those who accessed the first module (eg, first assessment) [51]. The same was seen in a study where possible participants were screened by general practitioners [54]. Among those who screened eligible for the intervention, only 50% accessed the first module and only 50% of those accessed the next module. Conversely, in studies where participants were recruited from patients in SUD treatment programs and where this information is reported, the rate of sustained use of the intervention seems to have been as high as 75% at the end of the study [50].

In the complex interventions, rates of use seemed to be fairly high at the beginning, typically around 90% in the first few weeks. But rates of use of the interventions dropped very quickly; for example, from a mean of 7.3 to a mean of 1.3 log-ins each week during the first 3 months in the intervention studied by Campbell [70] or to 18% after 6 weeks in the interventions studied by Schaub [83,89]. Two interventions had higher rates of use; 1 reported a drop from 63% completing the first module to 48% completing the last module during the 3-week intervention [85] and a second reported that on average the participants completed 7 of the 8 sessions in the intervention [57]. Two of the complex interventions seem to have high overall rates of use. In the My First Year of Recovery (MyFYR) intervention [82], 78% completed the 1-year-long program and 70% of those who relapsed during the intervention remained engaged or re-engaged and were able to complete the intervention. In the case of the A-CHESS intervention, 78% of the sample was still using the intervention after 4 months [71]. The intervention Check-In Program was combined with a computer-based psychoeducative program (Therapeutic Education System) in one of the study conditions [79]. In the combined condition, the retention rate was 84% by the end of the study, compared to 56% in the uncombined condition (P=.031).

Intensity and Duration of the Interventions

The simple interventions varied quite a bit in intensity and duration. The most intense intervention was studied by Reback et al [65]. In this intervention, the participants received on average about 10 messages each day and sent as many replies over the 2-month duration. The participants received additional support feedback in response to about a third of their messages. The shortest intervention was 2 weeks in duration [64]. The intensity was also very high in this intervention; on average the participants received 8 and sent 4 messages each day. The rest of the interventions were less intense and had a longer duration. In these, participants typically received and sent 2 messages each day for 3 months [48] or 10 months [63], 1 message each day for 2 to 4 months [52,55,58,67,68], or 1 each week for 6 months [62]. There is not much information in the papers documenting whether the participants actually read the messages they received. There are 2 papers reporting on this; in Agyapong [49], participants read 67% of the received messages, and in Haug [90], participants responded to 88% of the messages.

The complex interventions were generally more intense than most of the simple interventions. In most interventions, it was possible for participants to log in to several elements or modules each day, making the interventions more or less intense based on how many modules the participants accessed each day. In studies of the complex interventions, the duration of the interventions varied between 2 and 12 months.

Intervention Content and Use of Features

The text messages in the simple interventions covered a large range of topics. One article [49] reported on which topics were of greatest interest to the participants. Among those were messages on motivation for recovery and relapse prevention and reminders on why and how to stay abstinent. The same kinds of topics were recommended by participants in Gonzales [78] and in the survey by Tofighi [45].

Use of different kinds of modules varied between interventions and during the intervention in the 3 complex interventions for which information on this was provided. None of the participants in the LBMI-A intervention [78] used the skill modules for resisting urges to drink or drink refusals and very few used the psychoeducational modules after week 2 of the intervention, while most of the participants continued using the monitoring modules. The 25 participants in the experimental condition of the study of the Check-In Program [79] completed, on average, 21 self-management modules and 9 functional analysis modules during the 3-month intervention. Several of the studies on A-CHESS present information on the use of the different modules of the intervention over time. The study by Dennis [73] showed that adolescent participants using the intervention completed 89% of the assessments in the ecological momentary assessment module and accessed the ecological momentary intervention module 78% of the days of intervention. The most used ecological momentary interventions were recovery support, motivation, relaxation, and social networking. In Gustafson [76], it was reported that the participants, on average, used the intervention 41 days during the 8-month trial and that 72% of the participant pressed the panic button at least once. McTavish [77] presents the use of different modules of the intervention during the first 4 months of the first trial and relates it to the theoretical principles of the intervention (SDT). McTavish [77] found that the percentage of participants using the intervention dropped from 94% the first week to 78% the fourth month. Use of modules related to perceived competence dropped from 80% to 39%, modules related to autonomous motivation from 84% to 66%, and modules aimed at increasing the feeling of relatedness from 91% to 76%.

User Satisfaction

The articles reported high satisfaction with the simple interventions. For example, some studies reported that the participants were generally highly satisfied [49,56] or that they "felt connected" via the intervention [55]. Bradford [56], Gagnon [58], and Ingersoll [63] reported that 80% to 90% of participants were satisfied, finding the interventions easy to use and being confident and comfortable in using them, while

XSL•FO
RenderX

Gonzales [60] found that 70% were positive about the intervention (20% were ambivalent and 10% negative). In the intervention studied by Haug [90], the overall satisfaction was a bit lower: 63% of participants found the intervention generally helpful, but 75% wanted to do the program again, which suggests a somewhat higher level of satisfaction.

Three of the papers reported on the participants' expressed satisfaction and the usefulness of the complex interventions. Guarino [79] reported that participants found the Check-In Program intervention highly acceptable and useful (75 to 80 points on a 100-point scale). Hasin [80] reported that 80% or more of the participants in the Health Call intervention gave very positive feedback on user interface and satisfaction with the content. Campbell [91] reported that the participants in the Overcoming Addictions intervention found the social support and awareness reminders to be the most helpful element.

Intervention Effects on Substance Use Outcomes

Of the 24 studies of 18 simple interventions included in this review, 7 featured a control condition and produced positive effects on substance use outcomes. These were a stepped care intervention that included computerized feedback [92], My Assessment by Bradford et al [56], 2 HIV risk reduction interventions [58,65], ESQYIR by Gonzales et al [61,93], an in-home messaging device by Santa Anna et al [67], and an internet-based relapse prevention program [69]. Three of these interventions were delivered by text messaging [59,61,65]. Other positive interventions consisted of an integrated psychosocial assessment delivered through an app [56], a website that provided tailored audiovisual messages regarding safer injection practices [58], and 20 online lessons that provided information on addiction and relapse prevention skills for adolescents [69].

Conversely, 8 studies of simple interventions with control conditions found no positive effects on substance use outcomes. The interventions tested in these studies were a text messaging system [47,48], integrated online counseling intervention [50], computerized intervention for anger management [57], integrated text messaging systems that included online counseling [62-64], and a text messaging intervention that included medication monitoring and support [94]. Another 9 studies either did not include a control condition or did not examine substance use outcomes. By a simple box score calculation, these results indicate that 7 of 15 studies with control conditions (47%) produced positive effects on substance use outcomes. When considered at the level of the interventions, 7 produced positive results in at least 1 study, whereas 7 produced negative results in 1 or more studies with no positive results in other studies (ie, 50% of interventions positive).

The interventions in the 7 positive studies all had moderate effect size advantages over the control conditions on 1 of the primary outcomes. The control conditions were bona fide interventions, usually treatment as usual without the digital component, except in the Bischof study [92], which employed an untreated control condition, and Trudeau et al [69], which used a wait list control. Five of the interventions addressed drugs, and 2 focused on alcohol. Five interventions were integrated, while 2 were stand-alone. The studies with interventions that did not produce positive effects over control groups look similar to those that did on strength of the control groups, targeted substance, and stand-alone versus integrated format. All of the negative studies included bona fide active control conditions, primarily behavioral treatment as usual. Three studies focused on alcohol only, while 4 addressed mixed or poly substance use. Finally, most of the interventions were integrated, with 2 stand-alone. It should be noted that 2 studies were likely underpowered [50,64], as they produced positive but nonsignificant effects on primary SUD outcomes.

The 10 complex interventions were evaluated in a total of 19 publications included in the review. The A-CHESS intervention was studied in 7 reports; only 2 other interventions were examined in more than 1 report [85,88]. The Hasin et al [80] and Aharonovich et al [81] publications were of the same intervention in 2 separate studies, whereas the 2 Tait et al [85,88] publications reported results from different follow-ups in the same study. Of the studies included, 9 yielded positive results, 5 produced negative results, 2 did not include control conditions, and 3 did not examine SUD outcomes. Four of the 9 positive studies were of the A-CHESS system. In the one large scale A-CHESS randomized controlled trial, those randomized to A-CHESS reported fewer heavy drinking days over a 12-month follow-up than those who did not receive A-CHESS: 1.39 versus 2.75 out of the prior 30 days [76]. A second study showed that ecological momentary assessment data gathered on A-CHESS could predict upcoming relapse episodes [71]. A third study indicated that adolescents who accessed 2 or more supportive functions on A-CHESS within 1 hour after reporting elevated relapse risk were less likely to go on to relapse in the next 7 days that those who used fewer supportive A-CHESS functions [73]. It should be noted that the positive results in this paper could have simply reflected self-selection, with more motivated participants both accessing A-CHESS more frequently and having better outcomes. Finally, a fourth study found that the effects of A-CHESS on the risky drinking days outcome was mediated by participation in outpatient SUD treatment [75].

Other complex interventions that generated SUD outcomes superior to comparison conditions were a mobile phone–delivered CBT-like intervention that consisted of 7 modules [78], a mobile phone–based monitoring program (HealthCall) that graphs results and arranges for contact with a counselor [80], a mobile phone–based treatment extender compatible with the computerized Therapeutic Education System [79], a Web-based self-help intervention that included chat counseling [95], and a Web-based intervention that included self-monitoring and weekly feedback from counselors [86].

Complex interventions that did not produce positive effects on SUD outcomes were a Web application based on Smart Recovery [70], a study of A-CHESS where there was no difference in A-CHESS use between lapsers and nonlapsers [71], an initial pilot study of HealthCall [80], an 8-module Web-delivered self-help intervention based on CBT and MI [96], and a Web intervention based on CBT, MI, and harm avoidance approaches [85,88]. According to a simple box score calculation, these results indicate that 9 of 14 controlled studies (64%) produced positive effects on substance use outcomes. When considered at the level of the interventions, 6 produced

XSL•FO
RenderX

positive results in at least 1 study, whereas 3 produced negative results in 1 or more studies with no positive results in other studies (67% of interventions positive). There was 1 negative A-CHESS study but 4 positive ones, and 1 negative study of Healthcall but 1 positive one.

In the studies with positive effects, 6 featured integrated interventions and 3 stand-alone interventions. Four of the studies focused on alcohol, 2 on cannabis, 1 on both alcohol and cannabis, and 2 on mixed substance use. Four of the studies featured no treatment or waitlist control conditions, 1 included another online intervention, 1 included an active behavioral intervention control condition, and 1 had a standard methadone maintenance control condition (2 studies did not include a treatment control condition but rather focused on the ability of the A-CHESS system to predict relapse and deliver just-in-time interventions). Effect sizes were generally in the moderate range, although the major A-CHESS trial [76] produced a smaller effect (*d*=.18-.25) and the Tossman et al [86] study produced a large effect (*d*=.75). In the studies that did not produce positive treatment effects, 2 interventions were integrated and 3 were stand-alone. Two studies focused on alcohol, 1 on cannabis, 1 on alcohol and cannabis, and 1 on stimulants. The control conditions were generally fairly weak, including a waitlist control, psychoeducation, historical interactive voice response intervention group, and Smart Recovery.

Intervention Effects on Other Outcomes

There was not much information in the papers on effects on outcomes other than substance use. Four of the papers on simple interventions reported on changes in use of other services. Gonzales [61] reported significantly higher attendance at self-help meetings and recovery-oriented activities among those in the experimental condition compared to controls, and Ingersoll [63] reported that adherence to antiretroviral treatment increased by 19 percentage points in the experimental condition compared to a 9 percentage point increase in the treatment as usual condition. Bischof [92] reported a drop in face-to-face counseling time of 50% in the experimental condition, while Lucht [64] reported that the participants in the experimental condition spent significantly more days than the controls in psychiatric hospital. While the results in the first 3 of these articles may be evaluated as positive results, we are not sure that the result reported by Lucht could be evaluated in this way. Only 1 of the articles on the complex interventions reported on changes in use of other services. Tait [97] reported that those in the experimental condition significantly increased their general help seeking compared to controls.

Three of the articles on simple interventions also reported on other outcomes. Cougle et al [57] reported that hostile interpretation training led to greater improvements in interpretation bias, trait anger, and anger expression. Reback et al [65] reported that the participants in the experimental condition significantly reduced their risky sexual behavior, while Rooke [66] reported a significant reduction in depression in those in the experimental condition compared to controls. Four of the articles on complex interventions also reported some information on effects on outcomes other than substance use. Gustafson [76] found that A-CHESS had no impact on negative consequences of drinking, Glass et al [75] reported that A-CHESS increased participation in outpatient treatment following rehabilitation, Schaub [95] found that Can Reduce did not affect mental health measures, and Tait [97] found that the intervention they studied made no difference on psychological distress. On the other hand, Tait found that the intervention led to a significant reduction in days of general impairment.

Discussion

Principal Findings

The development and evaluation of digital interventions aimed at preventing or treating substance use–related problems and disorders is a rapidly growing field. A large number of articles were identified on this topic, and most of these reports focus on interventions developed and studied in the last few years. The concept of digital interventions includes such a wide variety of interventions with regard to aims, target groups, methods, and technological solutions that it is impossible to cover them all in one review. In this review, we therefore focused more narrowly on unique interventions aimed at supporting people in recovery from SUD. But as we did not want to miss any studies due to imprecise use of concepts in relevant studies, we started out with a wide collection of search terms. We found a large number of papers on interventions for hazardous or risky drinking and digital interventions in the SUD field that were not relevant to the aim of this review. However, including them in the first stage of the evaluation process gave us the opportunity to suggest a categorization of interventions, created specifically for digital platforms, that may become useful in further development, research, and review of such interventions.

Although not the focus of this review, it is clear that the field of digital interventions aimed at hazardous but not disorder-level alcohol or drug use is a large and well-reviewed area but with relatively modest results. Digitizing existing treatment such as CBT has also become an important and promising area, but to our knowledge without any systematic review done so far. Our quick reading of these studies gave us an impression that these kinds of interventions could make an important contribution to the development of more available and effective treatment of SUD.

The 43 articles that reported on the studies that were evaluated as relevant to the aim of this review seemed to cover a larger variety of interventions then the 235 and 87 articles in the other categories. This made it important to define scoring criteria so as to conduct a review as systematically as possible. Two of these criteria—integrated versus stand-alone in relation to other services and substances of abuse—did not appear to have a significant impact on the feasibility or efficacy of the interventions. There was insufficient variation in 2 of our other criterion—gender and age of the participants—to draw any conclusions regarding their impact on outcomes. Finally, the categorization of the interventions that was based on their complexity (number of elements or modules) did appear to make a difference on both feasibility and effects on substance use. However, the more positive substance use outcomes for the complex interventions may have been due to differences in the

XSL•FO
RenderX

strength of the control conditions used in these studies, which was another criterion examined in the review.

Feasibility Strengths and Weaknesses

International market figures show that smartphones have become the dominant digital device in the Western world, with an 80% to 90% share of the total mobile phone market, and are quickly also becoming the dominant device in the rest of the world, passing 50% of the market [98]. Mobile phones are also quickly replacing other digital devices, such as laptops and tablets, as the device most likely to be used daily. It is therefore understandable that mobile phones are becoming the dominant technological platform for digital interventions, making it even more appealing as a tool for offering effective and flexible solutions for recovery support. The simple interventions seemed relatively easy to develop and implement, especially those using standard text messaging and mobile phone apps. Complex interventions seemed also relatively easy to implement, but the study by Ford [74] showed that there are many organizational prerequisites to achieving a sustainable implementation of such interventions over time. In addition to addressing questions about the feasibility of various technical solutions, it is important to determine whether the organizational prerequisites are in place to sustain the implementation over time before implementing a digital intervention.

Another challenge in the implementation of digital interventions is, of course, that the participants need the required technical equipment. The initial cost of buying such equipment may be too high for many people with SUD, and mobile phones may easily be broken, lost, or stolen. Buying and replacing phones for participants may be a solution in a research project but appears not to be a sustainable solution in real-life contexts. Our impression is that owning a mobile phone, which increasingly means a smartphone, is regarded by people in recovery from SUD as highly desired and even viewed as an essential expense. Smartphones are not only replacing ordinary mobile phones but also personal computers. This means that interventions have to be built on mobile platforms, using the flexibility and technological possibilities of modern smartphones. But developers of digital interventions also need to adjust their methodological and design strategies to the particularities of modern mobile phone technology and user interface designs [99]. Interventions based on traditional desktop, website, or mobile phone technologies may already be out of date, as they may not contain the functionalities and user interface designs required to reach potential intervention participants and keep them engaged with the program.

Rates of sustained use of the apps and interventions varied to a considerable amount. Access to simple interventions via websites or mobile phone app stores made it possible to reach many potential participants, but retention in most interventions dropped quickly. This is a method often used in interventions aiming at hazardous but not disorder-level substance use. However, if the goal is to offer recovery support, it appears to be more effective to recruit people in treatment or self-help group/network settings to both reach the most relevant participants and keep rates of intervention use high over time.

In the complex interventions, retention was generally high both in the beginning and over time. Here it seems that the challenge was more that the frequency of use was very different for different kinds of elements or modules. Generally, the information and educational modules appeared to be most frequently used in the beginning of the intervention, while modules supporting continuous monitoring and communication with counselors and peers retained higher rates of use over time. Instead of evaluating this as a weakness of the intervention program, these findings could lead to the development of interventions in which different kinds of modules are presented to the participants in a planned “tunneled” sequence [100]. Or it might be advantageous to divide some interventions into separate modules, making it possible to directly access each module.

Effects

In formulating the aim of this review, we were interested in how digital interventions could support people in recovery from SUD with regard to their goals for achieving abstinence or reduced substance use as well as better health and life situations and use of other services. The review showed, however, that few studies reported on anything other than changes in substance use.

Across simple and complex interventions, slightly more than half (55%) of the studies with control conditions generated positive findings on 1 or more substance use outcomes, with 57% of the interventions also found to be efficacious in 1 or more studies. In the positive studies, effects were typically in the small to moderate range, with a few studies yielding larger effects. At first glance, the simple interventions appeared to be somewhat less effective than the complex ones. In studies of simple interventions that employed control groups, 47% yielded positive findings on 1 or more of the primary substance use outcomes, with 50% of the interventions producing positive results in at least 1 study. Studies testing complex interventions, on the other hand, generated positive effects in 64% of studies with control groups, with 67% of the interventions producing positive effects in 1 or more studies. However, studies of simple interventions were more likely to include stronger control conditions than studies of complex interventions. As was noted earlier, this might explain why simple interventions were less likely to produce positive effects. In addition, 2 studies of simple interventions appeared to be underpowered and might have shown positive effects with larger samples. It did not appear that the substance targeted in the study or whether the intervention featured a stand-alone versus integrated format accounted for the results.

Overall, these results do not provide consistent, strong support for the efficacy of these interventions. However, the heterogeneity in results, with some interventions appearing to be more promising than others, indicates that more work is needed to better understand the characteristics of efficacious digital recovery support interventions. Further research should also shed light on the kinds of individuals most likely to benefit from different digital interventions and at what points in their recoveries the largest effects are obtained.

XSL•FO
RenderX

In the scoring of the studies, we systematically searched for other outcomes than changes in substance use such as psychological health, medication, housing, employment, social functioning, and criminality. But as the articles in this review contained little information on such effects, it is not possible to draw strong conclusions on this issue. We have, however, the impression that the interventions had no or only modest effects on such outcomes. As recovery should focus on many more issues than just changes in substance use, it is a weakness in the studies that they did not focus more on such outcomes and a weakness in the interventions that they either are not aiming at contributing to such changes or that they have no effect when they try to do so.

Conclusions

The digital interventions included in this review are in general feasible but are not consistently effective in helping people in recovery from SUD reduce their substance use. It is questionable whether they are effective in supporting people to achieve other recovery goals, given the relative lack of information on this in the studies. Mobile phones appear to be the most feasible technological platform for such interventions. Single interventions, such as 1-way or interactive text messaging or text messaging in combination with a simple monitoring module are relatively easy to develop, implement, and sustain and can be an effective supplement in continuing care and support programs. Complex interventions appear to be feasible and some of them are also modestly effective. They require, however, more technological and organizational resources to develop, implement, and sustain. It also appears that they could benefit from being developed into more sequentially and individually tunneled programs or being divided into single, directly accessible interventions.

Participants' general satisfaction with the studied interventions should be regarded as the best inspiration to develop even more feasible and effective digital interventions, using all the technological possibilities and appealing user interface designs of modern mobile technologies. However, these technological solutions are only relevant if they are adjusted to the life situation of potential users and the organizational and knowledge-based framework of the support systems they are meant to be a part of and they make a difference in helping participants reach their recovery goals.

Acknowledgments

We thank Kari Hølland for her invaluable help in the search process and retrieval of relevant papers.

Conflicts of Interest

JRM is currently collaborating with the developers of the A-CHESS intervention in an National Institute on Alcohol Abuse and Alcoholism–funded grant. He also receives salary support from Caron Treatment Centers, the developer of My First Year in Recovery. SN reports no conflicts.

Multimedia Appendix 1

Intervention and study characteristics.

[PDF File (Adobe PDF File), 56KB - jmir_v20i8e255_app1.pdf]

Multimedia Appendix 2

Intervention retention, feasibility, and effects.

[PDF File (Adobe PDF File), 63KB - jmir_v20i8e255_app2.pdf]

References

1. McKay JR, Franklin TR, Patapis N, Lynch KG. Conceptual, methodological, and analytical issues in the study of relapse. Clin Psychol Rev 2006 Mar;26(2):109-127. [doi: 10.1016/j.cpr.2005.11.002] [Medline: 16371242]
2. Kalivas PW. Neurobiology of cocaine addiction: implications for new pharmacotherapy. Am J Addict 2007;16(2):71-78. [doi: 10.1080/10550490601184142] [Medline: 17453607]
3. Koob GF, Volkow ND. Neurocircuitry of addiction. Neuropsychopharmacology 2010 Jan;35(1):217-238 [FREE Full text] [doi: 10.1038/npp.2009.110] [Medline: 19710631]
4. Kreek MJ, Nielsen DA, Butelman ER, LaForge KS. Genetic influences on impulsivity, risk taking, stress responsivity and vulnerability to drug abuse and addiction. Nat Neurosci 2005 Nov;8(11):1450-1457. [doi: 10.1038/nn1583] [Medline: 16251987]
5. Miller WR, Westerberg VS, Harris RJ, Tonigan JS. What predicts relapse? Prospective testing of antecedent models. Addiction 1996 Dec;91 Suppl:S155-S172. [Medline: 8997790]
6. Weisner C, Delucchi K, Matzger H, Schmidt L. The role of community services and informal support on five-year drinking trajectories of alcohol dependent and problem drinkers. J Stud Alcohol 2003 Nov;64(6):862-873. [Medline: 14743951]
7. Moos RH, Moos BS. Treated and untreated alcohol-use disorders: course and predictors of remission and relapse. Eval Rev 2007 Dec;31(6):564-584. [doi: 10.1177/0193841X07306749] [Medline: 17986708]

XSL•FO
RenderX

8. Orford J, Hodgson R, Copello A, John B, Smith M, Black R, UKATT Research Team. The clients' perspective on change during treatment for an alcohol problem: qualitative analysis of follow-up interviews in the UK Alcohol Treatment Trial. Addiction 2006 Jan;101(1):60-68. [doi: 10.1111/j.1360-0443.2005.01291.x] [Medline: 16393192]
9. McKay JR. Treating Substance Use Disorders With Adaptive Continuing Care. Washington: American Psychological Association; 2009.
10. Dennis M, Scott CK, Funk R. An experimental evaluation of recovery management checkups (RMC) for people with chronic substance use disorders. Eval Program Plann 2003 Aug;26(3):339-352 [FREE Full text] [doi: 10.1016/S0149-7189(03)00037-5] [Medline: 30034059]
11. McLellan AT, Lewis DC, O'Brien CP, Kleber HD. Drug dependence, a chronic medical illness: implications for treatment, insurance, and outcomes evaluation. JAMA 2000 Oct 04;284(13):1689-1695. [Medline: 11015800]
12. Dennis M, Scott CK. Managing addiction as a chronic condition. Addict Sci Clin Pract 2007 Dec;4(1):45-55 [FREE Full text] [Medline: 18292710]
13. McLellan AT, McKay JR, Forman R, Cacciola J, Kemp J. Reconsidering the evaluation of addiction treatment: from retrospective follow-up to concurrent recovery monitoring. Addiction 2005 Apr;100(4):447-458. [doi: 10.1111/j.1360-0443.2005.01012.x] [Medline: 15784059]
14. Wagner EH, Austin BT, Davis C, Hindmarsh M, Schaefer J, Bonomi A. Improving chronic illness care: translating evidence into action. Health Aff (Millwood) 2001;20(6):64-78. [doi: 10.1377/hlthaff.20.6.64] [Medline: 11816692]
15. Humphreys K. Circles of Recovery: Self-Help Organizations for Addictions. Cambridge: Cambridge University Press; 2004.
16. McKay JR, Merikle E, Mulvaney FD, Weiss RV, Koppenhaver JM. Factors accounting for cocaine use two years following initiation of continuing care. Addiction 2001 Feb;96(2):213-225. [doi: 10.1080/09652140020020946] [Medline: 11182866]
17. Silverman K, Holtyn AF, Morrison R. The therapeutic utility of employment in treating drug addiction: science to application. Transl Issues Psychol Sci 2016 Jun;2(2):203-212 [FREE Full text] [doi: 10.1037/tps0000061] [Medline: 27777966]
18. Milby JB, Schumacher JE, McNamara C, Wallace D, Usdan S, McGill T, et al. Initiating abstinence in cocaine abusing dually diagnosed homeless persons. Drug Alcohol Depend 2000 Jul 01;60(1):55-67. [Medline: 10821990]
19. Milby JB, Schumacher J, Wallace D, Frison S, McNamara C, Usdan S, et al. Day treatment with contingency management for cocaine abuse in homeless persons: 12-month follow-up. J Consult Clin Psychol 2003 Jun;71(3):619-621. [Medline: 12795584]
20. Milby JB, Schumacher JE, Wallace D, Freedman MJ, Vuchinich RE. To house or not to house: the effects of providing housing to homeless substance abusers in treatment. Am J Public Health 2005 Jul;95(7):1259-1265. [doi: 10.2105/AJPH.2004.039743] [Medline: 15983278]
21. Scott CK, Dennis ML. Results from two randomized clinical trials evaluating the impact of quarterly recovery management checkups with adult chronic substance users. Addiction 2009 Jun;104(6):959-971 [FREE Full text] [doi: 10.1111/j.1360-0443.2009.02525.x] [Medline: 19344441]
22. Morgenstern J, Hogue A, Dauber S, Dasaro C, McKay JR. A practical clinical trial of coordinated care management to treat substance use disorders among public assistance beneficiaries. J Consult Clin Psychol 2009 Apr;77(2):257-269 [FREE Full text] [doi: 10.1037/a0014489] [Medline: 19309185]
23. Morgenstern J, Hogue A, Dauber S, Dasaro C, McKay JR. Does coordinated care management improve employment for substance-using welfare recipients? J Stud Alcohol Drugs 2009 Nov;70(6):955-963 [FREE Full text] [Medline: 19895774]
24. Blodgett JC, Maisel NC, Fuh IL, Wilbourne PL, Finney JW. How effective is continuing care for substance use disorders? A meta-analytic review. J Subst Abuse Treat 2014 Feb;46(2):87-97 [FREE Full text] [doi: 10.1016/j.jsat.2013.08.022] [Medline: 24075796]
25. Tofighi B, Nicholson J, McNeely J, Muench F, Lee J. Mobile phone messaging for illicit drug and alcohol dependence: A systematic review of the literature. Drug Alcohol Rev 2017 Dec;36(4):477-491. [doi: 10.1111/dar.12535] [Medline: 28474374]
26. Dedert EA, McDuffie J, Stein R, McNiel J, Kosinski A, Freiermuth C, et al. Electronic interventions for alcohol misuse and alcohol use disorders: a systematic review. Ann Intern Med 2015 Aug 04;163(3):205-214 [FREE Full text] [doi: 10.7326/M15-0285] [Medline: 26237752]
27. Tait R, Spijkerman R, Riper H. Internet and computer based interventions for cannabis use: a meta-analysis. Drug Alcohol Depend 2013 Dec 01;133(2):295-304. [doi: 10.1016/j.drugalcdep.2013.05.012] [Medline: 23747236]
28. Kaner E, Beyer F, Garnett C, Crane D, Brown J, Muirhead C, et al. Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations. Cochrane Database Syst Rev 2017 Dec 25;9:CD011479. [doi: 10.1002/14651858.CD011479.pub2] [Medline: 28944453]
29. Cadigan J, Haeny A, Martens M, Weaver C, Takamatsu S, Arterberry B. Personalized drinking feedback: a meta-analysis of in-person versus computer-delivered interventions. J Consult Clin Psychol 2015 Apr;83(2):430-437 [FREE Full text] [doi: 10.1037/a0038394] [Medline: 25486373]
30. Dotson K, Dunn M, Bowers C. Stand-alone personalized normative feedback for college student drinkers: a meta-analytic review, 2004 to 2014. PLoS One 2015;10(10):e0139518 [FREE Full text] [doi: 10.1371/journal.pone.0139518] [Medline: 26447792]

XSL•FO
RenderX

31. Leeman R, Perez E, Nogueira C, DeMartini KS. Very-brief, web-based interventions for reducing alcohol use and related problems among college students: a review. Front Psychiatry 2015;6:129 [FREE Full text] [doi: 10.3389/fpsyt.2015.00129] [Medline: 26441690]
32. Huh D, Mun E, Larimer M, White H, Ray A, Rhew I, et al. Brief motivational interventions for college student drinking may not be as powerful as we think: an individual participant-level data meta-analysis. Alcohol Clin Exp Res 2015 May;39(5):919-931 [FREE Full text] [doi: 10.1111/acer.12714] [Medline: 25872599]
33. Berman A, Gajecki M, Sinadinovic K, Andersson C. Mobile interventions targeting risky Drinking Among University Students: A Review. Curr Addict Rep 2016;3:166-174 [FREE Full text] [doi: 10.1007/s40429-016-0099-6] [Medline: 27226948]
34. Cunningham J, Shorter G, Murphy M, Kushnir V, Rehm J, Hendershot C. Randomized controlled trial of a brief versus extended internet intervention for problem drinkers. Int J Behav Med 2017 Dec;24(5):760-767 [FREE Full text] [doi: 10.1007/s12529-016-9604-5] [Medline: 27770293]
35. Boyle SC, Earle AM, LaBrie JW, Smith DJ. PNF 2.0? Initial evidence that gamification can increase the efficacy of brief, web-based personalized normative feedback alcohol interventions. Addict Behav 2017 Apr;67:8-17 [FREE Full text] [doi: 10.1016/j.addbeh.2016.11.024] [Medline: 27978426]
36. Braitman AL, Henson JM. Personalized boosters for a computerized intervention targeting college drinking: the influence of protective behavioral strategies. J Am Coll Health 2016 Dec;64(7):509-519 [FREE Full text] [doi: 10.1080/07448481.2016.1185725] [Medline: 27148633]
37. Wurdak M. Effectiveness of a brief intervention with and without booster session for adolescents hospitalized due to alcohol intoxication. J Substance Use 2016;21(1):72-77. [doi: 10.3109/14659891.2014.951415]
38. Gajecki M, Andersson C, Rosendahl I, Sinadinovic K, Fredriksson M, Berman A. Skills training via smartphone app for university students with excessive alcohol consumption: a randomized controlled trial. Int J Behav Med 2017 Dec;24(5):778-788 [FREE Full text] [doi: 10.1007/s12529-016-9629-9] [Medline: 28224445]
39. Haug S, Paz CR, Kowatsch T, Filler A, Dey M, Schaub MP. Efficacy of a web- and text messaging-based intervention to reduce problem drinking in adolescents: results of a cluster-randomized controlled trial. J Consult Clin Psychol 2017 Dec;85(2):147-159. [doi: 10.1037/ccp0000138] [Medline: 27606700]
40. Ridout B, Campbell A. Using Facebook to deliver a social norm intervention to reduce problem drinking at university. Drug Alcohol Rev 2014 Nov;33(6):667-673. [doi: 10.1111/dar.12141] [Medline: 24689339]
41. Carroll K, Kiluk B, Nich C, Gordon M, Portnoy G, Marino D, et al. Computer-assisted delivery of cognitive-behavioral therapy: efficacy and durability of CBT4CBT among cocaine-dependent individuals maintained on methadone. Am J Psychiatry 2014 Apr;171(4):436-444 [FREE Full text] [doi: 10.1176/appi.ajp.2013.13070987] [Medline: 24577287]
42. Aharonovich E, Greenstein E, O'Leary A, Johnston B, Seol SG, Hasin DS. HealthCall: technology-based extension of motivational interviewing to reduce non-injection drug use in HIV primary care patients—a pilot study. AIDS Care 2012;24(12):1461-1469 [FREE Full text] [doi: 10.1080/09540121.2012.663882] [Medline: 22428809]
43. McGeary JE, Meadows SP, Amir N, Gibb BE. Computer-delivered, home-based, attentional retraining reduces drinking behavior in heavy drinkers. Psychol Addict Behav 2014 Jun;28(2):559-562 [FREE Full text] [doi: 10.1037/a0036086] [Medline: 24955674]
44. Andreu C, Cosmelli D, Slagter H, Franken I. Effects of a brief mindfulness-meditation intervention on neural measures of response inhibition in cigarette smokers. PLoS One 2018;13(1):e0191661 [FREE Full text] [doi: 10.1371/journal.pone.0191661] [Medline: 29370256]
45. Tofighi B, Grossman E, Bereket S. Text message content preferences to improve buprenorphine maintenance treatment in primary care. J Addict Dis 2016;35(2):92-100. [doi: 10.1080/10550887.2015.1127716] [Medline: 26670868]
46. Hennemann S, Beutel M, Zwerenz R. Drivers and barriers to acceptance of web-based aftercare of patients in inpatient routine care: a cross-sectional survey. J Med Internet Res 2016 Dec 23;18(12):e337 [FREE Full text] [doi: 10.2196/jmir.6003] [Medline: 28011445]
47. Agyapong VIO, Ahern S, McLoughlin DM, Farren CK. Supportive text messaging for depression and comorbid alcohol use disorder: single-blind randomised trial. J Affect Disord 2012 Dec 10;141(2-3):168-176. [doi: 10.1016/j.jad.2012.02.040] [Medline: 22464008]
48. Agyapong VIO, McLoughlin DM, Farren CK. Six-months outcomes of a randomised trial of supportive text messaging for depression and comorbid alcohol use disorder. J Affect Disord 2013 Oct;151(1):100-104. [doi: 10.1016/j.jad.2013.05.058] [Medline: 23800443]
49. Agyapong VIO, Milnes J, McLoughlin DM, Farren CK. Perception of patients with alcohol use disorder and comorbid depression about the usefulness of supportive text messages. Technol Health Care 2013;21(1):31-39. [doi: 10.3233/THC-120707] [Medline: 23358057]
50. Alemi F, Haack M, Nemes S, Harge A, Baghi H. Impact of online counseling on drug use: a pilot study. Qual Manag Health Care 2010;19(1):62-69. [doi: 10.1097/QMH.0b013e3181d1394c] [Medline: 20042934]
51. Andrade A, de Lacerda Roseli B, Gomide H, Ronzani T, Sartes L, Martins L, WHO e-health project on Alcohol and Health Investigators Group. Web-based self-help intervention reduces alcohol consumption in both heavy-drinking and dependent alcohol users: A pilot study. Addict Behav 2016 Dec;63:63-71. [doi: 10.1016/j.addbeh.2016.06.027] [Medline: 27424165]

XSL•FO
RenderX

52. Aschbrenner KA, Naslund JA, Gill LE, Bartels SJ, Ben-Zeev D. A qualitative study of client-clinician text exchanges in a mobile health intervention for individuals with psychotic disorders and substance use. J Dual Diagn 2016;12(1):63-71 [FREE Full text] [doi: 10.1080/15504263.2016.1145312] [Medline: 26829356]
53. Ben-Zeev D, Kaiser SM, Krzos I. Remote hovering with individuals with psychotic disorders and substance use: feasibility, engagement, and therapeutic alliance with a text-messaging mobile interventionist. J Dual Diagn 2014;10(4):197-203 [FREE Full text] [doi: 10.1080/15504263.2014.962336] [Medline: 25391277]
54. Bischof G, Grothues J, Reinhardt S, Meyer C, John U, Rumpf H. Evaluation of a telephone-based stepped care intervention for alcohol-related disorders: a randomized controlled trial. Drug Alcohol Depend 2008 Mar 01;93(3):244-251. [doi: 10.1016/j.drugalcdep.2007.10.003] [Medline: 18054443]
55. Bjerke T. It made me feel connected: an exploratory study on the use of mobile SMS in follow-up care for substance abusers. J Addictions Nurs 2008;19(4):195-200. [doi: 10.1080/10884600802504735]
56. Bradford S, Rickwood D. Acceptability and utility of an electronic psychosocial assessment (myAssessment) to increase self-disclosure in youth mental healthcare: a quasi-experimental study. BMC Psychiatry 2015 Dec 01;15:305 [FREE Full text] [doi: 10.1186/s12888-015-0694-4] [Medline: 26627041]
57. Cougle J, Summers B, Allan N, Dillon K, Smith H, Okey S, et al. Hostile interpretation training for individuals with alcohol use disorder and elevated trait anger: a controlled trial of a web-based intervention. Behav Res Ther 2017 Dec;99:57-66. [doi: 10.1016/j.brat.2017.09.004] [Medline: 28941810]
58. Gagnon H, Godin G, Alary M, Bruneau J, Otis J. A randomized trial to evaluate the efficacy of a computer-tailored intervention to promote safer injection practices among drug users. AIDS Behav 2010 Jun;14(3):538-548. [doi: 10.1007/s10461-009-9651-x] [Medline: 20033276]
59. Gonzales R, Ang A, Murphy DA, Glik DC, Anglin MD. Substance use recovery outcomes among a cohort of youth participating in a mobile-based texting aftercare pilot program. J Subst Abuse Treat 2014 Jul;47(1):20-26 [FREE Full text] [doi: 10.1016/j.jsat.2014.01.010] [Medline: 24629885]
60. Gonzales R, Douglas AM, Glik DC. Exploring the feasibility of text messaging to support substance abuse recovery among youth in treatment. Health Educ Res 2014 Feb;29(1):13-22 [FREE Full text] [doi: 10.1093/her/cyt094] [Medline: 24038196]
61. Gonzales R, Hernandez M, Murphy DA, Ang A. Youth recovery outcomes at 6 and 9 months following participation in a mobile texting recovery support aftercare pilot study. Am J Addict 2016 Jan;25(1):62-68 [FREE Full text] [doi: 10.1111/ajad.12322] [Medline: 26689171]
62. Haug S, Lucht MJ, John U, Meyer C, Schaub MP. A pilot study on the feasibility and acceptability of a text message-based aftercare treatment programme among alcohol outpatients. Alcohol Alcohol 2015 Mar;50(2):188-194. [doi: 10.1093/alcalc/agu107] [Medline: 25600249]
63. Ingersoll KS, Dillingham RA, Hettema JE, Conaway M, Freeman J, Reynolds G, et al. Pilot RCT of bidirectional text messaging for ART adherence among nonurban substance users with HIV. Health Psychol 2015 Dec;34S:1305-1315 [FREE Full text] [doi: 10.1037/hea0000295] [Medline: 26651472]
64. Lucht MJ, Hoffman L, Haug S, Meyer C, Pussehl D, Quellmalz A, et al. A surveillance tool using mobile phone short message service to reduce alcohol consumption among alcohol-dependent patients. Alcohol Clin Exp Res 2014 Jun;38(6):1728-1736. [doi: 10.1111/acer.12403] [Medline: 24730528]
65. Reback C, Fletcher J. Theory-based text-messaging to reduce methsex risks among MSM. Drug Alcohol Dep 2015;156:e186-e187. [doi: 10.1016/j.drugalcdep.2015.07.502]
66. Rooke S. A smartphone application for managing cannabis use. Drug Alc Dep 2015;156:e192. [doi: 10.1016/j.drugalcdep.2015.07.516]
67. Santa Ana EJ. Impact of telehealth in-home-messaging devices on alcohol use in dually diagnosed veterans. Drug Alc Dep 2015;156:e197. [doi: 10.1016/j.drugalcdep.2015.07.531]
68. Stoner SA, Arenella PB, Hendershot CS. Randomized controlled trial of a mobile phone intervention for improving adherence to naltrexone for alcohol use disorders. PLoS One 2015;10(4):e0124613 [FREE Full text] [doi: 10.1371/journal.pone.0124613] [Medline: 25909320]
69. Trudeau K. A randomized controlled trial of an online relapse prevention program for adolescents in substance abuse treatment. Child Youth Care Forum 2017;46(3):437-454. [doi: 10.1007/s10566-016-9387-5]
70. Campbell W. Overcoming addictions, a Web-based application, and SMART recovery, an online and in-person mutual help group for problem drinkers, part 2: six-month outcomes of a randomized controlled trial and qualitative feedback from participants. J Med Internet Res 2016;18(10):3-23 [FREE Full text] [doi: 10.2196/jmir.2565] [Medline: 23846588]
71. Chih M. Exploring the use patterns of a mobile health application for alcohol addiction before the initial lapse after detoxification. AMIA Annu Symp Proc 2014:385-394 [FREE Full text] [Medline: 25954342]
72. Chih M, Patton T, McTavish FM, Isham AJ, Judkins-Fisher CL, Atwood AK, et al. Predictive modeling of addiction lapses in a mobile health application. J Subst Abuse Treat 2014 Jan;46(1):29-35 [FREE Full text] [doi: 10.1016/j.jsat.2013.08.004] [Medline: 24035143]
73. Dennis ML, Scott CK, Funk RR, Nicholson L. A pilot study to examine the feasibility and potential effectiveness of using smartphones to provide recovery support for adolescents. Subst Abus 2015;36(4):486-492 [FREE Full text] [doi: 10.1080/08897077.2014.970323] [Medline: 25310057]

XSL•FO
RenderX

74. Ford J, Alagoz E, Dinauer S, Johnson K, Pe-Romashko K, Gustafson D. Successful organizational strategies to sustain use of A-CHESS: a mobile intervention for individuals with alcohol use disorders. J Med Internet Res 2015 Aug 18;17(8):e201 [FREE Full text] [doi: 10.2196/jmir.3965] [Medline: 26286257]
75. Glass JE, McKay JR, Gustafson DH, Kornfield R, Rathouz PJ, McTavish FM, et al. Treatment seeking as a mechanism of change in a randomized controlled trial of a mobile health intervention to support recovery from alcohol use disorders. J Subst Abuse Treat 2017 Dec;77:57-66 [FREE Full text] [doi: 10.1016/j.jsat.2017.03.011] [Medline: 28476273]
76. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
77. McTavish FM, Chih M, Shah D, Gustafson DH. How patients recovering from alcoholism use a Smartphone intervention. J Dual Diagn 2012;8(4):294-304 [FREE Full text] [doi: 10.1080/15504263.2012.723312] [Medline: 23316127]
78. Gonzalez V, Dulin P. Comparison of a smartphone app for alcohol use disorders with an Internet-based intervention plus bibliotherapy: a pilot study. J Consult Clin Psychol 2015 Apr;83(2):335-345 [FREE Full text] [doi: 10.1037/a0038620] [Medline: 25622202]
79. Guarino H, Acosta M, Marsch LA, Xie H, Aponte-Melendez Y. A mixed-methods evaluation of the feasibility, acceptability, and preliminary efficacy of a mobile intervention for methadone maintenance clients. Psychol Addict Behav 2016 Feb;30(1):1-11 [FREE Full text] [doi: 10.1037/adb0000128] [Medline: 26618796]
80. Hasin DS, Aharonovich E, Greenstein E. HealthCall for the smartphone: technology enhancement of brief intervention in HIV alcohol dependent patients. Addict Sci Clin Pract 2014 Feb 17;9:5 [FREE Full text] [doi: 10.1186/1940-0640-9-5] [Medline: 24533631]
81. Aharonovich E, Stohl M, Cannizzaro D, Hasin D. HealthCall delivered via smartphone to reduce co-occurring drug and alcohol use in HIV-infected adults: a randomized pilot trial. J Subst Abuse Treat 2017 Dec;83:15-26 [FREE Full text] [doi: 10.1016/j.jsat.2017.09.013] [Medline: 29129192]
82. McKay JR, Knepper C, Deneke E, O'Reilly C, DuPont RL. An initial evaluation of a comprehensive continuing care intervention for clients with substance use disorders: My First Year of Recovery (MyFYR). J Subst Abuse Treat 2016 Dec;67:50-54 [FREE Full text] [doi: 10.1016/j.jsat.2016.04.003] [Medline: 27296662]
83. Schaub M, Sullivan R, Haug S, Stark L. Web-based cognitive behavioral self-help intervention to reduce cocaine consumption in problematic cocaine users: randomized controlled trial. J Med Internet Res 2012 Nov 28;14(6):e166 [FREE Full text] [doi: 10.2196/jmir.2244] [Medline: 23192752]
84. Schaub MP, Wenger A, Berg O, Beck T, Stark L, Buehler E, et al. A Web-based self-help intervention with and without chat counseling to reduce cannabis use in problematic cannabis users: three-arm randomized controlled trial. J Med Internet Res 2015 Oct 13;17(10):e232 [FREE Full text] [doi: 10.2196/jmir.4860] [Medline: 26462848]
85. Tait R, McKetin R, Kay-Lambkin F, Carron-Arthur B, Bennett A, Bennett K, et al. A Web-based intervention for users of amphetamine-type stimulants: 3-month outcomes of a randomized controlled trial. JMIR Ment Health 2014;1(1):e1 [FREE Full text] [doi: 10.2196/mental.3278] [Medline: 26543901]
86. Tossmann H, Jonas B, Tensil M, Lang P, Strüber E. A controlled trial of an internet-based intervention program for cannabis users. Cyberpsychol Behav Soc Netw 2011 Nov;14(11):673-679. [doi: 10.1089/cyber.2010.0506] [Medline: 21651419]
87. Hendershot CS, Witkiewitz K, George WH, Marlatt GA. Relapse prevention for addictive behaviors. Subst Abuse Treat Prev Policy 2011 Jul 19;6:17 [FREE Full text] [doi: 10.1186/1747-597X-6-17] [Medline: 21771314]
88. Tait RJ, McKetin R, Kay-Lambkin F, Carron-Arthur B, Bennett A, Bennett K, et al. Six-month outcomes of a Web-based intervention for users of amphetamine-type stimulants: randomized controlled trial. J Med Internet Res 2015 Apr 29;17(4):e105 [FREE Full text] [doi: 10.2196/jmir.3778] [Medline: 25925801]
89. Schaub MP, Blankers M, Lehr D, Boss L, Riper H, Dekker J, et al. Efficacy of an internet-based self-help intervention to reduce co-occurring alcohol misuse and depression symptoms in adults: study protocol of a three-arm randomised controlled trial. BMJ Open 2016 Dec 25;6(5):e011457 [FREE Full text] [doi: 10.1136/bmjopen-2016-011457] [Medline: 27225652]
90. Haug S, Lucht MJ, John U, Meyer C, Schaub MP. A pilot study on the feasibility and acceptability of a text message-based aftercare treatment programme among alcohol outpatients. Alcohol Alcohol 2015 Mar;50(2):188-194. [doi: 10.1093/alcalc/agu107] [Medline: 25600249]
91. Campbell W, Hester RK, Lenberg KL, Delaney HD. Overcoming Addictions, a Web-based application, and Smart Recovery, an online and in-person mutual help group for problem drinkers, part 2: six-month outcomes of a randomized controlled trial and qualitative feedback from participants. J Med Internet Res 2016 Oct 04;18(10):e262 [FREE Full text] [doi: 10.2196/jmir.5508] [Medline: 27701064]
92. Bischof G, Grothues J, Reinhardt S, Meyer C, John U, Rumpf H. Evaluation of a telephone-based stepped care intervention for alcohol-related disorders: a randomized controlled trial. Drug Alcohol Depend 2008 Mar 01;93(3):244-251. [doi: 10.1016/j.drugalcdep.2007.10.003] [Medline: 18054443]
93. Gonzales R, Ang A, Murphy DA, Glik DC, Anglin MD. Substance use recovery outcomes among a cohort of youth participating in a mobile-based texting aftercare pilot program. J Subst Abuse Treat 2014 Jul;47(1):20-26 [FREE Full text] [doi: 10.1016/j.jsat.2014.01.010] [Medline: 24629885]

XSL•FO
RenderX

94. Stoner S, Arenella PB, Hendershot CS. Randomized controlled trial of a mobile phone intervention for improving adherence to naltrexone for alcohol use disorders. PLoS One 2015;10(4):e0124613 [FREE Full text] [doi: 10.1371/journal.pone.0124613] [Medline: 25909320]
95. Schaub MP, Maier LJ, Wenger A, Stark L, Berg O, Beck T, et al. Evaluating the efficacy of a web-based self-help intervention with and without chat counseling in reducing the cocaine use of problematic cocaine users: the study protocol of a pragmatic three-arm randomized controlled trial. BMC Psychiatry 2015 Jul 10;15:156 [FREE Full text] [doi: 10.1186/s12888-015-0518-6] [Medline: 26159624]
96. Schaub M, Sullivan R, Haug S, Stark L. Web-based cognitive behavioral self-help intervention to reduce cocaine consumption in problematic cocaine users: randomized controlled trial. J Med Internet Res 2012 Nov 28;14(6):e166 [FREE Full text] [doi: 10.2196/jmir.2244] [Medline: 23192752]
97. Tait R, McKetin R, Kay-Lambkin F, Carron-Arthur B, Bennett A, Bennett K, et al. Six-month outcomes of a Web-based intervention for users of amphetamine-type stimulants: randomized controlled trial. J Med Internet Res 2015 Apr 29;17(4):e105 [FREE Full text] [doi: 10.2196/jmir.3778] [Medline: 25925801]
98. Bhardwaj P. There's a stark contrast between the smartphone market today and what it was in 2010. 2018 May 09. URL: https://nordic.businessinsider.com/smartphone-market-evolutions-winners-and-losers-charts-2018-5 [accessed 2018-08-08] [WebCite Cache ID 71WiadkpT]
99. Danaher BG, Brendryen H, Seeley JR, Tyler MS, Woolley T. From black box to toolbox: outlining device functionality, engagement activities, and the pervasive information architecture of mHealth interventions. Internet Interv 2015 Mar 01;2(1):91-101 [FREE Full text] [doi: 10.1016/j.invent.2015.01.002] [Medline: 25750862]
100. Brendryen H, Johansen A, Nesvåg S, Kok G, Duckert F. Constructing a theory- and evidence-based treatment rationale for complex ehealth interventions: development of an online alcohol intervention using an intervention mapping approach. JMIR Res Protoc 2013 Jan 23;2(1):e6 [FREE Full text] [doi: 10.2196/resprot.2371] [Medline: 23612478]

Abbreviations

A-CHESS: Addiction Comprehensive Health Enhancement Support System
CBT: cognitive behavioral therapy
LBMI-A: Location-Based Monitoring and Intervention System for Alcohol Use Disorders
MyFYR: My First Year of Recovery
MI: motivational interviewing
SBI: Screening and Brief Interaction
SDT: self-determination theory
SUD: substance use disorders

Edited by G Eysenbach; submitted 18.01.18; peer-reviewed by D Epstein, R Ashford, N Unciano; comments to author 08.03.18; revised version received 01.05.18; accepted 18.06.18; published 23.08.18.

Please cite as:
Nesvåg S, McKay JR
Feasibility and Effects of Digital Interventions to Support People in Recovery From Substance Use Disorders: Systematic Review
J Med Internet Res 2018;20(8):e255
URL: http://www.jmir.org/2018/8/e255/
doi:10.2196/jmir.9873
PMID:30139724

©Sverre Nesvåg, James R McKay. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 23.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Participant Recruitment and Retention in Remote eHealth Intervention Trials: Methods and Lessons Learned From a Large Randomized Controlled Trial of Two Web-Based Smoking Interventions

Noreen L Watson[1], PhD; Kristin E Mull[1], MS; Jaimee L Heffner[1], PhD; Jennifer B McClure[2], PhD; Jonathan B Bricker[1,3], PhD

[1]Fred Hutchinson Cancer Research Center, Seattle, WA, United States

[2]Kaiser Permanente Washington Health Research Institute, Seattle, WA, United States

[3]University of Washington, Seattle, WA, United States

Corresponding Author:
Noreen L Watson, PhD
Fred Hutchinson Cancer Research Center
1100 Fairview Avenue North, M3-B232
Seattle, WA, 98109
United States
Phone: 1 2066672942
Email: nlwatson@fredhutch.org

Abstract

Background: Despite having many advantages, online eHealth trials are not without challenges—notably, participant recruitment, and outcome data retention. Moreover, publications from these trials rarely provide detailed information on the methods used for recruitment and retention or discuss implications of the methods for future studies.

Objective: To address this need for empirical guidance regarding recruitment and outcome data retention planning, we aim to describe the methods and lessons learned from the recruitment and retention procedures used in a large randomized trial of 2 Web-based smoking cessation interventions.

Methods: To ensure a demographically and geographically diverse participant sample, we used the recruitment strategies (1) traditional, (2) Web-based, and (3) online survey panel methods and adaptively modified each in response to recruitment success. At baseline, participants indicated how they heard about the study and answered demographic questions. To maximize trial retention at each of the 3-, 6-, and 12-month assessment points, 4 survey modalities (first Web, followed by phone, mail, and postcard) were sequentially timed over a 30-day period. Participants received US $25 for submitting their responses, regardless of modality, and received an additional US $10 bonus for completing the Web survey within 24h of electronic notification.

Results: We randomized 2637 smokers in 16 months and achieved 88% retention at 12-months. Participants (79.26% female, 72.60% Caucasian) were recruited from all 50 states. The majority of participants were recruited through Facebook (49.43%), followed by the survey panel (20.85%), free internet sources (14.54%), traditional media (11.34%), and Google ads (3.84%). Descriptively, participant demographics varied by recruitment source. Of the completed follow-up surveys, most were completed by Web (92%). Retention rates did not vary by recruitment source.

Conclusions: Continuous monitoring and refinement of multiple recruitment methods, particularly of online advertising campaigns, allowed us to maximize the effectiveness of recruitment strategies in recruiting a large, diverse sample of smokers. Likewise, offering multiple follow-up survey modalities in sequential order along with time-dependent bonus incentives enabled us to obtain outcome data from a very high level of enrolled participants for the duration of the trial protocol. These strategies may be similarly useful in other trials.

Trial Registration: ClinicalTrials.gov NCT01812278; https://clinicaltrials.gov/ct2/show/NCT01812278 (Archived by WebCite at http://www.webcitation.org/71gy5GLvO)

(J Med Internet Res 2018;20(8):e10351) doi:10.2196/10351

XSL•FO
RenderX

KEYWORDS

recruitment; retention; randomized controlled trial; RCT; smoking cessation; web intervention

Introduction

In recent years, access to the internet has climbed exponentially worldwide [1]. In parallel with this global trend, more eHealth interventions are being developed and tested in online trials. In fact, as of June 2018, searching "eHealth" in PubMed yields over 29,000 results. Considering the potential for high reach and opportunity to provide low-cost interventions, one of the prominent allures of eHealth interventions is the possibility for them to make population-level health impacts. Moreover, conducting eHealth intervention research online offers a multitude of advantages including automated data collection, high control over intervention content and format, low cost, maximizing external validity, and potential for rapid recruitment of large numbers of participants [2,3]. However, online eHealth trials are not without challenges—most notably, recruitment of research participants and outcome data retention.

Although not unique to online trials [4,5], many online eHealth studies have difficulty obtaining adequate sample size [2,6-9] and recruiting a representative sample of their target population [10-12]. Such difficulties with participant recruitment can result in extended recruitment time, increased costs for recruitment, as well as inadequate statistical power if accrual targets are not met [13]. In fact, failure to meet accrual targets is the primary reason for premature trial termination [4,14,15].

Another major methodological challenge faced by online eHealth trials is keeping participants engaged in the study protocol and achieving adequate rates of outcome data retention [2,9,16-19]. Low retention rates increase the risk of selection bias (particularly if there are differential retention rates between arms), threaten validity, and lead to loss of statistical power [5,17,20]. Moreover, using tobacco treatment studies as an example, methods commonly used to address issues of low retention (eg, imputing all missing outcome data as smoking) are also problematic in that they can lead to inflated type-I and type-II errors [21]. For these reasons, drawing conclusions from studies with low retention rates, even when using conventional methods of imputation [21], can be misleading.

At present, best-practice standards do not exist for how to recruit or retain participants in remotely conducted eHealth intervention trials and most trials do not detail their recruitment and retention methods sufficiently so that others can learn from their successes or mistakes. Among the remotely conducted trials of eHealth interventions that report a more significant amount of detail, there is considerable variability in the amount and type of information provided on the trial recruitment and retention strategies [3,22-26]. As a result, much of the information is not generalizable, it is difficult to determine which methods are most effective under what circumstances, and little is known about how recruitment sources might affect participant characteristics or data retention. Greater transparency of these issues has, therefore, been strongly encouraged in the existing literature [6,11,14,27,28] as detailed reports could inform plans for recruitment and retention in future trials of eHealth interventions.

Thus, the primary aim of this paper is to provide a detailed description of the recruitment and outcome data retention methods used in a successful online eHealth trial (WebQuit), including lessons learned that might be useful for future eHealth trials. The WebQuit trial compared the effectiveness of 2 online smoking cessation programs in a diverse sample of smokers (N=2637) recruited across the US; 88% of participants completed the one-year follow-up survey [29]. To inform future strategies, we also examine the effects of recruitment source on participant characteristics and outcome data retention as well as the effects of participant characteristics on data retention.

Methods

Overview of WebQuit Trial Design

The WebQuit trial was conducted to compare the effectiveness of 2 Web-based interventions for smoking cessation among adult smokers [29]. The websites evaluated in the trial were (1) a Web intervention based on Acceptance and Commitment Therapy (WebQuit.org), and (2) the National Cancer Institute's Smokefree.gov website, which is the most accessed cessation website worldwide. The interventions were designed to be stand-alone interventions. Thus, the trial involved minimal contact with study personnel and did not provide pharmacotherapy to study participants. In addition, to access to their randomly assigned program, participants in both arms could receive up to four messages per day (via text or email) designed to encourage engagement with their assigned website, unless they opted out. These messages were sent for the first 28 days after randomization. Participant follow-up data were collected at 3, 6, and 12 months after randomization.

Target Population.

Participants (N=2637) were adult smokers living in the US. The eligibility requirements for the study included: (1) ≥ 18 years old, (2) smoke ≥5 cigarettes per day for the last year, (3) desire to quit smoking within 30 days, (4) have access to high-speed internet and email, (5) not participating in other cessation interventions or treatment, (6) never having used Smokefree.gov, (7) never having participated in one of our previous studies, (8) have no other household member participating in the study, (9) willingness to be randomized to treatment, complete 3 outcome surveys, and to provide contact information for themselves and 2 relatives, (10) live in the US, and (11) the ability to read in English. To recruit a diverse sample with adequate representation of smokers identifying as racial and ethnic minorities, we aimed to recruit a sample comprised of at least 25% smokers identifying as a racial/ethnic minority (ie, smokers who do not identify as non-Hispanic Caucasian). The target sample size for the study, which was based on having 80% power to detect a two-tailed significant difference between quit rates estimated for the 2 arms from our pilot study and relapse rates [29], was met.

XSL•FO
RenderX

Recruitment Strategy

We recruited participants for 18 months, from March 2014 to August 2015. We implemented a multi-pronged recruitment strategy that encompassed traditional, Web-based, and online survey panel methods (detailed below). Recruitment methods were monitored on an ongoing basis and modified as needed based on recruitment success. Our recruitment strategy was based on methods used in the pilot trial of WebQuit [23,30]. Also, we consulted with the Dana-Farber/Harvard Cancer Center Health Communication Core (DF/HCC) to develop advertisements, a recruitment website, and logo that would be relevant, sensitive, and appealing to our target population and distinguishable from other tobacco cessation websites. The DF/HCC also provided ongoing consultation regarding online advertising strategies for Facebook and Google. To ensure that our data retention operations team would have the capacity to handle the procedures for collecting follow-up data (discussed below), we limited the number of participants able to enroll in the study per month. We were able to closely monitor the study flow and ensure data quality by putting restraints on how quickly we recruited participants. All sources of recruitment directed interested individuals to a study website. The purpose of the website was to establish the credibility of the study and communicate the purpose of the study, how to enroll, and what would be asked of participants. Individuals who consented to be screened for the study were directed to a short screening survey to determine basic eligibility criteria. Individuals deemed eligible by the screening survey were immediately sent an email directing them to an online informed consent.

Traditional Recruitment Methods

Traditional methods for recruitment included press releases as well as newspaper, radio, and television interviews with the principal investigator (JBB). Participants recruited through targeted mailed invitations to known smokers within a large regional health plan (approximately 10,000 letters mailed in batches from December 2014 to August 2015), hospitals, or word of mouth by friends, family, or health care providers were also considered as being recruited through traditional methods.

Free Internet Methods

Participants were recruited through several free Web-based sources including Craigslist, Reddit, visibility of our recruitment website on search engines, and information about the study seen or shared on local and national websites (eg, news websites, fredhutch.org, the Penny Hoarder, Twitter). For recruitment on Craigslist, 296 advertisements were made throughout the recruitment period, with most posts made during the first 10 months. Two advertisements were made per day until all predetermined areas across 18 states were posted in. Predetermined areas were selected based on smoking prevalence [31], economic status, and rural areas (Alaska, Arizona, Arkansas, Indiana, Kentucky, Louisiana, Michigan, Mississippi, Missouri, Nevada, North Dakota, Ohio, Oklahoma, South Carolina, South Dakota, Tennessee, Virginia, and West Virginia). Additional postings were made in 18 cities with high concentrations of smokers [32] (eg, Wilkes Barre-Scranton, PA; Birmingham, AL; Grand Rapids, MI) and in 24 cities with high percentages of Black and African American populations (eg, Miami Gardens, FL; Birmingham, AL; Baltimore, MD). Some areas received up to three posts over the recruitment period.

In addition to a general study advertisement, a season-specific advertisement was placed in several cities around January 2015. Ads were placed in the "Community" tab under both the "General Community" and "Volunteers" subsections of all cities. An example of the wording used in the Craigslist ads can be found in Figure 1. For recruitment on Reddit, 9 posts with similar wording were made throughout the recruitment period in different Reddit subgroups (eg, SampleSize, stopsmoking, addiction).

Facebook Advertisements

In consultation with the DF/HCC, 17 Facebook ads were created that varied in the images and wording used (see Figure 2 for an example ad) to be relevant to our target population. Most images were of cigarettes, the Fred Hutchinson Cancer Research Center logo, or people. Images of people varied by gender, race/ethnicity, the age of the person/people in the photo, and in how many people were pictured (1 to 3). Wording varied around eight content categories: (1) health ("You heard your doctor. And you're ready to quit smoking."), (2) readiness to quit ("You're ready to quit smoking and we're ready to help!"), (3) relationships ("You cherish your time with your children. So you're going to quit smoking."), (4) freedom from cigarettes ("You're ready to be free, You're ready to quit smoking."), (5) research ("Earn up to $105 to quit with us—free, online quit smoking study from the Fred Hutchinson Cancer Research Center!"), (6) financial ("You're ready to spend money on what YOU want, not on cigarettes!"), (7) appearance ("You're ready to look better, smell better—BE better! Are you ready to quit smoking?"), and (8) help/altruism ("You're ready to quit smoking. We're ready to share free skills and support to help you quit!").

Creating multiple ads allowed us to determine which were most effective in real time. For the first 16 days of the ad campaign, advertisements were run in two-day intervals to determine which permutations of images and wording were most successful.

Afterward, the top 3 ads were run one at a time so that Facebook's embedded algorithm could optimize ad performance. One of the most successful ads read, "You heard your doctor. And you're ready to quit smoking. We have a great opportunity—earn up to $105!" Ads were turned on and off in response to which were yielding the highest rates of randomization into the trial at the lowest cost per randomization. To increase the likelihood of reaching our desired population, we set several targeting parameters including ages 18-65 or older, English speaking, US, and people who identified relevant interests (eg, cigarette, quit smoking, electronic cigarette). Facebook ads ran from March 2014 until April 2015. Per recommendations from the DF/HCC and Facebook, we continued to monitor the performance of the ad campaign as response drop-offs are common over time. Minor adjustments were made as needed to boost ad performance.

XSL•FO
RenderX

Figure 1. Content used for general Craigslist advertisements.

Ad Title: Quit smoking and earn up to $105

Do you want to quit smoking in the next 30 days, start living a smoke-free life and earn up to $105?

Fred Hutchinson Cancer Research Center is testing two methods to help daily smokers quit in a study called WebQuit. Participants in the WebQuit study will be randomly assigned to one of two online programs. The goal is to learn which program is most useful for helping people quit smoking.

Participants in either program will receive:

- Tools for dealing more effectively with urges to smoke
- A step-by-step guide for quitting smoking
- Help staying motivated while you are quitting

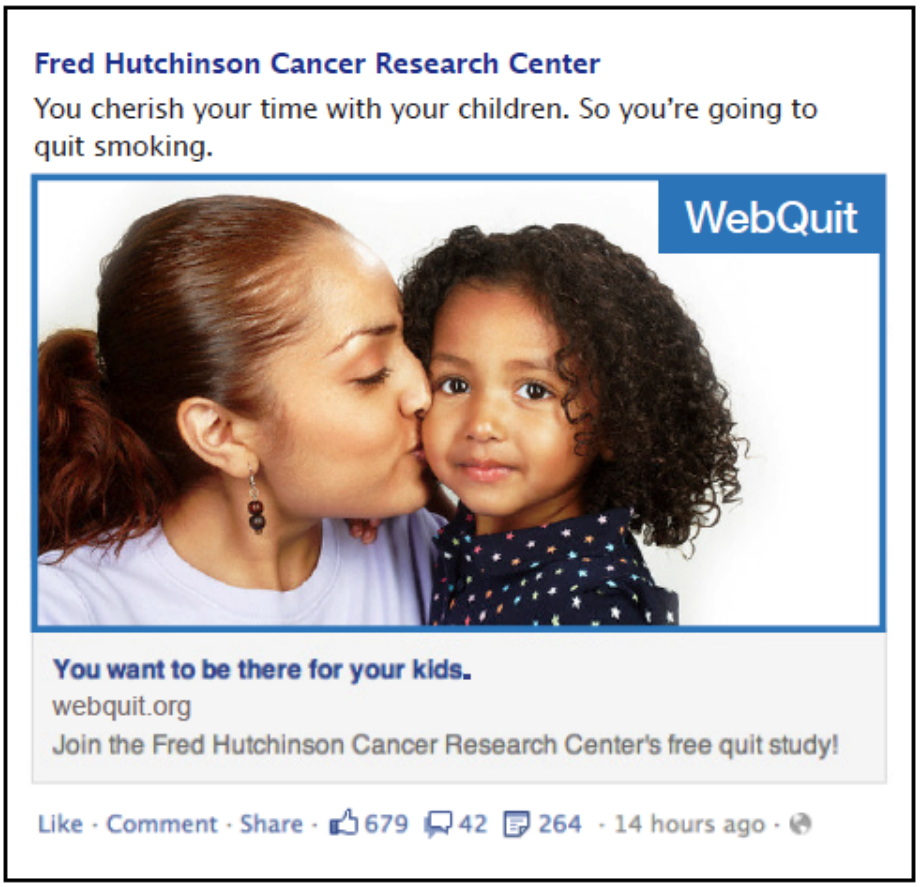
Both programs are free. Participants can earn up to $105 for filling out three brief follow-up surveys over the next 12 months.

Participants must:

- Be willing to complete all three follow-up surveys
- Provide email, phone, and mailing address
- Provide contact information for two relatives.

For more information, visit http://webquit.org or email webquit@fhcrc.org.

Figure 2. Example Facebook ad with image.



XSL•FO
RenderX

Google Advertisements

Google ads were created with the same wording as the Facebook ads. Twenty keywords determined by the study team and DF/HCC that were associated with the study theme (eg, "quit smoking," "help to quit smoking," "stop smoking," "how do I stop smoking," "tips to quit smoking") were rotated through to optimize response. Similar to the strategy used for Craigslist postings, we set parameters to target areas based on high prevalence of smoking, economic status, and rural areas. However, in monitoring conversions from Google ads, it quickly became apparent that the ads were not performing as well as Facebook ads (ie, lower rates of randomization). As a result, and to maximize recruitment efforts and resources, we discontinued the Google ads after approximately two months (March-May 2014).

Online Survey Panel

To help boost recruitment of minority smokers, we utilized the online survey panel company Survey Sampling International (SSI) [33] beginning in April 2015. Such companies can send study announcements to members of their online panels who meet specific criteria. For this study, we requested that they recruit racial/ethnic minority smokers living in the US. The panels are comprised of verified individuals who have voluntarily signed up to participate in online surveys and research in exchange for incentives. SSI was able to identify members of their panel who were likely to be eligible for the study and targeted the audience based on behavioral and demographic characteristics. Respondents recruited from the survey panel answered a brief screener to ensure they were the target audience before being directed to the study website.

In addition to utilizing SSI to boost enrollment of minority smokers, we also limited enrollment on non-minority smokers beginning in February 2015 because we would otherwise not have reached our goal of enrolling at least 25% minority smokers. Specifically, the screening survey was programmed with an algorithm that would randomly reject a set proportion of non-Hispanic Caucasian smokers that were otherwise eligible. For example, if the rejection rate was 30%, non-minority smokers were randomly assigned a number between 0 and 1. If that number were less than 0.30, they would be ineligible. To ensure we met our recruitment goal, the rate of rejection varied over time.

All activities for participant recruitment were coordinated by our project manager in consultation with the study team. On average, the project manager spent an estimated hour per day on recruitment activities throughout the recruitment period, with more time spent at the beginning of the recruitment period. Primary recruitment-related activities included communicating with outside vendors (DF/HCC, SSI, media) and setting up contracts as necessary, reviewing enrollment reports and communicating the enrollment status to key stakeholders, making daily posts to Craigslist (as discussed above), monitoring and tweaking Google and Facebook ads, and communications with our data retention operations team.

Enrollment

All enrollment procedures occurred online by way of the study website noted above. Interested individuals completed a screening survey to assess eligibility criteria. Antifraud measures were also implemented to decrease the likelihood of fraudulent participation (eg, enrolling more than once, changing survey responses to become eligible). Specifically, these measures included reviewing internet protocol addresses for duplicates or non-US origin, CAPTCHA authentication, and review of survey logs for suspicious response times (ie, completing the screening assessment in <90 seconds or completing the baseline survey in <10 minutes). Study staff contacted individuals with suspicious responses to confirm their information. If their information could not be confirmed, they were not enrolled in the study [29]. Eligible individuals were sent an e-mail from the study inviting them to return to the study website to provide informed consent, complete a baseline survey, and provide contact information. Eligible individuals had 14 days to complete the online enrollment process. Automated reminder emails were sent to eligible individuals on days 5 and 11 if they had not returned to the website. On day 7, a personalized email from study staff was sent in case the emails from the study email address were sent to participants' spam folder. After completing the baseline assessment, participants had 30 days to return to the study website for randomization. Up to three weekly email reminders were sent from the study email address to participants who had not returned to the website. Ultimately, 2637 participants were randomized into the trial.

Measures

In the initial screening survey, participants reported how they learned about the study by selecting from one of 13 response options, including an "other" category in which they were able to write out an answer. For this manuscript, response options were grouped into the 5 recruitment methods described above: (1) traditional, (2) free-internet, (3) Facebook ads, (4) Google ads, and (5) online survey panel. We were able to classify all but 9 participants into 1 of the recruitment categories. While we were able to confirm participants recruited through the survey panel, all other responses are self-reported, which is typical for research on recruitment methods [22,34,35]. The baseline assessment also included questions about demographic characteristics and validated self-report screening measures of the following mental health conditions: depression (Center for Epidemiologic Studies Depression scale) [36], generalized anxiety (Generalized Anxiety Disorder 7-item scale) [37], panic disorder (Autonomic Nervous System Questionnaire) [38], posttraumatic stress disorder (PTSD; PTSD Checklist) [39], and social anxiety disorder (Mini-Social Phobia Inventory) [40]. A geographic classification of participants was determined by linking the participants' zip codes to Rural-Urban Commuting Area (RUCA) codes [41]. There are 10 primary classifications based on population density, urbanization, and daily commuting. Definitions of the 10 RUCA codes can be found on the US Department of Agriculture website [41]. Zip codes associated with RUCA codes 1-6 were classified as metropolitan/micropolitan while those associated with RUCA codes 7-10 were classified as small town/rural areas.

XSL•FO
RenderX

Retention Strategy

Follow-up data were collected at 3, 6, and 12 months after randomization. To maximize data retention at each follow-up assessment, participants had up to 30 days to complete the assessments. Four survey modalities were sequentially timed until the survey was completed. Sequentially timing survey modalities has been shown to improve response rates compared to offering multiple survey modalities in parallel [42]. The modalities were (1) Web, (2) telephone, (3) mailed survey, and (4) a postcard with selected outcomes.

For each follow-up assessment, we utilized the following strategy until participants completed the survey. Two weeks before the Web-based follow-up survey was available participants were mailed a survey invitation with a US $2 preincentive. Participants were then sent up to three automated emails with a link to the Web version of the survey on days 0 (exactly 3, 6, or 12 months after randomization), 5, and 9. Afterward, participants had the opportunity to complete the survey via phone. Study staff called participants up to eight times, once per day on days 10-17. On day 18, if participants had not completed the survey, study staff mailed a paper version of the survey with a prestamped and addressed return envelope. If participants did not respond to any of the previous modalities by day 30, they were mailed a postcard that only inquired about primary outcomes and a few selected secondary outcomes.

We incentivized participants with US $25 for completing a survey, regardless of modality. Additionally, to encourage timely responses, participants who completed the Web-based survey within 24 hours of any email received a US $10 bonus. Thus, participants received up to US $105 in incentives for completing follow-up surveys.

Analyses

Descriptives regarding participant demographics across recruitment sources are reported. To examine the association between participant characteristics and recruitment source and 12-month data retention, we used logistic regression models with a covariate for treatment arm and accounted for multiple comparisons by adjusting *P* values to control the false discovery rate [43]. To assess differences in data retention across recruitment sources, we conducted chi-square tests for total response rates at the 3-, 6-, and 12-month follow-up assessments.

Results

Participants

We recruited and randomized 2637 smokers. Of these, 2628 could be classified into 1 of the 5 recruitment categories and were thus retained for these analyses. The mean age of the sample at baseline was 46.15 (SD 13.36) years, and most of the sample was female (2083/2628, 79.26%). A large proportion of participants identified as Caucasian (1908, 72.60%), 278 (10.58%) identified as Black or African American, and 442 (16.82%) identified as another race (Asian, Native American, Native Hawaiian, or more than one race). A total of 222 (8.44%) identified their ethnicity as Hispanic. The remaining demographic characteristics of the overall sample can be found in Table 1.

Recruitment and Demographic Variation by Recruitment Source

Most of the sample was recruited from Facebook (1299/2628, 49.43%), followed by the survey panel (548, 20.85%), free internet sources (382, 14.54%), traditional methods (298, 11.35%), and Google ads (101, 3.84%). Using these recruitment channels, we recruited participants from all 50 states (Figure 3).

Most demographic characteristics of the participants varied across recruitment sources (Table 1). Facebook advertisements recruited the oldest smokers with a mean of 52.87 (SD 10.54) years of age, while the survey panel recruited the youngest with a mean of 35.67 (SD 9.81) years of age. Although most smokers from all recruitment sources were women, traditional recruitment sources yielded the highest percentage of males (85/298, 28.52%), while Facebook ads resulted in the lowest (198/1299, 15.24%). The highest proportion of Black and African American smokers were recruited by both Google ads (18/101, 17.82%) and the online survey panel (98/548, 17.88%). As in the targeted recruitment plan, the online survey panel recruited the highest proportion of smokers identifying as Hispanic ethnicity (111/548, 20.26%) and as a race other than Caucasian or Black/African American (196, 35.77%). Google ads recruited the highest proportion of smokers with a high school education or less (37/101, 36.63%) while traditional recruitment sources recruited the highest proportion of smokers with a bachelor's degree or more (77/298, 25.84%). Free internet sources recruited the highest percentage of participants who identified as lesbian, gay, or bisexual (56/382, 14.66%); with the lowest percentage recruited from Facebook (84/1229, 6.83%). Regarding income, participants recruited from Google ads were most likely to report an income that is greater than US $20,000 (36/101, 35.64%) while participants recruited from the survey panel were most likely to report their income as higher than US $20,000 (432/548, 78.83%). Although most participants in the study lived in metropolitan or micropolitan areas, Facebook recruited the highest proportion of participants from small towns or rural areas (155/1299, 11.93%), whereas traditional methods recruited the lowest proportion from these areas (15/298, 5.03%). All recruitment sources yielded similar proportions of smokers who screened positive for one or more mental health conditions.

XSL•FO
RenderX

Table 1. Demographic characteristics by recruitment source.

Parameter	Total (N=2628), n (%)	Recruitment source					*P* value[a]
		Traditional[b] (n=298), n (%)	Free internet[c] (n=382), n (%)	Facebook ad[b] (n=1299), n (%)	Google ad[c] (n=101), n (%)	Survey panel[d] (n=548), n (%)	
Age, (years), mean (SD)	46.15 (13.36)	46.29 (12.93)	39.56 (12.97)	52.87 (10.54)	40.96 (14.04)	35.67 (9.81)	<.001
Age (years)							<.001
18-24	145 (5.52)	9 (3.02)	42 (10.99)	24 (1.85)	12 (11.88)	58 (10.58)	
25-44	984 (37.44)	117 (39.26)	204 (53.40)	229 (17.63)	47 (46.53)	387 (70.62)	
45-64	1312 (49.92)	148 (49.66)	125 (32.72)	902 (69.44)	38 (37.62)	99 (18.07)	
>65	187 (7.12)	24 (8.05)	11 (2.88)	144 (11.09)	4 (3.96)	4 (0.73)	
Gender							<.001
Male	545 (20.74)	85 (28.52)	108 (28.27)	198 (15.24)	26 (25.74)	128 (23.36)	
Female	2083 (79.26)	213 (71.48)	274 (71.73)	1101 (84.76)	75 (74.26)	420 (76.64)	
Race							<.001
Caucasian	1908 (72.60)	229 (76.85)	246 (64.40)	1116 (85.91)	63 (62.38)	254 (46.35)	
Black/African American	278 (10.58)	22 (7.38)	56 (14.66)	84 (6.47)	18 (17.82)	98 (17.88)	
Other[e]	442 (16.82)	47 (15.77)	80 (20.94)	99 (7.62)	20 (19.80)	196 (35.77)	
Ethnicity							<.001
Hispanic	222 (8.45)	18 (6.04)	40 (10.47)	43 (3.31)	10 (9.90)	111 (20.26)	
Non-Hispanic	2406 (91.55)	280 (93.96)	342 (89.53)	1256 (96.69)	91 (90.10)	437 (79.74)	
Education							<.001
≤High school	733 (27.89)	67 (22.48)	99 (25.92)	414 (31.87)	37 (36.63)	116 (21.17)	
Some college or junior college	1363 (51.86)	154 (51.68)	205 (53.66)	638 (49.11)	45 (44.55)	321 (58.58)	
≥Bachelor's degree	532 (20.24)	77 (25.84)	78 (20.42)	247 (19.01)	19 (18.81)	111 (20.26)	
Sexual orientation							<.001
Heterosexual	2375 (90.37)	264 (88.59)	326 (85.34)	1215 (93.53)	91 (90.19)	479 (87.41)	
LGB[f]	253 (9.63)	34 (11.41)	56 (14.66)	84 (6.47)	10 (9.90)	69 (12.59)	
Mental health							.13
Screened positive for MHC[g]	1930 (73.44)	209 (70.13)	273 (71.47)	933 (71.82)	75 (74.26)	440 (80.29)	
Did not screen positive	574 (21.84)	72 (24.16)	92 (24.08)	284 (21.86)	23 (22.77)	103 (18.80)	
Income (US $)							<.001
≤20,000	735 (27.97)	83 (27.85)	95 (24.87)	405 (31.18)	36 (35.64)	116 (21.17)	
>20,000	1892 (71.99)	215 (72.15)	287 (75.13)	893 (68.75)	65 (64.36)	432 (78.83)	
Location							<.001
Metropolitan or micropolitan	2365 (89.99)	280 (93.96)	356 (93.19)	1137 (87.53)	94 (93.07)	498 (90.88)	
Small town or rural	246 (9.36)	15 (5.03)	24 (6.28)	155 (11.93)	6 (5.94)	46 (8.39)	

[a]Adjusted for false discovery rate.

[b]All advertisements were designed to be culturally sensitive and appealing to our target population; however, no unique population targeting was used for traditional recruitment sources or Facebook ads.

[c]Minimal targeting was used for Craigslist and Google ads in that some locations for posts were chosen to encompass areas with greater smoking prevalence, greater rural areas, and lower economic status. Craigslist ads were also posted in cities with high proportions of Black and African American populations.

[d]The online survey panel was used specifically to boost recruitment of minority (non-Caucasian) smokers.

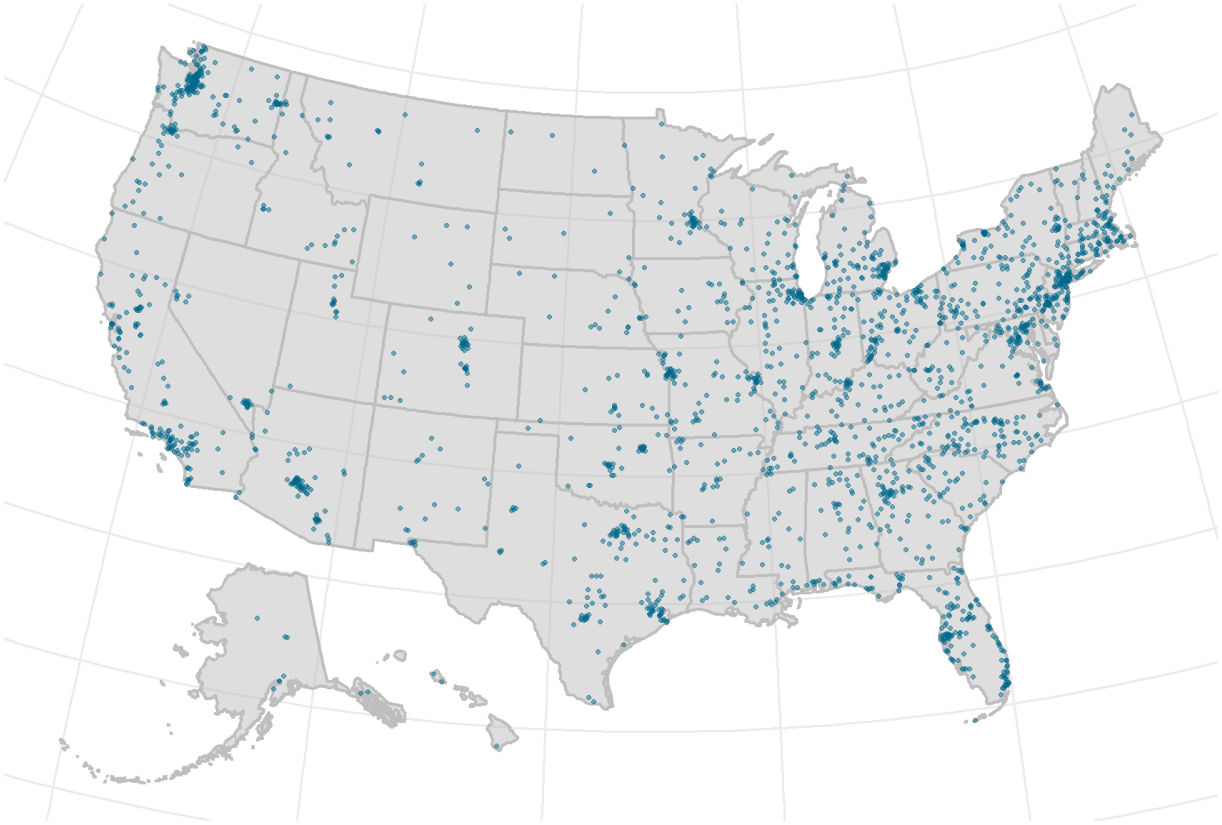
XSL•FO
RenderX

[e]Asian, Native American, Native Hawaiian, or more than one race.

[f]LGB: lesbian, gay, or bisexual.

[g]MHC: mental health condition.

Figure 3. Geographic location of participants from all 50 states; each participant is represented by a single dot.



In assessing baseline participant characteristics associated with data retention (see Table 2), only gender emerged as a significant predictor, with males being less likely to complete the 12-month follow-up assessment than females (81% versus 89%, respectively; OR=0.52, 95% CI 0.40-0.67, *P*<.001). Data retention rates did not differ by recruitment sources at any follow-up assessment (*P*>.05, data not shown).

Costs for Participant Recruitment Advertisements

Excluding costs for personnel, the total cost (US $) of recruitment was approximately $84,083.59, or $31.89 per randomized participant. These costs included $1,995.00 for press releases, $4,054.00 for costs associated with mailed letters (ie, postage, printing, mailing supplies, graphic design), $49,791.49 for Facebook ads, $3,506.00 for Google ads, and $7,644.00 for services provided by SSI. The cost per randomized participant for each recruitment source from highest to lowest was $40.51 for Facebook, $34.71 for Google, $20.30 for traditional sources, and $13.95 for the survey panel.

Rates of Participant Recruitment

Collapsing across all recruitment sources, we enrolled an average of 146 participants per month for the 18-month recruitment period. Recruitment sources varied in rates of participant recruitment (see Table 3).

Outcome Data Retention

Outcome data retention rates for all assessment points and modalities can be found in Table 4. Overall, data retention rates were 88.85%, 89.16%, and 88.17% for the 3-, 6-, and 12-month follow-ups, respectively.

Collapsing across follow-up assessments and assessment modality, a total of 6995 follow-up assessments were completed. The majority of surveys (6386/6995, 91.29%) were completed online. Of the surveys completed online, most (4261/6386, 66.91%) were completed within 24 hours of an email, earning the US $10 bonus incentive. An additional 894 (14.04%) of online surveys were completed prior to any phone calls; the remaining 1231 (19.33%) were completed after phone calls that began on day 10 of the follow-up period. Of the surveys not completed online, (160/609, 26.27%) were completed by phone, 289 (47.45%) were completed via mailed paper versions, and 160 (26.27%) by postcard. In other words, paper surveys accounted for 289/6995 (4.13%) of all survey responses and the phone and postcard surveys each accounted for 160 (2.28%).

XSL•FO
RenderX

Table 2. Baseline predictors of 12-month data retention.

Parameters	Baseline (n)	Not retained (n=327), n (%)	Retained (n=2301), n (%)	Odds ratio (95% CI)	*P* value[a]
Age (years)					.18
18-24	145	24 (16.55)	121 (83.45)	Reference group	
24-44	984	121 (12.30)	863 (87.70)	1.41 (0.87-2.27)	
45-64	1312	151 (11.51)	1161 (88.49)	1.52 (0.93-2.39)	
>65	187	31 (16.58)	156 (83.42)	1.00 (0.55-1.79)	
Gender					<.001
Male	545	103 (18.90)	442 (81.10)	0.52 (0.40-0.67)	
Female	2083	224 (10.75)	1859 (89.25)	Reference group	
Race					.06
Caucasian	1908	255 (13.36)	1653 (86.64)	Reference group	
Black/African American	278	21 (7.55)	257 (92.45)	1.88 (1.18-3.00)	
Other[b]	442	51 (11.54)	391 (88.46)	1.18 (0.86-1.62)	
Ethnicity					.93
Hispanic	222	27 (12.16)	195 (87.84)	1.02 (0.67-1.55)	
Non-Hispanic	2406	300 (12.47)	2106 (87.53)	Reference group	
Education					.14
≥High school	733	108 (14.73)	625 (85.27)	Reference group	
Some college or junior college	1363	161 (11.81)	1202 (88.19)	1.29 (0.99-1.67)	
≥Bachelor's degree	532	58 (10.90)	474 (89.10)	1.42 (1.01-1.99)	
Sexual orientation					.12
Heterosexual	2375	286 (12.04)	2089 (87.96)	1.42 (1.00-2.04)	
LGB[c]	253	41 (16.21)	212 (83.79)	Reference group	
Mental health					.07
Screened positive for MHC[d]	1930	251 (13.01)	1679 (86.99)	1.43 (1.05-1.95)	
Did not screen positive	574	54 (9.41)	520 (90.59)	Reference group	
Income (US $)					.52
≤20,000	735	85 (11.56)	650 (88.44)	0.89 (0.69-1.16)	
>20,000	1892	242 (12.79)	1650 (87.21)	Reference group	
Location					.85
Metropolitan or micropolitan	2635	293 (12.39)	2072 (87.61)	0.94 (0.64-1.39)	
Small town or rural	246	32 (13.01)	214 (86.99)	Reference group	

[a]Adjusted for false discovery rate.

[b]Asian, Native American, Native Hawaiian, more than one race.

[c]LGB: lesbian, gay, or bisexual.

[d]MHC: mental health condition.

Table 3. Rates of recruitment by source.

Parameter	Recruitment source				
	Traditional (n=298)	Free internet (n=382)	Facebook ad (n=1299)	Google ad (n=101)	Survey panel (n=548)
Duration of use (months)	18 months	18 months	14 months	2 months	5 months
Recruitment rate (persons per month)	16.56	21.22	92.79	50.50	109.60

XSL•FO
RenderX

Table 4. Outcome retention rates at 3-, 6-, and 12-month follow-ups by survey modality.

Parameter	Follow-up assessment[a]		
	3-month, n (%)	6-month, n (%)	12-month, n (%)
Online			
≤24 hours of an email	1290 (49.09)	1460 (55.56)	1511 (57.50)
Before any calls	290 (11.0)	308 (11.72)	296 (11.26)
After 1-2 calls	292 (11.04)	243 (9.25)	176 (6.70)
After ≥3 calls	237 (9.02)	142 (5.40)	141 (5.37)
Online total	2109 (80.25)	2153 (81.93)	2124 (80.82)
Phone			
Within 1-2 calls	34 (1.29)	25 (0.95)	36 (1.37)
After ≥3 calls	28 (1.07)	14 (0.53)	23 (0.88)
Phone total	62 (2.36)	39 (1.48)	59 (2.25)
Paper	111 (4.22)	95 (3.61)	83 (3.16)
Postcard	53 (2.02)	56 (2.13)	51 (1.94)
Total number of surveys completed	2335 (88.85)	2343 (89.16)	2317 (88.17)

[a]Data include all randomized participants except for nine that we were unable to classify into one of the five recruitment strategies.

Discussion

Study Objectives

Despite the growing popularity of remotely conducted eHealth trials and difficulties regarding participant recruitment and data retention, most studies do not provide detailed accounts or implications regarding such methodologies. As a result, researchers are left with little guidance when planning these methods for eHealth trials. We sought to add to the literature by explicating the recruitment and retention methods used in the WebQuit trial and discussing implications for future online eHealth intervention trials.

Recruitment and Implications

By implementing a flexible, multi-modal strategy, we recruited and randomized 2637 geographically and demographically diverse adult smokers across the US into a Web-based smoking cessation trial in 18 months. This strategy enabled us to: (1) reallocate resources to methods that were most effective (eg, discontinuing Google ads when Facebook ads were performing better), (2) choose which advertisements to use by monitoring advertisement performance (eg, comparing response rates to Facebook ads), and (3) implement alternative strategies as needed (eg, using an online survey panel to boost recruitment of racial/ethnic minorities).

Unlike findings from our pilot trial [23], participant characteristics varied across recruitment sources. While some differences were intended (eg, a higher proportion of racial and ethnic minority smokers from the online survey panel), other differences were not expected (eg, a more significant proportion of sexual minority smokers from free internet sources; a greater proportion of Black and African American smokers from Google ads). Other studies (eHealth and otherwise) have also found sources of recruitment to be differentially associated with demographic characteristics [22,34,35,44-46]. However, not all studies found the same differences, which may be attributable to differences in recruitment sources used, target population, type of research, and more granular details of recruitment methods (eg, images and words used in advertisements). The variability in participants' demographic characteristics by recruitment source has implications worth considering for future trials and suggests that restricting recruitment to a single recruitment source may limit the sample diversity and, therefore, the generalizability of trial findings.

A further examination of the recruitment methods lends useful insights for future eHealth trials. For example, although previous studies, including the WebQuit pilot [23], have successfully recruited participants with Google ads [30,47,48], relative to Facebook ads with the same text, Google ads significantly underperformed in terms of the number of participants recruited into the present study, leading to our decision to stop implementing these ads early on. Others have also reported poorer performance of Google relative to Facebook ads [49]. Interestingly, however, Google ads recruited the highest proportions of smokers with lower education, income, and higher proportions of smokers who identified as Black or African American, even though only minimal targeting parameters were used to display some ads in areas based on smoking prevalence, economic status, and rural areas. This suggests that while non-targeted Facebook ads may recruit more substantial numbers of people in shorter amounts of time, Google ads may be more effective at recruiting specific subgroups of individuals with only minimal targeting. Future studies should compare these methods systematically to determine if unique targeting strategies can be used to make the platforms equally effective regarding the rate of recruitment and participant demographics.

XSL•FO
RenderX

Nearly half of the participants in this study were recruited from Facebook. Participants recruited from Facebook were predominantly Caucasian (1116/1299, 85.91%), and female (1101, 84.76%)—characteristics also found in other studies of Facebook-recruited participants [10,12,50]. However, relative to our other sources, Facebook recruited the lowest proportion of young adults aged 18-24 years (24, 1.85%), the highest proportion of older participants aged 45-64 years (902, 69.43%) those aged greater than 65 years (144, 11.09%), and the largest proportion (155, 11.93%) of participants living in small towns or rural areas. Contrary to previous reports that Facebook (and other social media platforms) tends to recruit younger samples [11,50,51], these findings add further support [10,52] that Facebook can recruit older participants, even without targeting for specific age groups. In the context of smoking cessation research, this suggests that, without targeting younger age groups, Facebook ads may be more effective for recruiting an older demographic of smokers wanting to quit. It may also be that Facebook's algorithm for optimizing ad performance detected that ads worked best among people aged 45-64, then displayed ads to this group more frequently. These findings also suggest that, even without unique targeting parameters, Facebook may effectively reach some hard-to-reach populations, such as the 19% of the US population living in rural areas [53]. Although not done for this study, research suggests that further targeting and adjusting Facebook ad campaigns may increase the likelihood of reaching the desired population [49,54]. For example, different targeting parameters, advertisement images, and wording may be used strategically to recruit highly-specific subsets of participants [27].

Our original sources of recruitment were not recruiting enough racial/ethnic minority smokers, prompting decisions to use an additional source to recruit minority participants (ie, online survey panel) and program our recruitment website to limit enrollment of non-minority smokers. By implementing this combination of strategies, we met our goal of recruiting at least 25% racial/ethnic minority smokers. A closer look at the other characteristics of participants recruited by the survey panel indicates that they tended to be younger and more likely to have an income higher than US $20,000 relative to participants from other sources. This may be due to the nature of individuals who participate in online survey panels. However, since the only specifications we provided SSI were smokers who identified as a racial/ethnic minority, it is possible that providing additional specifications regarding whom we were seeking to recruit would have produced a different sample.

Most participants in this trial were women. Although some studies have successfully recruited samples comprised entirely of males for eHealth research through targeted advertisement campaigns [27,49,54], it is quite common among eHealth intervention trials to have a greater proportion of women compared to men [3,9,24,25,30,55-58]. This is not surprising as women are more likely to utilize eHealth programs [59]. Future studies, particularly those in which specific demographic variables (eg, gender) are deemed essential, should carefully monitor enrollment of participants with selected demographics to help ensure the desired sample is recruited. Much as we limited enrollment of Caucasian participants in this study, future studies might consider limiting enrollment of participants with certain demographics and creating targeted advertising campaigns to recruit participants meeting specific demographic criteria (eg, males). In making decisions regarding which demographic variables to monitor closely and put enrollment limitations on, we encourage researchers to determine who their target population is because a sample's representativeness is dependent on the characteristics of the target population. For example, characteristics of a representative sample would be different for each of the following target populations: the US adult population, the population of adult smokers, the population of adult smokers who are interested in quitting, and the population of adult smokers interested in quitting with an eHealth intervention.

Overall, when selecting methods for recruitment, researchers should consider many factors including target population, cost, level of reach, targeting abilities, level of ongoing effort required, possible rate of recruitment, and demographic characteristics likely to be recruited from a particular source. For example, in the current study, although the cost per participant enrolled from traditional and free internet sources was relatively low, these methods had limited reach and targeting ability and relatively slow rates of recruitment. In contrast, advertisements through Facebook and Google had a much broader reach, greater ability to target certain populations, faster rates of recruitment, and required low levels of ongoing effort after an initial learning curve. In summary, we highly recommend not only implementing a multi-modal recruitment plan to increase sample diversity, but also monitoring enrollment of participants with characteristics deemed essential for the research question(s). Such practices will not only help ensure that the desired sample is recruited, but can also help researchers determine when alternative recruitment strategies should be implemented to obtain the desired sample.

Retention Strategy and Implications

Yielding an 88.17% 12-month outcome data retention rate, our sequential, multi-modal participant retention strategy was highly effective at obtaining participant follow-up data. We believe 2 key factors contributed to the high rates of data retention. The first is offering bonus incentives for participants who complete their survey early on. Offering a US $10 bonus to the base incentive resulted in 60.91% (4,261/6,995) of all surveys being completed within 24 hours of receiving 1 of the 3 emails for the Web-based survey. The second is sequentially offering different survey modalities [42], thereby offering multiple opportunities and alternative ways of completing the follow-up surveys. Although very few participants completed the surveys via phone, phone calls appear to have prompted many participants to complete their survey online. Interestingly, of the surveys not completed online, more were completed by mail (the third modality offered) than by phone (the second modality). Future research should empirically examine the possibility of reducing participant- and personnel-related costs associated with these methods (eg, reducing the number of phone calls; sending paper versions of the surveys earlier), and still achieve the high rates of data retention found in this study.

XSL•FO
RenderX

Data retention was not related to recruitment source or participant characteristics measured, with the exception that men were less likely to provide outcome data than women. This gender difference in outcome data retention is consistent with some previous eHealth studies [26,60], though results are mixed [3,57,61]. Reasons for this discrepancy are unclear and are beyond the scope of the present study. Despite this difference, 81.10% of men were retained in the present study at 12-months, which is still quite high for remotely conducted eHealth intervention trials. Future research is needed to understand under what conditions men might be less likely to complete follow-up assessments in eHealth intervention trials and to empirically test strategies to improve their retention rates under those conditions.

Limitations

The key limitation of these findings is that the study was not designed to compare the effectiveness or cost-effectiveness of the recruitment or retention strategies. Thus, the findings here are meant to describe the methods we used to recruit and retain adult smokers in the WebQuit trial [29], as well as discuss possible methodological implications for future studies. The field would greatly benefit from empirical research designed to test the efficacy and necessity of different recruitment and retention strategies for remotely conducted eHealth trials as well as from more detailed reports regarding recruitment and retention methods to provide generalizable knowledge. Second, online recruitment methods are rapidly evolving. Thus, it is essential to keep in mind that recruitment for this study occurred from 2014-2015. Higher demand for and advancements in technology-based advertising campaigns may limit the generalizability of these findings in the market today. Differences in recruitment data from the WebQuit pilot study conducted in 2010 [23,30] as compared to the present study help emphasize this point. For example, in contrast to the present study, Google ads outperformed Facebook ads in our pilot study. This change may, in part, be due to updates in Facebook advertising options and their proprietary targeting algorithm between the two studies, which have been updated in many ways since this study was completed and can change without notice. Relatedly, our findings regarding Facebook recruitment may not generalize to future trials as user demographics and use trends evolve. Third, because we recruited adult smokers ready to quit smoking, the findings presented here may not generalize to other eHealth trials seeking to recruit other populations. Fourth, although we were able to validate participants recruited from the online panel, all other reports of recruitment source were self-reported, which is subject to recall bias. Finally, as discussed above, despite being diverse in many other ways, our sample was predominately female, which is consistent with a pattern in eHealth trials overall. While our sample may be representative of the population of smokers interested in quitting with an eHealth intervention, it may not be representative of the entire population of smokers.

Conclusions

Continuous monitoring and refinement of multiple recruitment methods, particularly of online advertising campaigns, was key to our success in recruiting a large, diverse sample of smokers from across the US. Relatedly, offering multiple follow-up survey modalities in sequential order along with time-dependent bonus incentives enabled us to retain most enrolled participants for the duration of the 12-month protocol. Our findings suggest that recruitment sources are associated with demographic differences among participants, but not with differential rates of outcome data retention. Based on the overall success of our participant recruitment and data retention efforts, our experience may serve as an example to others interested in conducting randomized, online clinical intervention trials.

Acknowledgments

We would like to thank the study participants, the invaluable involvement of our study staff at the Fred Hutchinson Cancer Research Center, and the staff at the Kaiser Permanente Washington Health Research Institute for their assistance in recruiting participants. We also thank the National Cancer Institute (R01CA166646 to JBB; R01CA192849 to JBB) and the National Institute on Drug Abuse (R01DA038411 to JBB). NLW's work on the project was supported by a grant from NIDA (F32 DA041100-02). We would also like to thank the Dana-Farber/Harvard Cancer Center for the use of the Health Communication Core, which provided design services for the trial recruitment website. Dana-Farber/Harvard Cancer Center is supported in part by an NCI Cancer Center Support Grant (# NIH 5 P30 CA06516). Funders had no role in the trial conduct or interpretation of results.

Conflicts of Interest

In July 2016, JBB was a consultant to Glaxo Smith Kline, the makers of a nicotine replacement therapy. He now serves on the Scientific Advisory Board of Chrono Therapeutics, the makers of a nicotine replacement therapy device. JLH has received research support from Pfizer, the makers of a smoking cessation medication. Other authors have no declarations.

References

1. Poushter J. Pew Research Center.: Pew Research Center; 2016 Feb 22. Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies URL: http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/ [accessed 2018-08-16] [WebCite Cache ID 71izk5zpA]
2. Murray E, Khadjesari Z, White IR, Kalaitzaki E, Godfrey C, McCambridge J, et al. Methodological challenges in online trials. J Med Internet Res 2009;11(2):e9 [FREE Full text] [doi: 10.2196/jmir.1052] [Medline: 19403465]

XSL•FO
RenderX

3. Anguera JA, Jordan JT, Castaneda D, Gazzaley A, Areán PA. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. BMJ Innov 2016 Jan;2(1):14-21 [FREE Full text] [doi: 10.1136/bmjinnov-2015-000098] [Medline: 27019745]
4. Kasenda B, von EE, You J, Blümle A, Tomonaga Y, Saccilotto R, et al. Prevalence, characteristics, and publication of discontinued randomized trials. JAMA 2014 Mar 12;311(10):1045-1051. [doi: 10.1001/jama.2014.1361] [Medline: 24618966]
5. Gul RB, Ali PA. Clinical trials: the challenge of recruitment and retention of participants. J Clin Nurs 2010 Jan;19(1-2):227-233. [doi: 10.1111/j.1365-2702.2009.03041.x] [Medline: 20500260]
6. 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]
7. Haines-Saah RJ, Kelly MT, Oliffe JL, Bottorff JL. Picture Me Smokefree: a qualitative study using social media and digital photography to engage young adults in tobacco reduction and cessation. J Med Internet Res 2015;17(1):e27 [FREE Full text] [doi: 10.2196/jmir.4061] [Medline: 25624064]
8. Powell J, Newhouse N, Martin A, Jawad S, Yu L, Davoudianfar M, et al. A novel experience-based internet intervention for smoking cessation: feasibility randomised controlled trial. BMC Public Health 2016 Dec 11;16(1):1156 [FREE Full text] [doi: 10.1186/s12889-016-3821-3] [Medline: 27835953]
9. van Spijker BA, Werner-Seidler A, Batterham PJ, Mackinnon A, Calear AL, Gosling JA, et al. Effectiveness of a Web-Based Self-Help Program for Suicidal Thinking in an Australian Community Sample: Randomized Controlled Trial. J Med Internet Res 2018 Feb 14;20(2):e15 [FREE Full text] [doi: 10.2196/jmir.8595] [Medline: 29444769]
10. Carter-Harris L, Bartlett ER, Warrick A, Rawl S. Beyond Traditional Newspaper Advertisement: Leveraging Facebook-Targeted Advertisement to Recruit Long-Term Smokers for Research. J Med Internet Res 2016;18(6):e117 [FREE Full text] [doi: 10.2196/jmir.5502] [Medline: 27306780]
11. Topolovec-Vranic J, Natarajan K. The Use of Social Media in Recruitment for Medical Research Studies: A Scoping Review. J Med Internet Res 2016 Nov 7;18(11):e286 [FREE Full text] [doi: 10.2196/jmir.5698] [Medline: 27821383]
12. Thornton L, Batterham PJ, Fassnacht DB, Kay-Lambkin F, Calear AL, Hunt S. Recruiting for health, medical or psychosocial research using Facebook: Systematic review. Internet Interventions 2016 May;4:72-81. [doi: 10.1016/j.invent.2016.02.001]
13. Sully BGO, Julious SA, Nicholl J. A reinvestigation of recruitment to randomised, controlled, multicenter trials: a review of trials funded by two UK funding agencies. Trials 2013;14:166 [FREE Full text] [doi: 10.1186/1745-6215-14-166] [Medline: 23758961]
14. Carlisle B, Kimmelman J, Ramsay T, MacKinnon N. Unsuccessful trial accrual and human subjects protections: an empirical analysis of recently closed trials. Clin Trials 2015 Feb;12(1):77-83 [FREE Full text] [doi: 10.1177/1740774514558307] [Medline: 25475878]
15. Pica N, Bourgeois F. Discontinuation and Nonpublication of Randomized Clinical Trials Conducted in Children. Pediatrics 2016 Sep;138(3):1-11 [FREE Full text] [doi: 10.1542/peds.2016-0223] [Medline: 27492817]
16. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The Use and Effectiveness of Mobile Apps for Depression: Results From a Fully Remote Clinical Trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
17. 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]
18. Melville KM, Casey LM, Kavanagh DJ. Dropout from Internet-based treatment for psychological disorders. Br J Clin Psychol 2010 Nov;49(Pt 4):455-471. [doi: 10.1348/014466509X472138] [Medline: 19799804]
19. Riper H, Kramer J, Smit F, Conijn B, Schippers G, Cuijpers P. Web-based self-help for problem drinkers: a pragmatic randomized trial. Addiction 2008 Feb;103(2):218-227. [doi: 10.1111/j.1360-0443.2007.02063.x] [Medline: 18199300]
20. Fewtrell MS, Kennedy K, Singhal A, Martin RM, Ness A, Hadders-Algra M, et al. How much loss to follow-up is acceptable in long-term randomised trials and prospective studies? Arch Dis Child 2008 Jun;93(6):458-461. [doi: 10.1136/adc.2007.127316] [Medline: 18495909]
21. Hall SM, Delucchi KL, Velicer WF, Kahler CW, Ranger-Moore J, Hedeker D, et al. Statistical analysis of randomized trials in tobacco treatment: longitudinal designs with dichotomous outcome. Nicotine Tob Res 2001 Aug;3(3):193-202. [doi: 10.1080/14622200110050411] [Medline: 11506764]
22. Gordon JS, Akers L, Severson HH, Danaher BG, Boles SM. Successful participant recruitment strategies for an online smokeless tobacco cessation program. Nicotine Tob Res 2006 Dec;8 Suppl 1:S35-S41. [Medline: 17491169]
23. Sandy M, Butler A. Microbial iron acquisition: marine and terrestrial siderophores. Chem Rev 2009 Oct;109(10):4580-4595 [FREE Full text] [doi: 10.1021/cr9002787] [Medline: 19772347]
24. Ritterband LM, Thorndike FP, Ingersoll KS, Lord HR, Gonder-Frederick L, Frederick C, et al. Effect of a Web-Based Cognitive Behavior Therapy for Insomnia Intervention With 1-Year Follow-up: A Randomized Clinical Trial. JAMA Psychiatry 2017 Jan 01;74(1):68-75. [doi: 10.1001/jamapsychiatry.2016.3249] [Medline: 27902836]
25. Squiers L, Brown D, Parvanta S, Dolina S, Kelly B, Dever J, et al. The SmokefreeTXT (SFTXT) Study: Web and Mobile Data Collection to Evaluate Smoking Cessation for Young Adults. JMIR Res Protoc 2016 Jun 27;5(2):e134 [FREE Full text] [doi: 10.2196/resprot.5653] [Medline: 27349898]

XSL•FO
RenderX

26. Geraghty AWA, Torres LD, Leykin Y, Pérez-Stable EJ, Muñoz RF. Understanding attrition from international Internet health interventions: a step towards global eHealth. Health Promot Int 2013 Sep;28(3):442-452 [FREE Full text] [doi: 10.1093/heapro/das029] [Medline: 22786673]
27. Choi SH, Mitchell J, Lipkus I. Lessons Learned From an Online Study with Dual-smoker Couples. Am J Health Behav 2017 Jan;41(1):61-66. [Medline: 27935791]
28. Lane TS, Armin J, Gordon JS. Online Recruitment Methods for Web-Based and Mobile Health Studies: A Review of the Literature. J Med Internet Res 2015;17(7):e183 [FREE Full text] [doi: 10.2196/jmir.4359] [Medline: 26202991]
29. Bricker JB, Mull KE, McClure JB, Watson NL, Heffner JL. Improving quit rates of web-delivered interventions for smoking cessation: full-scale randomized trial of WebQuit.org versus Smokefree.gov. Addiction 2017 Dec 13:914-923. [doi: 10.1111/add.14127] [Medline: 29235186]
30. Bricker J, Wyszynski C, Comstock B, Heffner JL. Pilot randomized controlled trial of web-based acceptance and commitment therapy for smoking cessation. Nicotine Tob Res 2013 Oct;15(10):1756-1764 [FREE Full text] [doi: 10.1093/ntr/ntt056] [Medline: 23703730]
31. Nguyen K, Marshall L, Hu S, Neff L. State-specific prevalence of current cigarette smoking and smokeless tobacco use among adults aged ≥ 18 years--United States, 2011-2013. Morbidity and Mortality Weekly Report 2015;64(19):532-536.
32. The Daily Beast. 2011 Jan 03. 30 cities with worst smoking problems, from Memphis to Honolulu URL: https://www.thedailybeast.com/30-cities-with-worst-smoking-problems-from-memphis-to-honolulu?ref=scroll [accessed 2018-08-16] [WebCite Cache ID 71j0G8rv6]
33. Survey Sampling International. URL: https://www.surveysampling.com/ [accessed 2018-06-22] [WebCite Cache ID 70N5gduC7]
34. Brodar KE, Hall MG, Butler EN, Parada H, Stein-Seroussi A, Hanley S, et al. Recruiting Diverse Smokers: Enrollment Yields and Cost. Int J Environ Res Public Health 2016 Dec 16;13(12):1-9 [FREE Full text] [doi: 10.3390/ijerph13121251] [Medline: 27999280]
35. Frandsen M, Thow M, Ferguson SG. The Effectiveness Of Social Media (Facebook) Compared With More Traditional Advertising Methods for Recruiting Eligible Participants To Health Research Studies: A Randomized, Controlled Clinical Trial. JMIR Res Protoc 2016;5(3):e161 [FREE Full text] [doi: 10.2196/resprot.5747] [Medline: 27511829]
36. Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. Applied Psychological Measurement 1977 Jun 01;1(3):385-401. [doi: 10.1177/014662167700100306]
37. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
38. Stein MB, Roy-Byrne PP, McQuaid JR, Laffaye C, Russo J, McCahill ME, et al. Development of a brief diagnostic screen for panic disorder in primary care. Psychosom Med 1999;61(3):359-364. [Medline: 10367617]
39. Lang AJ, Wilkins K, Roy-Byrne PP, Golinelli D, Chavira D, Sherbourne C, et al. Abbreviated PTSD Checklist (PCL) as a guide to clinical response. Gen Hosp Psychiatry 2012;34(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2012.02.003] [Medline: 22460001]
40. Connor KM, Kobak KA, Churchill LE, Katzelnick D, Davidson JR. Mini-SPIN: A brief screening assessment for generalized social anxiety disorder. Depress Anxiety 2001;14(2):137-140. [Medline: 11668666]
41. United States Department of Agriculture. Rural-Urban Commuting Area Codes URL: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx [accessed 2018-06-22] [WebCite Cache ID 70N5zeV2f]
42. Dillman DA, Phelps G, Tortora R, Swift K, Kohrell J, Berck J, et al. Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. Social Science Research 2009 Mar;38(1):1-18. [doi: 10.1016/j.ssresearch.2008.03.007]
43. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practicalpowerful approach to multiple testing. Journal of the Royal Statistics Society B 1995;57(1):289-300.
44. Graham AL, Milner P, Saul JE, Pfaff L. Online advertising as a public health and recruitment tool: comparison of different media campaigns to increase demand for smoking cessation interventions. J Med Internet Res 2008;10(5):e50 [FREE Full text] [doi: 10.2196/jmir.1001] [Medline: 19073542]
45. Laws RA, Litterbach EV, Denney-Wilson EA, Russell CG, Taki S, Ong K, et al. A Comparison of Recruitment Methods for an mHealth Intervention Targeting Mothers: Lessons from the Growing Healthy Program. J Med Internet Res 2016 Sep 15;18(9):e248 [FREE Full text] [doi: 10.2196/jmir.5691] [Medline: 27634633]
46. Lindner P, Nyström MB, Hassmén P, Andersson G, Carlbring P. Who seeks ICBT for depression and how do they get there? Effects of recruitment source on patient demographics and clinical characteristics. Internet Interventions 2015 May;2(2):221-225. [doi: 10.1016/j.invent.2015.04.002]
47. Morgan AJ, Jorm AF, Mackinnon AJ. Internet-based recruitment to a depression prevention intervention: lessons from the Mood Memos study. J Med Internet Res 2013;15(2):e31 [FREE Full text] [doi: 10.2196/jmir.2262] [Medline: 23403043]
48. Gross MS, Liu NH, Contreras O, Muñoz RF, Leykin Y. Using Google AdWords for international multilingual recruitment to health research websites. J Med Internet Res 2014 Jan 20;16(1):e18 [FREE Full text] [doi: 10.2196/jmir.2986] [Medline: 24446166]

XSL•FO
RenderX

49. Prescott TL, Phillips IG, DuBois LZ, Bull SS, Mustanski B, Ybarra ML. Reaching Adolescent Gay, Bisexual, and Queer Men Online: Development and Refinement of a National Recruitment Strategy. J Med Internet Res 2016;18(8):e200 [FREE Full text] [doi: 10.2196/jmir.5602] [Medline: 27492781]
50. Whitaker C, Stevelink S, Fear N. The Use of Facebook in Recruiting Participants for Health Research Purposes: A Systematic Review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]
51. Guillory J, Wiant KF, Farrelly M, Fiacco L, Alam I, Hoffman L, et al. Recruiting Hard-to-Reach Populations for Survey Research: Using Facebook and Instagram Advertisements and In-Person Intercept in LGBT Bars and Nightclubs to Recruit LGBT Young Adults. J Med Internet Res 2018 Jun 18;20(6):e197 [FREE Full text] [doi: 10.2196/jmir.9461] [Medline: 29914861]
52. Cowie JM, Gurney ME. The Use of Facebook Advertising to Recruit Healthy Elderly People for a Clinical Trial: Baseline Metrics. JMIR Res Protoc 2018 Jan 24;7(1):e20 [FREE Full text] [doi: 10.2196/resprot.7918] [Medline: 29367186]
53. United States Census Bureau. 2016 Dec 08. New Census Data Show Differences Between Urban and Rural Populations URL: https://www.census.gov/newsroom/press-releases/2016/cb16-210.html [accessed 2018-08-16] [WebCite Cache ID 71j0uHoGs]
54. Reiter PL, Katz ML, Bauermeister JA, Shoben AB, Paskett ED, McRee A. Recruiting Young Gay and Bisexual Men for a Human Papillomavirus Vaccination Intervention Through Social Media: The Effects of Advertisement Content. JMIR Public Health Surveill 2017 Jun 02;3(2):e33 [FREE Full text] [doi: 10.2196/publichealth.7545] [Medline: 28576758]
55. Abroms LC, Boal AL, Simmens SJ, Mendel JA, Windsor RA. A randomized trial of Text2Quit: a text messaging program for smoking cessation. Am J Prev Med 2014 Sep;47(3):242-250. [doi: 10.1016/j.amepre.2014.04.010] [Medline: 24913220]
56. Bricker JB, Copeland W, Mull KE, Zeng EY, Watson NL, Akioka KJ, et al. Single-arm trial of the second version of an acceptance & commitment therapy smartphone application for smoking cessation. Drug Alcohol Depend 2016 Nov 04;170:37-42. [doi: 10.1016/j.drugalcdep.2016.10.029] [Medline: 27870987]
57. Christoforou M, Sáez FJA, Tsakanikos E. Two Novel Cognitive Behavioral Therapy-Based Mobile Apps for Agoraphobia: Randomized Controlled Trial. J Med Internet Res 2017 Nov 24;19(11):e398 [FREE Full text] [doi: 10.2196/jmir.7747] [Medline: 29175809]
58. Cha S, Erar B, Niaura RS, Graham AL. Baseline Characteristics and Generalizability of Participants in an Internet Smoking Cessation Randomized Trial. Ann Behav Med 2016 Oct;50(5):751-761. [doi: 10.1007/s12160-016-9804-x] [Medline: 27283295]
59. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
60. Murray E, White IR, Varagunam M, Godfrey C, Khadjesari Z, McCambridge J. Attrition revisited: adherence and retention in a web-based alcohol trial. J Med Internet Res 2013;15(8):e162 [FREE Full text] [doi: 10.2196/jmir.2336] [Medline: 23996958]
61. Jahangiry L, Shojaeizadeh D, Montazeri A, Najafi M, Mohammad K, Yaseri M. Adherence and Attrition in a Web-Based Lifestyle Intervention for People with Metabolic Syndrome. Iran J Public Health 2014;43(9):1248-1258 [FREE Full text] [Medline: 26175979]

Abbreviations

DF/HCC: Dana-Farber/Harvard Cancer Center Health Communication Core
LGB: lesbian, gay, or bisexual
MHC: mental health condition
RUCA: Rural-Urban Commuting Area
SSI: Survey Sampling International

Edited by G Eysenbach; submitted 08.03.18; peer-reviewed by C Short, N Chavannes; comments to author 29.03.18; revised version received 25.07.18; accepted 01.08.18; published 24.08.18.

Please cite as:
Watson NL, Mull KE, Heffner JL, McClure JB, Bricker JB
Participant Recruitment and Retention in Remote eHealth Intervention Trials: Methods and Lessons Learned From a Large Randomized Controlled Trial of Two Web-Based Smoking Interventions
J Med Internet Res 2018;20(8):e10351
URL: http://www.jmir.org/2018/8/e10351/
doi:10.2196/10351
PMID:30143479

XSL•FO
RenderX

©Noreen L Watson, Kristin E Mull, Jaimee L Heffner, Jennifer B McClure, Jonathan B Bricker. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 24.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Influence of Technology Delivery Mode on Intervention Outcomes: Analysis of a Theory-Based Sexual Health Program

Nicole Levitz[1*], MPH; Erica Wood[1*], MPH; Leslie Kantor[1,2*], MPH, PhD

[1]Planned Parenthood Federation of America, New York, NY, United States

[2]Rutgers School of Public Health, Rutgers University, Newark, NJ, United States

[*]all authors contributed equally

Corresponding Author:
Nicole Levitz, MPH
Planned Parenthood Federation of America
123 William Street
New York, NY, 10038
United States
Phone: 1 9142623408
Email: nicole.levitz@ppfa.org

Abstract

Background: There are few studies on the role of technology delivery mode on health intervention outcomes. Furthermore, the opportunity to examine potential mode effects on a program that is theory-based and integrates principles of communication and decision-making science to influence sexual and reproductive health outcomes is a new contribution to the literature.

Objective: Planned Parenthood Federation of America's national Chat/Text program can be accessed via short message service (SMS; more commonly referred to as text messaging), Web-based desktop chatting, and mobile phone chatting. The program has been in existence since 2010 and has conducted over 1,000,000 conversations. In this study, we examined whether the mode used to access the program (SMS text, desktop chat, or mobile phone chat) affected program users' intention to act on the action plan established in their conversation.

Methods: Data were examined for a 6-month period from January 2016 to June 2016. The data were collected as a part of the monitoring and evaluation of an ongoing program. We limited our sample to the program's priority audience of 15-24 years residing within the United States, which resulted in a sample of 64,939 conversations. Available data items for analysis included user demographics, delivery mode, topic discussed, helpfulness rating (on a 4-point scale), user confidence in following through on the intentions made during the conversation (on a 4-point scale), and educator confidence in whether the user would follow through on the stated intention. Linear and multinomial robust regression analyses were conducted to examine the relationships between conversation delivery mode and confidence.

Results: No significant relationships between users' confidence to carry out their intentions and gender or race were found. None of the 3 modalities (SMS text, desktop chat, or mobile phone chat) were significantly associated with user confidence. All the 3 modalities had significant associations with educator confidence and showed similar effect sizes to those of user confidence. Educator confidence was significantly associated with all the topics discussed.

Conclusions: The Planned Parenthood Chat/Text program was designed as a tool to improve access to sexual and reproductive health care among young people. The mode of intervention delivery was not associated with users' confidence in their ability to carry out their stated intention, suggesting that all modes are legitimate for delivering this intervention. Furthermore, each mode worked across gender and race or ethnicity, indicating that this is a modality that can work across groups.

(J Med Internet Res 2018;20(8):e10398) doi:10.2196/10398

KEYWORDS

sexual and reproductive health; public health; text messaging; instant messaging; behavior theory; internet

XSL•FO
RenderX

Introduction

The nearly ubiquitous use of the internet among adolescents and young adults in the United States has created a potential platform for health programs to reach these populations. Over 75% of teenagers have a smartphone, and an estimated 93% of teenagers in the United States are users of the internet [1]. Many of these adolescents and young adults use Web and mobile phone technology to find health-related information. Northwestern University found that 84% of teenagers searched health information on the Web [2] and that 34% of these individuals reported a change in health behavior because of Web-based health information [3]. However, many digitally-based health programs to date lack a theoretical structure despite evidence that greater effect sizes are observed on intended populations from interventions with a theoretical underpinning as compared with those without one [3-6]. Including behavioral theory within program design helps to address the complex psychological processes that underlie the execution of behavior and is important for the creation of health programs that contribute to sustainable behavioral change. One theory, in particular, that has been underutilized in Web- and mobile phone-based program design is the unified theory of behavior (UTB) [7,8].

The UTB provides a valuable framework for conceptualizing the complex pathway that underlies behavior. This theory emphasizes the determinative role that intention plays in the execution of behavior and draws on several interconnected constructs that affect the formation of intention and that mediate the relationship between intention and behavior [9]. The key constructs that influence the formation of intention are (1) the perceived advantages and disadvantages of performing the behavior, (2) social norms surrounding the behavior, (3) social image repercussions (ie, does this behavior align with the view people hold of themselves?), (4) emotions and affect toward the behavior, and (5) self-efficacy. However, a set intention does not always translate into behavior. Rather, it is mediated by the (1) knowledge and skills needed to carry out the behavior, (2) environmental constraints or facilitators, (3) salience of the behavior, and (4) automatic and habitual psychological processes [7-9]. Altogether, these variables interact to facilitate or hinder one's ability to transform intention into behavior.

Planned Parenthood Federation of America's national Chat/Text Program (the program) incorporates behavioral science by honing in on several of the key UTB constructs that contribute to the formation of intention and by using communication theory to structure conversations to encourage healthy sexual and reproductive health behaviors. Elements of the communication theory included in the program focus on establishing trustworthiness, expertise, and accessibility by the health educators; these elements have a long history in the communication literature and have been found to influence attitude change [10]. Trustworthiness is established through a name exchange; a statement showing expertise is shared immediately following the name exchange, and accessibility is demonstrated by sharing local health center information and proactively inviting users to come back whenever they want.

The program is staffed by trained health educators who freely tailor scripted message library health information to individual user's needs. The scripted message library is a tool for educators to rely on to ensure the accuracy and consistency of health information shared. Educators are trained on health content and how to apply communication and behavioral science, initially with a robust, in-person 40-hour training and then quarterly booster trainings thereafter or more frequently if the quality assurance process identifies a gap or need.

Program conversations are structured to begin with the health educator establishing trustworthiness, expertise, and accessibility with the user and then working with the user to explicitly identify their health concerns and goals. These concerns and goals are then incorporated into the construction of a personalized action plan to address their specific sexual and/or reproductive health concerns. Within each conversation, health educators typically address several UTB constructs, including (1) beliefs surrounding the health behavior in discussion, (2) emotions related to this behavior, (3) self-image, and (4) self-efficacy. In addition, the program incorporates measures that capture users' intentions to act on action plans discussed in conversations. At the end, health educators have the option to ask users to state their intended action plan for addressing their specific health concerns and how confident they are in executing this plan. Once the intention is affirmed by the user, the program helps the user to make a plan that will help ensure that intentions are translated into behavior by working with the user through habits and environmental constraints while addressing knowledge and obstacles. Obstacles that are frequently reported by users are transportation or cost, and health educators can typically assist users with addressing these concerns. Knowledge and skills are generally addressed with how to execute the intention, such as how to take birth control pills at the same time each day. Using the UTB as a framework, confidence in the intention to act on the personalized plan is used as a proxy for behavior because intention is theorized to be its strongest predictor [7-9].

Each month, the program conducts approximately 20,000 conversations with users, most of whom find the program using search and come through the organization's website. The average user identifies as an 18-year-old female, and about half of the users identify as people of color. Each conversation takes an average of 15 min (mean, median, and mode), and there is no time constraint placed on the conversations. Quality assurance is conducted on an on-going basis to assure that educators follow the behavioral science-informed protocol consistently. A team of 5 experienced educators score 3% of all conversations each month on a 19-point metric that includes all key elements of the protocol.

There has been little focus on the impact of the mode of a digitally-based health intervention (eg, text messaging vs Web-based design) on health programs' intended outcomes. Because the program can be accessed via text messaging or short message service (SMS), Web-based desktop chatting, or mobile phone chatting, we sought to examine how a user's choice of mode for accessing the program affected their intention to act on the action plan established in their conversation. In this study, we examined the effects of mode of delivery on both

XSL•FO
RenderX

the user's and the health educator's confidence in executing the user-determined intention set during the course of the conversation. Asking more probing questions to hit on all 5 intention factors may lead to improved outcomes. Adherence to theory within the intervention via content analysis and reviewing differences in effects may clarify how theory can translate into the digital sphere and how it can translate back into outcomes. In addition to examining intentions, we looked at the effects that confidence had on moderating the relationship between mode of delivery and perceived helpfulness by the user.

Methods

Data Collection

Data were collected over a 6-month period from January 2016 to June 2016 as part of the monitoring and evaluation of the ongoing program. Three surveys were the main sources of data for this study. Pre- and postconversation surveys were filled out by the users and a postsurvey was filled out by the health educators. The presurveys gathered users' demographic information and were required for all Web-based users but were optional for text users. The user postsurvey is optional and asks the user to evaluate how helpful the program was. The educator postsurveys included measures for educators to capture the level of confidence the user expressed in executing the discussed action plan when users were asked about it explicitly during the course of the conversation. If users did not share a response about confidence, their data were excluded. Educators also had to indicate how confident they were in the user's ability to act on their plan. In addition, health educators were required to indicate the sexual and reproductive health topics discussed throughout the course of each chat in the postsurvey. As the data were deidentified and did not contain personal health information, this study did not undergo review by an institutional review board.

Measures

Demographic information was obtained from the user prechat surveys. However, demographic data collection varied according to the mode of delivery. Desktop and mobile phone users were required to complete these prechat fields before speaking to a health educator, whereas SMS (text) users were asked for this information but it was not compulsory to provide it before using the service.

Educators recorded the topics discussed during the conversation in their postsurveys. Topics included birth control, emergency contraception, sexually transmitted infection (STI) testing, pregnancy testing, abortion, and other. All topics except for *other* were categorized as dichotomous (*yes discussed* and *no not discussed*). The *other* selection was collected via a qualitative open field, but the specific entries were not analyzed in this study.

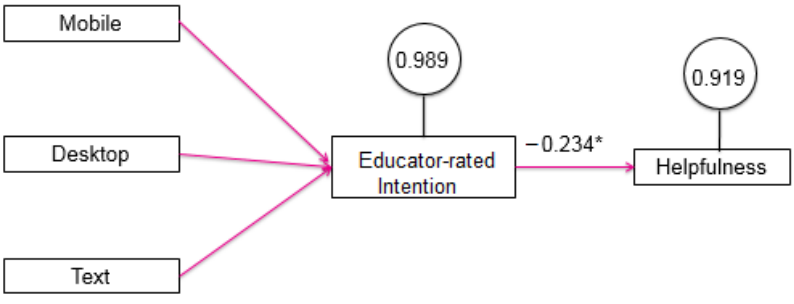
Confidence in the intention to follow through with personalized action plans was measured during and after the conversation by both the user and the educator, respectively. A 4-point Likert scale ranging from *not at all confident* to *very confident* was sent to users by the health educator before the conclusion of the conversation. Users directly indicated to the educator their level of confidence, and the educator then reported their response in their postsurvey. This measure was not used in the final model because of missing data. The educator's confidence scale that measured perceived confidence in the user's intention of acting on the plan discussed was the same 4-point Likert scale. These confidence scales were reverse coded so that higher scores reflected greater confidence (ie, *not at all confident* had a value of 0 and *very confident* had a value of 3). There were also options included for the health educators to indicate if *no next steps [were] discussed* and if the message was not sent but were not excluded from these analyses. The measures of intention were designed for this study.

Analysis

The sample was limited to the program's target audience of 15- to 24-year-olds residing in the United States and excluded users identifying as genders other than male or female, as they made up less than 0.50% (328/65,627) of users. The sample consisted of 64,939 conversations between January and June 2016.

Descriptive statistics, bivariate analyses, and linear regression models were all conducted using IBM SPSS (SPSS Inc, Chicago, IL) Version 21. Correlations were run between demographic variables, topic discussed, user confidence, educator confidence, mode of delivery, and helpfulness. The linear models regressed topics discussed and delivery modes used onto confidence and were created for both user confidence in the execution of their behavioral intention set during the course of conversation and educator confidence in the user executing this behavioral intention. Linear models were adjusted for user age, gender, and race or ethnicity.

Figure 1. Mediation model examining the path from modality to helpfulness. Note that 0.989 is the disturbance term for educator-rated intention, and 0.919 is the disturbance term for helpfulness.



XSL•FO
RenderX

MPlus (Muthén & Muthén, Los Angeles, CA) Version 7.11 was used to conduct the multinomial robust regression analyses. We wanted to see if the topic of conversation moderated the level of confidence an individual user felt by delivery mode (see Figure 1 for path model). This modeling also provided the ability to connect helpfulness and examine its interplay with confidence as a consecutive sequence. First, the user chooses a mode of delivery, and then, they have a conversation about a given topic, establish an action plan, and identify a certain level of confidence in that plan and the degree of helpfulness overall.

Results

Sample

Table 1 displays details about the demographic composition of the sample. Demographic data are sorted by modality to further explicate the factors that contribute to setting a behavioral intention in the course of a conversation. The 3 modes of delivery to use the program are SMS text message, desktop Web-based chat, and mobile phone Web-based chat.

Bivariate Analysis

Correlations were run between demographic variables, topic discussed, user confidence, educator confidence, mode of delivery, and helpfulness. Most of the observed relationships were close to zero and only a few were significant. Significant relationships were observed between user confidence and educator confidence (r=.82, P<.001) and between user confidence and helpfulness (r=.39, P<.001). Users who identified as female constituted the vast majority of the user base during this period; however, there were no significant gender differences by mode of delivery. Users who identified as Hispanic or Latino were 1.8 times more likely to access the program via mobile phone messaging as compared with users who identified as white. Similarly, users who identified as black were 1.6 times more likely to access the program via mobile phone than whites. Users who identified as Asian were 1.8 times more likely to be desktop users as compared with whites. Users who identified as American Indian or Alaska Native or as *other* race or ethnicity category were over 5 times more likely to be text users as compared with users who identified as whites. No other significant demographic distinctions by mode of delivery were observed.

Observed correlations between age groups and topics discussed were close to zero and nonsignificant. There were no significant differences in the observed relationships for each of the 5 main sexual health topics the health educators discuss (ie, birth control, pregnancy testing, abortion, STI testing, and emergency contraception) by age, gender, and race or ethnicity and user confidence, educator confidence, mode of delivery, and helpfulness. These correlations were also all close to zero.

Linear Regressions

No significant relationships between user confidence and gender or race were found (see Table 2 below). However, there were statistically significant differences in educator confidence by both user race and gender. Educators had minimally higher confidence in users who identified as Hispanic (beta=.043, P=.002) and multiracial (beta=.065, P=.003) as compared with whites.

Age did have a small but significant effect on both user and educator confidence. The older the user, the higher the confidence in executing the intention from both the user's (beta=.28, P<.001) and educator's perspective (beta=.27, P<.001).

Table 1. Background data for respondents by modality.

Characteristics	Number of users by modality, n (%)		
	Mobile phone Web-based chat (n=34,136)	Desktop Web-based chat (n=30,328)	Short message service text message (n=475)
Gender			
Female	30,841 (90.35)	27,248 (89.84)	423 (89.05)
Male	3295 (9.65)	3080 (10.16)	52 (10.95)
Race or ethnicity			
White	15,621 (47.37)	16,419 (54.55)	208 (46.43)
African American or black	4670 (14.16)	3192 (10.61)	50 (11.16)
Hispanic or Latino	8291 (25.14)	4851 (16.12)	67 (14.96)
Asian	1546 (4.69)	3046 (10.12)	55 (12.28)
Hawaiian Islander	188 (0.57)	204 (0.68)	0 (0.0)
American Indian or Alaska Native	174 (0.53)	147 (0.49)	11 (2.46)
Multiracial	2487 (7.54)	1868 (6.21)	44 (9.82)
Other	0 (0.0)	370 (1.23)	448 (2.90)
Age group (years)			
15 to 19	22,479 (65.85)	17,774 (58.61)	352 (74.11)
20 to 24	11,657 (34.15)	12,554 (41.39)	123 (25.89)

XSL•FO
RenderX

Table 2. Multivariate regression models predicting user and educator confidence.

Mode of intervention[a]	Beta	Standard error	Margins of error	*P* value
Mobile phone chat (compared with text)				
User	−.210	0.130	±0.25	.11 (NS[b])
Educator	−.226	0.065	±0.13	.02
Text (compared with desktop)				
User	.255	0.130	±0.25	.05 (NS)
Educator	.159	0.065	±0.13	.02
Desktop chat (compared with mobile phone)				
User	.045	0.028	±0.05	.10 (NS)
Educator	.067	0.011	±0.22	<.001
Abortion				
User	.070	0.042	±0.08	.10 (NS)
Educator	.212	0.018	±0.04	<.001
Birth control				
User	.171	0.031	±0.06	<.001
Educator	.171	0.012	±0.02	<.001
Emergency contraception				
User	−.016	0.032	±0.06	.67 (NS)
Educator	.036	0.014	±0.03	.01
Pregnancy testing				
User	−.026	0.027	±0.05	.34 (NS)
Educator	.058	0.012	±0.03	<.001
STI[c] testing				
User	.070	0.029	±0.06	.02
Educator	.171	0.014	±0.03	<.001
Other				
User	−.024	0.054	±0.11	.66 (NS)
Educator	−.103	0.021	±0.04	<.001

[a]Models above control for age, race or ethnicity, and gender.

[b]NS: not significant.

[c]STI: sexually transmitted infection.

About half of the users selected to access the program via mobile phone chat and the other half by desktop. Only 0.70% (475/64,939) of users accessed the program via SMS. All 3 modes were not significantly associated with user confidence. Desktop chat had a nonsignificant negative effect on confidence compared with mobile phone (beta=−.210, *P*=.11), and mobile phone chat was marginally significant with a positive effect (beta=−.255, *P*=.05). All 3 modes of delivery had significant associations with educator confidence and showed similar effect sizes as user confidence. All effects were very small changes on the 4-point Likert scale.

Across all 3 modes, helpfulness gives a sense of overall program satisfaction, 52.44% (4914/9371) strongly agreed and 31.35% (2938/9371) agreed the conversation was helpful. In addition, 7.50% (703/9371) disagreed it was helpful and 8.70% (815/9371) strongly disagreed it was helpful. User confidence in executing the action plan they came up with as part of the conversation was generally high. Overall, 63.99% (1964/3069) of users were very confident, 23.17% (711/3069) were somewhat confident, 12.09% (371/3069) were a little confident, and 0.75% (23/3069) were not at all confident. Educators' confidence in user execution of their action plan was lower than that of the users' own evaluation of confidence. Among educators, 18.02% (3943/21,877) were very confident, 45.7% (10,005/21,877) were somewhat confident, 30.51% (6674/21,877) were a little confident, and 5.74% (1255/21,877) were not at all confident. Only 33.69% (21,877/64,939) of conversations got to the end of the protocol so that educators could evaluate confidence.

XSL•FO
RenderX

Birth control and STI testing were the only topics that significantly affected user confidence. Educator confidence was significantly associated with all of the topics discussed. However, none of these associations were very strong. Users were less confident about next steps after emergency contraception and pregnancy testing conversations, but this relationship was not observed among the educators' responses. Both users and educators had reduced confidence in conversations that discussed topics outside the program scope (ie, *other* topics).

Mediation Model

To take these analyses one step further, we explored a model that regressed mode of delivery onto confidence and confidence onto helpfulness. The addition of helpfulness was to help further assess intention as a proxy of behavior to see whether the theory-based protocol supported users in achieving their immediate goals with the program. As we found with the regressions, there was a great deal of missing data on user intention. We first tried the model using user intention, but there were insufficient data to complete the covariance matrix, and hence, we moved to relying on the educator-reported intention. The covariance matrix needs to be at 0.5 or above to invoke full information maximum likelihood, and user confidence was below this threshold at 0.05. We found the user and educator intention measures to be correlated 0.95 and as such felt comfortable with the use of this proxy. Unfortunately, there was a similar problem with the helpfulness measure and missing data, but there was no logical proxy for helpfulness, so we ran the model with it (see Figure 1 for the final model).

We found a just identified model in which the number of free parameters exactly equals the number of known values (ie, a model with 0 degrees of freedom). As such, no assessment of model fit is possible. In addition, the joint significance test failed for mobile phone and desktop, so we could not move forward with the interpretation of these paths. Text use did pass the joint significance test. The total effect size was 0.857, but these results must be interpreted with a high level of caution given the missing data.

Discussion

Principal Findings

The vast use of the internet and mobile devices has allowed for the creation of health interventions and programs designed to increase health-promoting behaviors among the most vulnerable populations in need of health care access, such as youth of color living in economically disadvantaged areas [3,11]. This program was designed as a tool to improve access to sexual and reproductive health care among young people, and we were eager to see if it could be used as a tool to promote health equity. The equality effects indistinguishable by identity seen among the observed relationships between confidence and across racial and ethnic groups indicate that it can be used as a tool to promote health equity. Previous research conducted by the Pew Research Center found differences in the use of internet and mobile technology by race and ethnicity [1]. African-Americans and Latinos are more likely to rely on the internet for health information than whites [1]. Furthermore, African-Americans and Latinos are more likely to rely on their mobile phones for internet access as compared with whites [1]. These differences are further supported by a process evaluation of this program conducted by Giorgio et al, who found that there are significant demographic differences between those using mobile phone chat and SMS versus desktop chat to access the program [12]. Although we did not see significant differences here in the demographics accessing the program by mode, it is important to consistently monitor and assure that this program is equally accessible to help assure greater access to the populations that need it most.

The lack of effects from these analyses tells an important story. Most of the observed relationships between mode of delivery and intention were close to zero. The strongest observed relationships were between user and educator confidence in following through on intentions and how helpful the user perceived the program to be. This association indicates that intention setting can help meet users' needs. Small, negative correlations were observed among user confidence and the following 3 topics: birth control, STI testing, and pregnancy testing. However, these correlations fall out with more rigorous analysis. In addition, race and gender do not appear to play a significant role in how intention setting works for users.

In general, the older the user, the higher their confidence in executing intention. However, the program leads to essentially the same confidence in intention to follow through on their action plan no matter the user's gender or race or ethnicity. This equality observed by race or ethnicity and gender in user confidence and the minimal effects of these variables on educator confidence indicate that program protocol resonates with users and helps them set an intention they believe they can execute across gender and race or ethnicity. The program was designed for young people aged 15 to 24 years of all identities.

Mode of delivery led to minimal changes on the 4-point Likert scale, indicating that mode does not significantly affect user confidence. Although there were small associations with Hispanic users using text more and Asian users using desktop chat more, mode did not appear to play a role in how the intervention was received. No matter which mode a user came through, they were just as likely to set an action plan that they intended to execute. This finding adds to a limited literature and is important in a world where there is increasing interest in developing digital interventions. This investigation suggests that program developers need to focus on the underlying theoretical approach of interventions and not simply on the mode of delivery when creating interventions.

Limitations

Use of existing program data comes with limitations. We would have liked to examine confidence with more nuance and consistency, but the measures are preset. Although the 4-point metric is considered sufficient for quality assurance, the models here may require scales with more points to illuminate other differences. Furthermore, users often exit the program before making their way through the entire program protocol. Only approximately 60% of users make it all the way through. Thus, many users have not had the opportunity to set an intention, build an action plan, and discuss obstacles and are therefore

XSL•FO
RenderX

never asked about their confidence after the conversation has ended. In addition, as the educator sees the user's response to the confidence question before entering their own, it is inherently biasing their response.

In addition, there were several technological limitations that affected the data collection process. The SMS platform used did not require that SMS users complete survey questions before or after entering a conversation, leaving us with even more missing data for this mode than the other modes. The postchat survey for desktop and mobile phone users was also not mandatory, and users had to opt in for completion, leaving us with another significant source of missing data. We were also limited to the existing cross-sectional data and could not follow-up with users to check if they executed their intended health behaviors. More robust analyses are required using a more complete dataset to further examine the relationships between these variables.

Generalizability of these findings is limited to only the program itself. Mode of delivery and topic may play a different role in other digital health programs using different protocols, theory-based or not.

Acknowledgments

The authors would like to thank the program staff who made this work possible and all the users. The findings and conclusions of this study are those of the authors and do not necessarily represent the views of Planned Parenthood Federation of America, Inc. The authors would also like to thank Dr James Jaccard and Dr Lindsay Bornheimer for their assistance.

Conflicts of Interest

None declared.

References

1. Lenhart A. Pew Internet. 2015. Teens, Social Media & Technology Overview 2015 URL: http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/ [accessed 2016-01-01] [WebCite Cache ID 6xtI1pO6A]
2. Wartella E, Rideout V, Montague H, Beaudoin-Ryan L, Lauricella A. Teens, Health and Technology: A National Survey. MaC 2016 Jun 16;4(3):13 [FREE Full text] [doi: 10.17645/mac.v4i3.515]
3. Guse K, Levine D, Martins S, Lira A, Gaarde J, Westmorland W, et al. Interventions using new digital media to improve adolescent sexual health: a systematic review. J Adolesc Health 2012 Dec;51(6):535-543. [doi: 10.1016/j.jadohealth.2012.03.014] [Medline: 23174462]
4. Portnoy DB, Scott-Sheldon LA, Johnson BT, Carey MP. Computer-delivered interventions for health promotion and behavioral risk reduction: a meta-analysis of 75 randomized controlled trials, 1988-2007. Prev Med 2008 Jul;47(1):3-16 [FREE Full text] [doi: 10.1016/j.ypmed.2008.02.014] [Medline: 18403003]
5. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
6. Bailey J, Mann S, Wayal S, Hunter R, Free C, Abraham C, et al. Sexual health promotion for young people via digital media: a scoping review. Public Health Research 2015;3(13). [Medline: 26583166]
7. Fishbein M, Ajzen I. Predicting and Changing Behavior: The Reasoned Action Approach. New York, NY: Psychology Press; 2015.
8. Jaccard J, Levitz N. Parent-based interventions to reduce adolescent problem behaviors: New directions for self-regulation approaches. In: Self-Regulation in Adolescence. Zurich, Switzerland: Cambridge University Press; 2015:357-388.
9. Jaccard J. 2009. Unlocking the Contraception Conundrum: Reducing Unplanned Pregnancies in Emerging Adulthood URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.365.1723&rep=rep1&type=pdf
10. Guilamo-Ramos V, Jaccard J, Dittus P, Bouris AM. Parental expertise, trustworthiness, and accessibility: parent-adolescent communication and adolescent risk behavior. J Marriage Fam 2006 Dec;68(5):1229-1246. [doi: 10.1111/j.1741-3737.2006.00325.x]
11. Montague E, Perchonok J. Health and wellness technology use by historically underserved health consumers: systematic review. J Med Internet Res 2012 May 31;14(3):e78 [FREE Full text] [doi: 10.2196/jmir.2095] [Medline: 22652979]
12. Giorgio MM, Kantor LM, Levine DS, Arons W. Using chat and text technologies to answer sexual and reproductive health questions: Planned Parenthood pilot study. J Med Internet Res 2013 Sep 20;15(9):e203 [FREE Full text] [doi: 10.2196/jmir.2619] [Medline: 24055754]

Abbreviations

SMS: short message service
STI: sexually transmitted infection
UTB: unified theory of behavior

XSL•FO
RenderX

Edited by G Eysenbach; submitted 13.03.18; peer-reviewed by D Levine, N Leonard; comments to author 12.04.18; revised version received 05.06.18; accepted 28.06.18; published 29.08.18.

Please cite as:
Levitz N, Wood E, Kantor L
The Influence of Technology Delivery Mode on Intervention Outcomes: Analysis of a Theory-Based Sexual Health Program
J Med Internet Res 2018;20(8):e10398
URL: http://www.jmir.org/2018/8/e10398/
doi:10.2196/10398
PMID:30158100

©Nicole Levitz, Erica Wood, Leslie Kantor. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 29.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Face-Aging App for Smoking Cessation in a Waiting Room Setting: Pilot Study in an HIV Outpatient Clinic

Titus Josef Brinker[1,2,3,4], MD; Christian Martin Brieske[3,4], MD; Stefan Esser[3], MD; Joachim Klode[3], MD; Ute Mons[5]; Anil Batra[6], MD; Tobias Rüther[7], MD; Werner Seeger[8], MD; Alexander H Enk[2], MD; Christof von Kalle[1], MD; Carola Berking[9], MD; Markus V Heppt[9], MD; Martina V Gatzka[10], MD; Breno Bernardes-Souza[11], MD; Richard F Schlenk[12], PhD; Dirk Schadendorf[3], MD

[1]National Center for Tumor Diseases, Department of Translational Oncology, German Cancer Research Center, Heidelberg, Germany

[2]Department of Dermatology, Heidelberg University Hospital, University of Heidelberg, Heidelberg, Germany

[3]Department of Dermatology, Essen University Hospital, University of Duisburg-Essen, Essen, Germany

[4]German Cancer Consortium, University of Heidelberg, Heidelberg, Germany

[5]Cancer Prevention Unit, German Cancer Research Center, Heidelberg, Germany

[6]Section for Addiction Medicine and Addiction Research, University Department of Psychiatry and Psychotherapy, University Hospital Tübingen, Tübingen, Germany

[7]Department of Psychiatry and Psychotherapy, Ludwig Maximilian University of Munich, Munich, Germany

[8]Universities of Giessen and Marburg Lung Center, Department of Internal Medicine, Justus-Liebig-University, Gießen, Germany

[9]Department of Dermatology, University Medical Center Munich, University of Munich, Munich, Germany

[10]Department of Dermatology and Allergic Diseases, University of Ulm, Ulm, Germany

[11]School of Medicine, Federal University of Ouro Preto, Ouro Preto, Brazil

[12]Trial Center, National Center for Tumor Diseases, German Cancer Research Center, Heidelberg, Germany

Corresponding Author:
Titus Josef Brinker, MD
National Center for Tumor Diseases
Department of Translational Oncology
German Cancer Research Center
Im Neuenheimer Feld 460
Heidelberg, 69120
Germany
Phone: 49 15175084347
Email: titus.brinker@nct-heidelberg.de

Abstract

Background: There is strong evidence for the effectiveness of addressing tobacco use in health care settings. However, few smokers receive cessation advice when visiting a hospital. Implementing smoking cessation technology in outpatient waiting rooms could be an effective strategy for change, with the potential to expose almost all patients visiting a health care provider without preluding physician action needed.

Objective: The objective of this study was to develop an intervention for smoking cessation that would make use of the time patients spend in a waiting room by passively exposing them to a face-aging, public morphing, tablet-based app, to pilot the intervention in a waiting room of an HIV outpatient clinic, and to measure the perceptions of this intervention among smoking and nonsmoking HIV patients.

Methods: We developed a kiosk version of our 3-dimensional face-aging app Smokerface, which shows the user how their face would look with or without cigarette smoking 1 to 15 years in the future. We placed a tablet with the app running on a table in the middle of the waiting room of our HIV outpatient clinic, connected to a large monitor attached to the opposite wall. A researcher noted all the patients who were using the waiting room. If a patient did not initiate app use within 30 seconds of waiting time, the researcher encouraged him or her to do so. Those using the app were asked to complete a questionnaire.

Results: During a 19-day period, 464 patients visited the waiting room, of whom 187 (40.3%) tried the app and 179 (38.6%) completed the questionnaire. Of those who completed the questionnaire, 139 of 176 (79.0%) were men and 84 of 179 (46.9%) were smokers. Of the smokers, 55 of 81 (68%) said the intervention motivated them to quit (men: 45, 68%; women: 10, 67%);

XSL•FO
RenderX

41 (51%) said that it motivated them to discuss quitting with their doctor (men: 32, 49%; women: 9, 60%); and 72 (91%) perceived the intervention as fun (men: 57, 90%; women: 15, 94%). Of the nonsmokers, 92 (98%) said that it motivated them never to take up smoking (men: 72, 99%; women: 20, 95%). Among all patients, 102 (22.0%) watched another patient try the app without trying it themselves; thus, a total of 289 (62.3%) of the 464 patients were exposed to the intervention (average waiting time 21 minutes).

Conclusions: A face-aging app implemented in a waiting room provides a novel opportunity to motivate patients visiting a health care provider to quit smoking, to address quitting at their subsequent appointment and thereby encourage physician-delivered smoking cessation, or not to take up smoking.

(J Med Internet Res 2018;20(8):e10976) doi:10.2196/10976

KEYWORDS

face aging; smoking cessation; HIV; mobile apps; HIV patients; HIV seropositivity; smoking; cessation; tobacco smoking; morphing

Introduction

There is strong evidence for the effectiveness of addressing tobacco use in health care settings [1-11]. However, few smokers receive cessation advice when visiting a hospital [12,13] which is caused by many different reasons [14] and is therefore difficult to change.

Face-aging interventions, in which a photograph of the user is altered to predict the user's future appearance, have been shown to motivate healthier behavioral choices in adiposity prevention, skin cancer prevention, and smoking cessation settings [15-35]. These preliminary results can be explained by the high importance of appearance for a person's self-concept, particularly during adolescence [36].

However, to the best of our knowledge, the only completed prospective randomized trial to investigate the effectiveness of a face-aging intervention on actual behavior (smoking) was that of Burford et al [37]. Burford and her team recruited 160 participants (80 allocated to the control group and 80 to the intervention group) from 8 metropolitan community pharmacies located around Perth city center in Western Australia. All the participants received standardized smoking cessation advice, but those in the intervention group were also digitally photoaged by the internet-based APRIL Face Aging software to show images of what they might eventually look like as a lifelong smoker and as a nonsmoker. At the 6-month follow-up, 5 (6%) of the 80 control group participants suggested they had quit smoking, although this was confirmed by carbon monoxide validation in only 1 of them. In contrast, 22 (27%) of the 80 intervention group participants reported quitting, with 11 confirmed by carbon monoxide testing, a statistically significant difference in confirmed quitting (χ^2_1=9.0; *P*=.003; test power=80%). However, the study had several limitations: the photographs now appear technologically outdated, they were taken in an over-the-counter setting that always required the time of another person, they were not available for free, and the approach did nothing to address the poor initiation of smoking cessation by doctors as recommended by guidelines [38].

We have developed a 3-dimensional face-aging, tablet-based app, Smokerface, that alters a self-taken image of the user's face to simulate what the user would look like in 1 to 15 years' time as either a smoker or a nonsmoker [39]. In this study, we hypothesized that hospital waiting rooms provide an effective setting for encouraging smoking cessation via the app because this would allow most patients visiting a health care provider to be passively exposed without the need for preluding action by health care personnel. To the best of our knowledge, no previous interventions in the field have implemented new technology for behavioral change in waiting rooms. We chose an HIV outpatient clinic for piloting our intervention, because HIV-positive patients are approximately twice as likely to smoke as the general population [40-45], ensuring that a comparably high number of the sample exposed to the intervention would be current smokers.

Therefore, the aim of this study was to develop an intervention that would make use of the waiting time of patients for smoking cessation by exposing them to the face-aging app and to measure the perceptions of smokers and nonsmokers after using the app.

Methods

Ethical Considerations

We planned this study at the University Hospital of Essen in Germany in early 2017 and implemented it in October 2017. All the participants were adults, and participation in the intervention and questionnaire survey was voluntary. The questionnaire was anonymous, and no personal data were stored. All images were instantly deleted automatically by the kiosk version of the app. We considered oral consent to be sufficient for participation in the survey. Before participants could use the app, they were informed about the screen-mirroring procedure by an information board placed adjacent to the tablet. The ethics committee of the Essen University Hospital, Essen, Germany, approved the study.

Experimental Setup

We developed a kiosk version of the Smokerface app (Figure 1). We placed an Apple iPad (iOS) tablet (Apple Inc, Cupertino, CA, USA) on which this version of the app was running on a table in the middle of the waiting room of our HIV outpatient clinic and connected it to a large monitor attached to the opposite wall, which mirrored the screen of the iPad (Figure 2). An explanatory note was displayed on a board next to the tablet.

XSL•FO
RenderX

Figure 1. Start screen of the kiosk version of our face-aging app Smokerface, running on an Apple iPad (iOS).

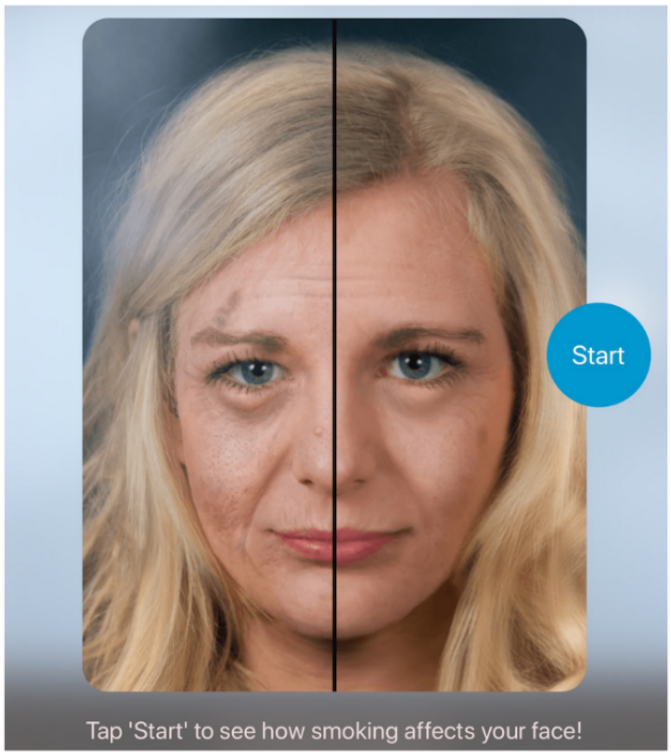
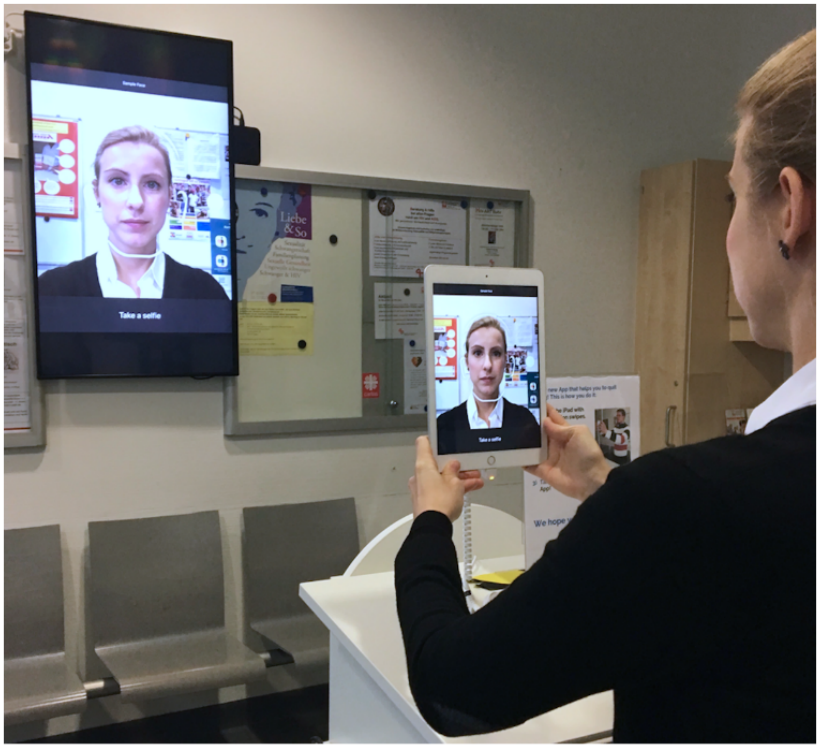


Figure 2. Setup of the smoking cessation face-aging intervention in the waiting room of our HIV outpatient clinic. After the home screen, users were instructed to "Tap 'Start' to see how smoking affects your face!" The original setup was in German.



The tablet screen displayed the instruction "Tap 'Start' to see how smoking affects your face!" (written in German) and was in guided access mode to ensure that patients could not quit the app. The app then displayed images of the patient's face simulating their appearance after 1 to 15 years of not smoking (Figure 3) or smoking (Figure 4).

Procedure

A researcher counted all the patients who visited the waiting room, noted their sex, and timed their total time spent in the waiting room. If a patient did not try the app within 30 seconds of starting their wait, the researcher encouraged the patient to use the app, following a standardized protocol. Patients who

XSL•FO
RenderX

used the app were then asked whether they were smokers or nonsmokers and were asked to complete, voluntarily, the appropriate one of 2 paper-and-pencil questionnaires. Both smoking status–specific questionnaires captured the age and sex of the participant, as well as the participant's perceptions of using the app, on 4-point Likert scales (from "absolutely true" to "absolutely false"). In addition, the participant was asked about the number of other patients in the room during the use of the app, the reactions of the other patients, and how the participant perceived those reactions (on 4-point Likert scales).

Figure 3. Example image of the user simulating how she might look after 9 years of aging without smoking. The screenshot was taken directly from the iPad.

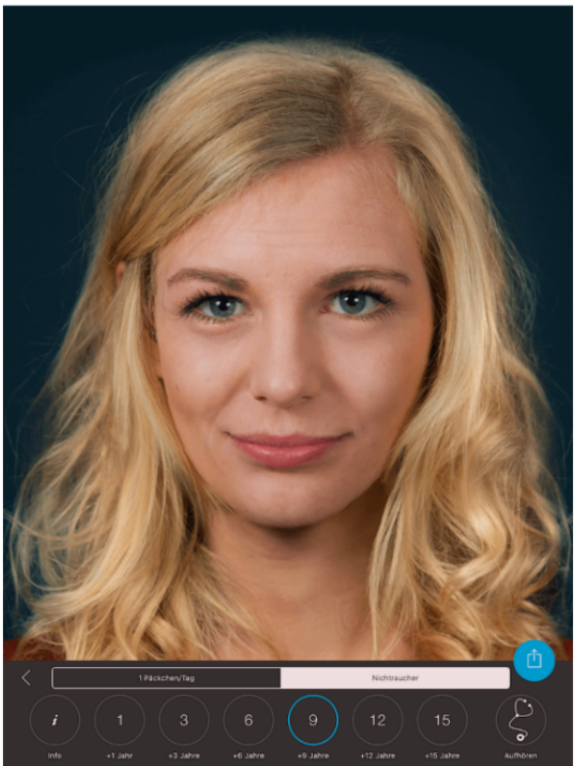
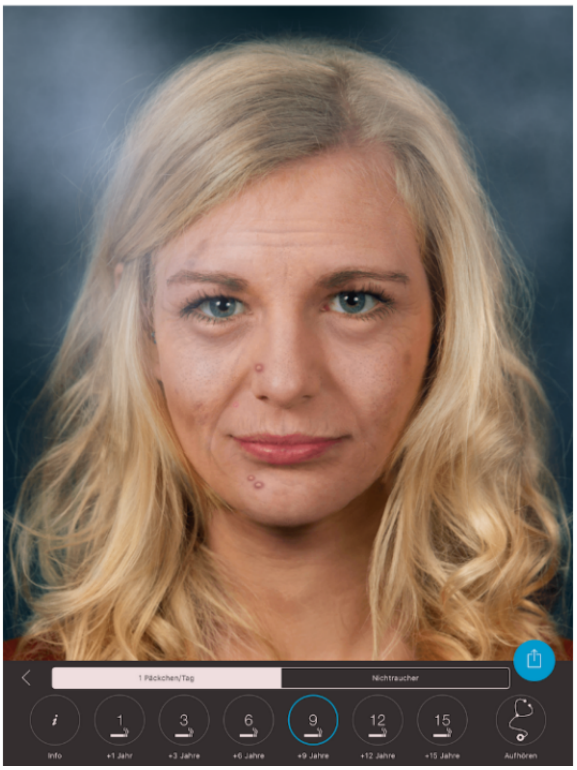


Figure 4. Example image of the user simulating how she might look after 9 years of aging with smoking a pack of cigarettes a day. The screenshot was taken directly from the iPad.



XSL•FO
RenderX

The smokers' questionnaire additionally included 2 standard Fagerström items to calculate the Heaviness of Smoking Index (HSI), a validated measure of smoking dependence that has also been shown to predict quit success [46]: "How many cigarettes do you smoke per day?" and "When do you smoke the first cigarette after waking up?" Prior to the study, we tested the questionnaires in a small subsample of 32 patients to ensure that the questions were understandable and to measure the time needed to complete them (approximately 4 minutes).

Data Analysis

We performed descriptive analysis of data with IBM SPSS Statistics version 25 (IBM Corporation). We undertook no tests for significance due to the explorative nature of the study.

Results

Sample Characteristics

The sample consisted of 464 patients (male: 355/464, 78.7%), of whom 179 filled out a questionnaire (male: 79.0%; median age 42 years; range 23-76 years). Of the 179 patients who completed the questionnaire, 84 (46.9%) smoked; of the smokers, 66/82 (80.5%) were men and 16/82 (19.5%) were women (2 smokers did not indicate their sex).

Among the 84 smokers, 25/83 (30%) had a low HSI, 43/83 (52%) had a medium HSI, and 15/83 (18%) had a high HSI. One participant did not answer both Fagerström items; therefore, we could not calculate the HSI for that person.

Participation

The intervention was implemented in our waiting room for 19 days, and Figure 5 illustrated study participation. The average waiting time for all patients was 21 minutes.

Perceptions About the Intervention

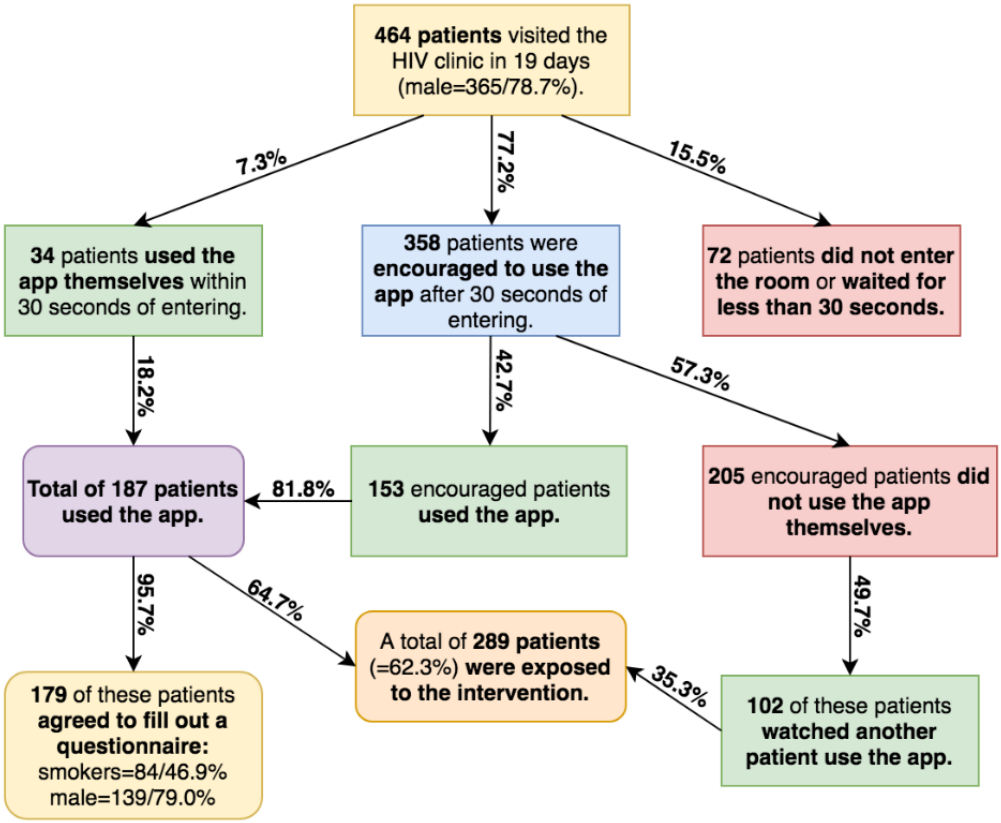
Among the 84 smokers, 55 of 81 (68%) reported that the intervention motivated them to quit (men: 45, 68%; women: 10, 67%, with 3 smokers not answering this question), 41 (51%) reported that it motivated them to discuss quitting with their doctor (men: 32, 49%; women: 9, 60%), and 72 (91%) perceived the intervention as fun (men: 57, 90%; women: 15, 94%). Of the nonsmokers, 92 of 94 (98%) reported that it motivated them to never take up smoking (men: 72, 99%; women: 20, 95%).

Other Patients in the Waiting Room

The numbers of other patients in the waiting room at the time a participant tried the app were as follows: no other patients, 30 (17%) cases; 1 to 3 other patients, 86 (49%) cases; 4 to 6 other patients, 56 (32%) cases; 7 to 10 other patients, 3 (2%) cases; 11 or more patients, 1 (0.6%) case.

Table 1 summarizes the participants' descriptions of the reactions of the other patients in the waiting room (in answer to the question "How did the other people in the room react to your public selfie?"). In a considerable proportion of cases (48/132, 36.4%), 1 or more of the other patients reacted by trying the app themselves.

Figure 5. Levels of participation, sex, and smoking status of the waiting room visitors.



XSL•FO
RenderX

Table 1. Reactions of other patients in the waiting room when a participant tried the app (only the cases where there was at least one other patient in the waiting room).

Patient group	Others tried the app themselves, n (%)	Quitting was a topic of discussion afterward, n (%)	They encouraged me to quit or stay a nonsmoker, n (%)	They were astonished, n (%)	Their reactions were very strong, n (%)
All patients	**N=132**	**N=138**	**N=126**	**N=132**	**N=131**
False/absolutely false	84 (63.6)	61 (44.2)	70 (55.6)	81 (61.4)	90 (68.7)
True/absolutely true	48 (36.3)	77 (55.8)	56 (44.4)	51 (38.6)	41 (31.3)
Smokers	**N=63**	**N=64**	**N=77**	**N=64**	**N=60**
False/absolutely false	40 (63)	23 (36)	24 (41)	39 (61)	42 (70)
True/absolutely true	23 (36)	41 (64)	53 (59)	25 (39)	18 (30)
Nonsmokers	**N=69**	**N=74**	**N=67**	**N=68**	**N=71**
False/absolutely false	44 (64)	38 (51)	46 (69)	42 (62)	48 (68)
True/absolutely true	25 (36)	36 (49)	21 (31)	26 (38)	23 (32)

Table 2. Participants' perceptions of other patients' reactions to the participant's use of the app (only the cases where there was at least one other patient in the waiting room).

Patient group	Motivated me to quit or remain a nonsmoker, n (%)	Helpful, n (%)	They gave me quitting advice, n (%)	There were no reactions, n (%)
Smokers	**N=49**	**N=47**	**N=42**	**N=15**
False/absolutely false	19 (48)	14 (30)	19 (45)	N/A[a]
True/absolutely true	30 (61)	33 (70)	23 (55)	15 (100)
Nonsmokers	**N=51**	**N=42**	**—**	**N=19**
False/absolutely false	11 (22)	15 (36)	N/A	N/A
True/absolutely true	40 (78)	27 (64)	N/A	19 (100)

[a]N/A: not applicable.

In most cases (77/138, 55.8%), the participant's use of the app initiated a discussion on quitting in the waiting room; this was even more the case (41/64, 64%) when the participant was a smoker. In addition, 59% (53/77) of the participants were encouraged to quit by the other patients in the waiting room after using the app, which appeared to be often accompanied by quit advice (23/42, 55%; Table 2).

Table 2 summarizes how the participants perceived the reactions of the other patients, answering the question "How did you perceive those reactions?" The reactions were largely perceived as helpful (by 33/47, 70% of smokers and 27/42, 64% of nonsmokers). Indeed, the reactions provided motivation for 61% (30/49) of the smokers to quit, with advice on quitting offered to 55% (23/42) of the smokers.

Discussion

Principal Findings

The face-aging app setup was successful in exposing the majority of patients visiting our HIV outpatient clinic to a smoking cessation intervention. The face-aging procedure itself and the public nature of the face-aging procedure that triggered reactions of other waiting patients were perceived as motivating to quit smoking and helpful by the majority of smoking as well as nonsmoking patients.

To the best of our knowledge, this is the first study to implement new technology in a waiting room in order to use the patients' waiting time to encourage smoking cessation. The results suggest a huge potential for the large-scale exposure of patients visiting health care providers to the technology. The effectiveness of this study remains subject to further study, and we aim to test the effectiveness in promoting smoking cessation in a randomized controlled trial. Further long-term studies should also examine the effects of group interactions and changes in the subjective norm because of setting the intervention in a waiting room.

Patients who are positive for HIV are approximately twice as likely to smoke as the general population (46.9% of our sample were smokers, compared with 23.9% in the general population) [40-45]. Effective interventions for this patient group remain scarce [4,47-55], and the population-attributable risk of death associated with smoking is double that of the general population [56]. In countries such as Germany where HIV care is well organized and antiretroviral therapy is free of charge, HIV-infected smokers lose more life-years to smoking than to HIV [56]. Hence, while our intervention is not specifically tailored to any patient group and could be applied in any patient

XSL•FO
RenderX

waiting room, an HIV outpatient clinic provided an ideal setting for piloting the intervention.

The nonadherence of physicians to smoking cessation guidelines is a situation that is prevalent in many countries because of role incongruence and a lack of time, financial reimbursement, and appropriate training [14,38]. Being a physician can be an extremely stressful occupation in modern times, with increases in the level of bureaucracy required and the number of patients per doctor. Interventions that help a physician identify and motivate smokers willing to quit have the potential to increase population health and thereby reduce the workload for the medical profession, helping in the fight against tobacco-attributable diseases. Our approach shows promise as a possible simple solution to help physicians meet their smoking cessation obligations without placing too great a burden on them. However, our results also point to limitations and raise questions that need to be addressed in future research.

Initiation of the Use of the App

In this study, the researcher intervened by inviting patients to try the app if they did not do so spontaneously within 30 seconds of starting their wait in the waiting room. We decided on this approach for 3 reasons: (1) the HIV waiting room is particularly busy, with an average waiting time of only 21 minutes; (2) the number of patients waiting there tends to be low, with just 24 patients per day on average between 7:30 AM and 4:45 M; and (3) HIV patients tend to be rather shy in health care settings, as described by the experienced head of our HIV outpatient clinic. Nevertheless, 34 (7.3%) of the 464 patients tried the app within 30 seconds without prompting, indicating the likelihood of successful passive exposure with more time or in settings with higher patient density and longer waiting times. However, it is possible that other patient groups are even more reluctant to have their photograph transmitted to a publicly visible screen, and there might be barriers to the use of such a technology, especially with older patients. In this study, 48 of 132 (36%) participants reported that another patient tried the app straight after seeing them use it, which further strengthens the hypothesis that there would be less need of external prompting in fuller waiting rooms.

Short- and Longer-Term Implications

We received no complaints about patients feeling bullied, according to the physicians who worked in our outpatient clinic at the time of the study, and a great majority of the smokers perceived the intervention as fun. However, the question of feeling bullied could be addressed more explicitly in future research because of the nature of the intervention. We observed that the intervention resulted in interaction between patients where there had been none. Usually, patients waiting in this waiting room sit silently using their mobile phones, reading a newspaper, or just staring at the ground. When the intervention was implemented, patients began talking to each other about the intervention and smoking cessation, finding a common topic they could discuss. Our researcher reported that the overall atmosphere of these conversations was encouraging and positive, and this was reflected in the questionnaire data. Those who had already quit smoking shared their advice and even encouraged addressing the topic at the participant's subsequent appointment; this was reported by 19 of 42 (45%) of the smokers in the questionnaire. In addition, 41 of 81 (51%) of the smokers reported that the intervention itself motivated them to address the topic at their upcoming appointment. Future studies should obtain information from the physicians treating those patients about whether smoking cessation was raised. Following this study, clinicians reported an increased rate of questions on how to quit, but this was not recorded in an objective fashion.

According to our data, 48 of 132 (36.4%) patients tried the app immediately after watching another person do it. In addition, it is reasonable to speculate that simply watching the intervention and perhaps engaging in a conversation arising from it in a full waiting room would motivate a patient to start quitting smoking due to a potential change in their subjective norm [57].

Projection of Potential Effects

During the 19 days of the study, a total of 289 patients in 1 waiting room used or were exposed to the smoking cessation app. This is equivalent to approximately 5500 patients per waiting room per year or to 176,000 patients per year if implemented in all 32 waiting rooms at our hospital. If we assume that the prospective effects measured by Burford et al (that 21.2% of smokers aged 18 to 30 years quit following the use of a similar method [37]) can be transferred to our intervention and that the prevalence of smoking among this hospital's patients is approximately 30%, then approximately 11,000 smoking patients would quit per year. In our sample, the median age was 42 years, meaning that 9 life-years would be saved per patient on average, equivalent to 99,000 saved life-years in total each year [58]. The total cost was US $1500 for 1 waiting room, equivalent to US $0.48 per saved life-year. However, transferability has not been proven, and the prospective effects might be weaker for older patients. In our sample, just 22 (12.3%) of the 179 participants who completed the questionnaire were aged 18 to 30 years. Thus, if the intervention had no effect at all for any patient other than those aged 18 to 30 years, the cost per saved life-year would be US $3.90, 1.320 patients would quit, and 13,200 life-years would be saved per year of implementation.

As for any smoking cessation intervention that has not yet been evaluated in large randomized trials, health systems and insurance companies may be hesitant to reimburse clinics for implementation of this technology. Funding opportunities for health care providers will improve with prospective research on the technology's influence on smoking behavior.

Study Limitations

This study had several limitations. Many of the participants were called by the nurse while still completing the questionnaire. We anticipated this problem and put the questions about individual perceptions of the intervention and important sociodemographic data and smoking status at the start of the questionnaire. The loss of data for these initial items was relatively low. Our study reported only cross-sectional data, and we could only estimate the influence on actual behavior. However, behavioral predictors, such as the behavioral intention to perform a certain behavior, indicate effectiveness in accordance with the theory of planned behavior [57]. In addition,

XSL•FO
RenderX

although anonymity decreases social desirability bias, the participants completing the questionnaires may nevertheless have felt pressure to answer in a socially desirable way because the researcher was present in the room. To minimize this, the participants were left to themselves for completing the questionnaire, which they could then drop into a sealed box to further reduce the risk of bias.

Other Studies That Help Physicians Identify Unhealthy Behaviors

It should be noted that other eHealth interventions can be found in the literature that at least help physicians to identify unhealthy behaviors of patients while only indirectly influencing that behavior [59-76]. These mostly comprise digitized screening and early detection tools. The majority of this work focuses on mental health or the prediction of mental disease [59,64-66,68,69,74,77], and only a few publications have focused on predictors of chronic disease in general, including substance abuse [63,65,68,70,77]. However, helping physicians to identify smokers is only one aim of the intervention presented here; we think it is at least likewise important to investigate its direct effect on quitting behavior in future studies.

Conclusion

The use of a face-aging smoking cessation app in waiting rooms provides a new, enjoyable opportunity to motivate the majority of smokers visiting a health care provider to quit smoking or to address quitting at their subsequent appointment and nonsmokers to never take up smoking. It thereby facilitates physician-delivered smoking cessation. We plan a cluster-randomized trial of the app in 10 waiting rooms. This will focus on long-term smoking abstinence rates, analyzing the impact on different patient subgroups and the interplay of waiting times and modes of initiation. In addition, we plan to repeat the experiment using the Sunface skin cancer awareness app [25] to determine if it shows similar promise.

Authors' Contributions

TJB initiated and designed the study, set up the intervention, instructed data collection, analyzed the data, and wrote the paper. SE, CMB, JK, CvK, AHE, RFS, WS, UM, AB, TR, CB, MvK, MVG, BBS and DS supported the design and conduct of the study, supported statistical analysis, and proofread the manuscript. All authors had full access to the data.

Conflicts of Interest

TJB is the owner of Smart Health Heidelberg GmbH (Handschuhsheimer Landstr. 9/1, 69120 Heidelberg), a technology company that develops and licenses health apps.

References

1. Rigotti NA, Chang Y, Rosenfeld LC, Japuntich SJ, Park ER, Tindle HA, et al. Interactive voice response calls to promote smoking cessation after hospital discharge: pooled analysis of two randomized clinical trials. J Gen Intern Med 2017 Sep;32(9):1005-1013. [doi: 10.1007/s11606-017-4085-z] [Medline: 28616847]
2. Lasser KE, Quintiliani LM, Truong V, Xuan Z, Murillo J, Jean C, et al. Effect of patient navigation and financial incentives on smoking cessation among primary care patients at an urban safety-net hospital: a randomized clinical trial. JAMA Intern Med 2017 Dec 01;177(12):1798-1807. [doi: 10.1001/jamainternmed.2017.4372] [Medline: 29084312]
3. Srivastava AB, Ramsey AT, McIntosh LD, Bailey TC, Fisher SL, Fox L, et al. Tobacco use prevalence and smoking cessation pharmacotherapy prescription patterns among hospitalized patients by medical specialty. Nicotine Tob Res 2018 Feb 22. [doi: 10.1093/ntr/nty031] [Medline: 29481616]
4. Mussulman LM, Faseru B, Fitzgerald S, Nazir N, Patel V, Richter KP. A randomized, controlled pilot study of warm handoff versus fax referral for hospital-initiated smoking cessation among people living with HIV/AIDS. Addict Behav 2018 Mar;78:205-208. [doi: 10.1016/j.addbeh.2017.11.035] [Medline: 29216569]
5. Hammett E, Veldheer S, Hrabovsky S, Yingst J, Berg A, Poole E, et al. TXT2STAYQUIT: pilot randomized trial of brief automated smoking cessation texting intervention for inpatient smokers discharged from the hospital. J Hosp Med 2018 Dec 01;13(7):488-489. [doi: 10.12788/jhm.2907] [Medline: 29370320]
6. Farley A, Koshiaris C, Oke J, Ryan R, Szatkowski L, Stevens R, et al. Physician support of smoking cessation after diagnosis of lung, bladder, or upper aerodigestive tract cancer. Ann Fam Med 2017 Dec;15(5):443-450 [FREE Full text] [doi: 10.1370/afm.2100] [Medline: 28893814]
7. Rice VH, Heath L, Livingstone-Banks J, Hartmann-Boyce J. Nursing interventions for smoking cessation. Cochrane Database Syst Rev 2017 Dec 15;12:CD001188. [doi: 10.1002/14651858.CD001188.pub5] [Medline: 29243221]
8. Pagidipati NJ, Hellkamp A, Thomas L, Gulati M, Peterson ED, Wang TY. Use of prescription smoking cessation medications after myocardial infarction among older patients in community practice. JAMA Cardiol 2017 Sep 01;2(9):1040-1042 [FREE Full text] [doi: 10.1001/jamacardio.2017.2369] [Medline: 28724116]
9. Rigotti NA, Tindle HA, Regan S, Levy DE, Chang Y, Carpenter KM, et al. A post-discharge smoking-cessation intervention for hospital patients: Helping Hand 2 randomized clinical trial. Am J Prev Med 2016 Oct;51(4):597-608 [FREE Full text] [doi: 10.1016/j.amepre.2016.04.005] [Medline: 27647060]
10. Thomas D, Abramson MJ, Bonevski B, George J. System change interventions for smoking cessation. Cochrane Database Syst Rev 2017 Dec 10;2:CD010742. [doi: 10.1002/14651858.CD010742.pub2] [Medline: 28185257]

XSL•FO
RenderX

11. Yudi MB, Farouque O, Andrianopoulos N, Ajani AE, Kalten K, Brennan AL, Melbourne Interventional Group. The prognostic significance of smoking cessation after acute coronary syndromes: an observational, multicentre study from the Melbourne interventional group registry. BMJ Open 2017 Oct 06;7(10):e016874 [FREE Full text] [doi: 10.1136/bmjopen-2017-016874] [Medline: 28988174]
12. Raupach T, Falk J, Vangeli E, Schiekirka S, Rustler C, Grassi MC, et al. Structured smoking cessation training for health professionals on cardiology wards: a prospective study. Eur J Prev Cardiol 2014 Jul;21(7):915-922. [doi: 10.1177/2047487312462803] [Medline: 23008136]
13. Keto J, Jokelainen J, Timonen M, Linden K, Ylisaukko-oja T. Physicians discuss the risks of smoking with their patients, but seldom offer practical cessation support. Subst Abuse Treat Prev Policy 2015 Nov 02;10:43 [FREE Full text] [doi: 10.1186/s13011-015-0039-9] [Medline: 26525302]
14. Balmford J, Leifert JA, Jaehne A. "Tobacco dependence treatment makes no sense because…": rebuttal of commonly-heard arguments against providing tobacco dependence treatment in the hospital setting. BMC Public Health 2014 Nov 19;14:1182 [FREE Full text] [doi: 10.1186/1471-2458-14-1182] [Medline: 25410166]
15. Brinker TJ, Seeger W, Buslaff F. Photoaging mobile apps in school-based tobacco prevention: the mirroring approach. J Med Internet Res 2016 Jun 28;18(6):e183 [FREE Full text] [doi: 10.2196/jmir.6016] [Medline: 27352819]
16. Flett K, Grogan S, Clark-Carter D, Gough B, Conner M. Male smokers' experiences of an appearance-focused facial-ageing intervention. J Health Psychol 2017 Mar;22(4):422-433. [doi: 10.1177/1359105315603477] [Medline: 26338489]
17. Jiwa M, Burford O, Parsons R. Preliminary findings of how visual demonstrations of changes to physical appearance may enhance weight loss attempts. Eur J Public Health 2015 Apr;25(2):283-285. [doi: 10.1093/eurpub/cku249] [Medline: 25644139]
18. Mahler HIM, Kulik JA, Butler HA, Gerrard M, Gibbons FX. Social norms information enhances the efficacy of an appearance-based sun protection intervention. Soc Sci Med 2008 Jul;67(2):321-329 [FREE Full text] [doi: 10.1016/j.socscimed.2008.03.037] [Medline: 18448221]
19. Olson AL, Gaffney CA, Starr P, Dietrich AJ. The impact of an appearance-based educational intervention on adolescent intention to use sunscreen. Health Educ Res 2008 Oct;23(5):763-769 [FREE Full text] [doi: 10.1093/her/cym005] [Medline: 18039727]
20. Owen AL, Grogan S, Clark-Carter D, Dhealthpsy EB. Effects of an appearance-focussed versus a health-focussed intervention on men's attitudes towards UV exposure. Int J Mens Health 2016;15:34 [FREE Full text]
21. Stapleton J, Turrisi R, Hillhouse J, Robinson JK, Abar B. A comparison of the efficacy of an appearance-focused skin cancer intervention within indoor tanner subgroups identified by latent profile analysis. J Behav Med 2010 Jun;33(3):181-190. [doi: 10.1007/s10865-009-9246-z] [Medline: 20058183]
22. Tuong W, Armstrong AW. Effect of appearance-based education compared with health-based education on sunscreen use and knowledge: a randomized controlled trial. J Am Acad Dermatol 2014 Apr;70(4):665-669. [doi: 10.1016/j.jaad.2013.12.007] [Medline: 24508292]
23. Brinker TJ, Brieske CM, Schaefer CM, Buslaff F, Gatzka M, Petri MP, et al. Photoaging mobile apps in school-based melanoma prevention: pilot study. J Med Internet Res 2017 Sep 08;19(9):e319 [FREE Full text] [doi: 10.2196/jmir.8661] [Medline: 28887295]
24. Brinker TJ, Holzapfel J, Baudson TG, Sies K, Jakob L, Baumert HM, et al. Photoaging smartphone app promoting poster campaign to reduce smoking prevalence in secondary schools: the Smokerface Randomized Trial: design and baseline characteristics. BMJ Open 2016 Nov 07;6(11):e014288 [FREE Full text] [doi: 10.1136/bmjopen-2016-014288] [Medline: 27821601]
25. Brinker TJ, Schadendorf D, Klode J, Cosgarea I, Rösch A, Jansen P, et al. Photoaging mobile apps as a novel opportunity for melanoma prevention: pilot study. JMIR Mhealth Uhealth 2017 Jul 26;5(7):e101 [FREE Full text] [doi: 10.2196/mhealth.8231] [Medline: 28747297]
26. Brinker TJ, Seeger W. Photoaging mobile apps: a novel opportunity for smoking cessation? J Med Internet Res 2015 Jul 27;17(7):e186 [FREE Full text] [doi: 10.2196/jmir.4792] [Medline: 26215210]
27. Burford O, Smith M, Jiwa M, Carter O. Photoageing intervention (PAINT): a proposal for a randomised controlled trial in Australian primary care. Australas Med J 2009;1(7):8-12. [doi: 10.4066/AMJ.2009.108]
28. Eastabrook S, Chang P, Taylor MF. Melanoma risk: adolescent females' perspectives on skin protection pre/post-viewing a ultraviolet photoaged photograph of their own facial sun damage. Glob Health Promot 2018 Mar 25;25(1):23-32. [doi: 10.1177/1757975916639871] [Medline: 27333916]
29. Faria BL, Brieske CM, Cosgarea I, Omlor AJ, Fries FN, de Faria COM, et al. A smoking prevention photoageing intervention for secondary schools in Brazil delivered by medical students: protocol for a randomised trial. BMJ Open 2017 Dec 10;7(12):e018589 [FREE Full text] [doi: 10.1136/bmjopen-2017-018589] [Medline: 29229659]
30. Mahler HIM, Kulik JA, Gerrard M, Gibbons FX. Effects of photoaging information and UV photo on sun protection intentions and behaviours: a cross-regional comparison. Psychol Health 2013;28(9):1009-1031. [doi: 10.1080/08870446.2013.777966] [Medline: 23537173]
31. Lo Presti LL, Chang P, Taylor MF. Young Australian adults' reactions to viewing personalised UV photoaged photographs. Australas Med J 2014;7(11):454-461 [FREE Full text] [doi: 10.4066/AMJ.2014.2253] [Medline: 25550717]

XSL•FO
RenderX

32. Brinker TJ, Owczarek AD, Seeger W, Groneberg DA, Brieske CM, Jansen P, et al. A medical student-delivered smoking prevention program, education against tobacco, for secondary schools in Germany: randomized controlled trial. J Med Internet Res 2017 Jun 06;19(6):e199 [FREE Full text] [doi: 10.2196/jmir.7906] [Medline: 28588007]
33. Brinker TJ, Heckl M, Gatzka M, Heppt MV, Resende RH, Schneider S, et al. A skin cancer prevention facial-aging mobile app for secondary schools in Brazil: appearance-focused interventional study. JMIR Mhealth Uhealth 2018 Mar 09;6(3):e60 [FREE Full text] [doi: 10.2196/mhealth.9794] [Medline: 29523502]
34. Brinker TJ, Alfitian J, Seeger W, Groneberg DA, von KC, Enk AH, et al. A Face-Aging Smoking Prevention/Cessation Intervention for Nursery School Students in Germany: An Appearance-Focused Interventional Study. Int J Environ Res Public Health 2018 Aug 04;15(8) [FREE Full text] [doi: 10.3390/ijerph15081656] [Medline: 30081549]
35. Brinker TJ, Klode J, Esser S, Schadendorf D. Facial-Aging App Availability in Waiting Rooms as a Potential Opportunity for Skin Cancer Prevention. JAMA Dermatol 2018 Jul 25. [doi: 10.1001/jamadermatol.2018.1907] [Medline: 30046805]
36. Baudson TG, Weber KE, Freund PA. More than only skin deep: appearance self-concept predicts most of secondary school students' self-esteem. Front Psychol 2016;7:1568 [FREE Full text] [doi: 10.3389/fpsyg.2016.01568] [Medline: 27803681]
37. Burford O, Jiwa M, Carter O, Parsons R, Hendrie D. Internet-based photoaging within Australian pharmacies to promote smoking cessation: randomized controlled trial. J Med Internet Res 2013 Mar 26;15(3):e64 [FREE Full text] [doi: 10.2196/jmir.2337] [Medline: 23531984]
38. Pipe A, Sorensen M, Reid R. Physician smoking status, attitudes toward smoking, and cessation advice to patients: an international survey. Patient Educ Couns 2009 Jan;74(1):118-123. [doi: 10.1016/j.pec.2008.07.042] [Medline: 18774670]
39. Brinker TJ, Enk A, Gatzka M, Nakamura Y, Sondermann W, Omlor AJ, et al. A dermatologist's ammunition in the war against smoking: a photoaging app. J Med Internet Res 2017 Sep 21;19(9):e326 [FREE Full text] [doi: 10.2196/jmir.8743] [Medline: 28935619]
40. Zeiher J, Kuntz B, Lange C. Smoking among adults in Germany. J Health Monit 2017;2(2):2017. [doi: 10.17886/RKI-GBE-2017-043]
41. Frazier EL, Sutton MY, Brooks JT, Shouse RL, Weiser J. Trends in cigarette smoking among adults with HIV compared with the general adult population, United States - 2009-2014. Prev Med 2018 Jun;111:231-234. [doi: 10.1016/j.ypmed.2018.03.007] [Medline: 29550303]
42. Mdodo R, Frazier EL, Dube SR, Mattson CL, Sutton MY, Brooks JT, et al. Cigarette smoking prevalence among adults with HIV compared with the general adult population in the United States: cross-sectional surveys. Ann Intern Med 2015 Mar 03;162(5):335-344. [doi: 10.7326/M14-0954] [Medline: 25732274]
43. Reddy KP, Kong CY, Hyle EP, Baggett TP, Huang M, Parker RA, et al. Lung cancer mortality associated with smoking and smoking cessation among people living with HIV in the United States. JAMA Intern Med 2017 Nov 01;177(11):1613-1621. [doi: 10.1001/jamainternmed.2017.4349] [Medline: 28975270]
44. Westreich D, Cates J, Cohen M, Weber KM, Seidman D, Cropsey K, et al. Smoking, HIV, and risk of pregnancy loss. AIDS 2017 Dec 20;31(4):553-560 [FREE Full text] [doi: 10.1097/QAD.0000000000001342] [Medline: 27902507]
45. Wilson SM, Pacek LR, Dennis PA, Bastian LA, Beckham JC, Calhoun PS. Veterans living with HIV: a high-risk group for cigarette smoking. AIDS Behav 2017 Jul;21(7):1950-1955. [doi: 10.1007/s10461-017-1717-6] [Medline: 28194586]
46. Zawertailo L, Voci S, Selby P. The Cigarette Dependence Scale and Heaviness of Smoking Index as predictors of smoking cessation after 10weeks of nicotine replacement therapy and at 6-month follow-up. Addict Behav 2018 Mar;78:223-227. [doi: 10.1016/j.addbeh.2017.11.036] [Medline: 29220743]
47. Vidrine DJ, Kypriotakis G, Li L, Arduino RC, Fletcher FE, Tamí-Maury I, et al. Mediators of a smoking cessation intervention for persons living with HIV/AIDS. Drug Alcohol Depend 2015 Feb 01;147:76-80 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.12.003] [Medline: 25542824]
48. Ledgerwood DM, Yskes R. Smoking cessation for people living with HIV/AIDS: a literature review and synthesis. Nicotine Tob Res 2016 Dec;18(12):2177-2184 [FREE Full text] [doi: 10.1093/ntr/ntw126] [Medline: 27245237]
49. Shuter J, Morales DA, Considine-Dunn SE, An LC, Stanton CA. Feasibility and preliminary efficacy of a web-based smoking cessation intervention for HIV-infected smokers: a randomized controlled trial. J Acquir Immune Defic Syndr 2014 Sep 01;67(1):59-66 [FREE Full text] [doi: 10.1097/QAI.0000000000000226] [Medline: 25118794]
50. Aigner CJ, Gritz ER, Tamí-Maury I, Baum GP, Arduino RC, Vidrine DJ. The role of pain in quitting among human immunodeficiency virus (HIV)-positive smokers enrolled in a smoking cessation trial. Subst Abus 2017;38(3):249-252. [doi: 10.1080/08897077.2017.1291466] [Medline: 28328303]
51. Stanton CA, Papandonatos GD, Shuter J, Bicki A, Lloyd-Richardson EE, de Dios MA, et al. Outcomes of a tailored intervention for cigarette smoking cessation among Latinos living with HIV/AIDS. Nicotine Tob Res 2015 Aug;17(8):975-982 [FREE Full text] [doi: 10.1093/ntr/ntv014] [Medline: 26180222]
52. Vidrine DJ, Frank SG, Savin MJ, Waters AJ, Li Y, Chen S, et al. HIV care initiation: a teachable moment for Smoking Cessation? Nicotine Tob Res 2017 Sep 25. [doi: 10.1093/ntr/ntx218] [Medline: 29059424]
53. Vijayaraghavan M, Yuan P, Gregorich S, Lum P, Appelle N, Napoles AM, et al. Disparities in receipt of 5As for smoking cessation in diverse primary care and HIV clinics. Prev Med Rep 2017 Jun;6:80-87 [FREE Full text] [doi: 10.1016/j.pmedr.2017.02.012] [Medline: 28271025]

XSL•FO
RenderX

54. Diaz P, Ferketich A. Smoking and HIV: confronting the epidemic. Lancet HIV 2018 Dec;5(3):e109-e110. [doi: 10.1016/S2352-3018(18)30001-8] [Medline: 29329762]
55. Grabovac I, Brath H, Schalk H, Degen O, Dorner TE. Clinical setting-based smoking cessation programme and the quality of life in people living with HIV in Austria and Germany. Qual Life Res 2017 Dec;26(9):2387-2395 [FREE Full text] [doi: 10.1007/s11136-017-1580-y] [Medline: 28429240]
56. Helleberg M, Afzal S, Kronborg G, Larsen CS, Pedersen G, Pedersen C, et al. Mortality attributable to smoking among HIV-1-infected individuals: a nationwide, population-based cohort study. Clin Infect Dis 2013 Mar;56(5):727-734. [doi: 10.1093/cid/cis933] [Medline: 23254417]
57. Montano D, 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: Theory, Research, and Practice. Fifth edition. San Francisco, CA: Jossey-Bass; 2015:95-124.
58. Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50 years' observations on male British doctors. BMJ 2004 Jun 26;328(7455):1519 [FREE Full text] [doi: 10.1136/bmj.38142.554479.AE] [Medline: 15213107]
59. Karasouli E, Adams A. Assessing the evidence for e-resources for mental health self-management: a systematic literature review. JMIR Ment Health 2014;1(1):e3 [FREE Full text] [doi: 10.2196/mental.3708] [Medline: 26543903]
60. Carey M, Noble N, Mansfield E, Waller A, Henskens F, Sanson-Fisher R. The role of eHealth in optimizing preventive care in the primary care setting. J Med Internet Res 2015 May 22;17(5):e126 [FREE Full text] [doi: 10.2196/jmir.3817] [Medline: 26001983]
61. Forjuoh SN, Ory MG, Wang S, des Bordes JK, Hong Y. Using the iPod touch for patient health behavior assessment and health promotion in primary care. JMIR Mhealth Uhealth 2014;2(1):e14 [FREE Full text] [doi: 10.2196/mhealth.2927] [Medline: 25100308]
62. Foraker RE, Kite B, Kelley MM, Lai AM, Roth C, Lopetegui MA, et al. EHR-based visualization tool: adoption rates, satisfaction, and patient outcomes. EGEMS (Wash DC) 2015;3(2):1159 [FREE Full text] [doi: 10.13063/2327-9214.1159] [Medline: 26290891]
63. Rose GL, Ferraro TA, Skelly JM, Badger GJ, MacLean CD, Fazzino TL, et al. Feasibility of automated pre-screening for lifestyle and behavioral health risk factors in primary care. BMC Fam Pract 2015 Oct 23;16:150 [FREE Full text] [doi: 10.1186/s12875-015-0368-9] [Medline: 26497902]
64. Krist AH, Glasgow RE, Heurtin-Roberts S, Sabo RT, Roby DH, Gorin SNS, et al. The impact of behavioral and mental health risk assessments on goal setting in primary care. Transl Behav Med 2016 Jun;6(2):212-219 [FREE Full text] [doi: 10.1007/s13142-015-0384-2] [Medline: 27356991]
65. Elley CR, Dawes D, Dawes M, Price M, Draper H, Goodyear-Smith F. Screening for lifestyle and mental health risk factors in the waiting room: feasibility study of the Case-finding Health Assessment Tool. Can Fam Physician 2014 Nov;60(11):e527-e534 [FREE Full text] [Medline: 25551137]
66. Ferrari M, Ahmad F, Shakya Y, Ledwos C, McKenzie K. Computer-assisted client assessment survey for mental health: patient and health provider perspectives. BMC Health Serv Res 2016 Sep 23;16(1):516 [FREE Full text] [doi: 10.1186/s12913-016-1756-0] [Medline: 27663508]
67. Dempsey AF, Maertens J, Beaty B, O'Leary ST. Characteristics of users of a tailored, interactive website for parents and its impact on adolescent vaccination attitudes and uptake. BMC Res Notes 2015 Dec 01;8:739 [FREE Full text] [doi: 10.1186/s13104-015-1721-8] [Medline: 26625932]
68. Goodyear-Smith F, Corter A, Suh H. Electronic screening for lifestyle issues and mental health in youth: a community-based participatory research approach. BMC Med Inform Decis Mak 2016 Dec 08;16(1):140 [FREE Full text] [doi: 10.1186/s12911-016-0379-z] [Medline: 27821128]
69. Ahmad F, Lou W, Shakya Y, Ginsburg L, Ng PT, Rashid M, et al. Preconsult interactive computer-assisted client assessment survey for common mental disorders in a community health centre: a randomized controlled trial. CMAJ Open 2017;5(1):E190-E197 [FREE Full text] [doi: 10.9778/cmajo.20160118] [Medline: 28401134]
70. Cohen-Silver J, Laher N, Freeman S, Mistry N, Sgro M. Family fIRST, an interactive risk screening tool for families in a school-based pediatric clinic: a look at feasibility and pilot data. Clin Pediatr (Phila) 2017 Mar;56(3):217-225. [doi: 10.1177/0009922816657152] [Medline: 27380792]
71. Foucher-Urcuyo J, Longworth D, Roizen M, Hu B, Rothberg MB. Patient-entered wellness data and tailored electronic recommendations increase preventive care. J Am Board Fam Med 2017;30(3):350-361 [FREE Full text] [doi: 10.3122/jabfm.2017.03.160231] [Medline: 28484067]
72. Bruun LL, Soendergaard J, Halling A, Thilsing T, Thomsen JL. A novel approach to population-based risk stratification, comprising individualized lifestyle intervention in Danish general practice to prevent chronic diseases: results from a feasibility study. Health Informatics J 2017 Dec;23(4):249-259. [doi: 10.1177/1460458216645149] [Medline: 27245672]
73. Staeheli M, Aseltine RH, Schilling E, Anderson D, Gould B. Using mHealth technologies to improve the identification of behavioral health problems in urban primary care settings. SAGE Open Med 2017;5:2050312117712656 [FREE Full text] [doi: 10.1177/2050312117712656] [Medline: 28634539]

XSL•FO
RenderX

74. Goodyear-Smith F, Martel R, Darragh M, Warren J, Thabrew H, Clark TC. Screening for risky behaviour and mental health in young people: the YouthCHAT programme. Public Health Rev 2017;38:20 [FREE Full text] [doi: 10.1186/s40985-017-0068-1] [Medline: 29450092]
75. Thabrew H, Corter A, Goodyear-Smith F, Goldfinch M. Randomized trial comparing the electronic composite psychosocial screener YouthCHAT with a clinician-interview assessment for young people: a study protocol. JMIR Res Protoc 2017 Jul 31;6(7):e135 [FREE Full text] [doi: 10.2196/resprot.7995] [Medline: 28760727]
76. Wachtler C, Coe A, Davidson S, Fletcher S, Mendoza A, Sterling L, et al. Development of a mobile clinical prediction tool to estimate future depression severity and guide treatment in primary care: user-centered design. JMIR Mhealth Uhealth 2018 Apr 23;6(4):e95 [FREE Full text] [doi: 10.2196/mhealth.9502] [Medline: 29685864]
77. Goodyear-Smith F, Warren J, Bojic M, Chong A. eCHAT for lifestyle and mental health screening in primary care. Ann Fam Med 2013;11(5):460-466 [FREE Full text] [doi: 10.1370/afm.1512] [Medline: 24019278]

Abbreviations

HSI: Heaviness of Smoking Index

Edited by G Eysenbach; submitted 04.05.18; peer-reviewed by N Walker, P Keller; comments to author 09.06.18; revised version received 22.06.18; accepted 10.07.18; published 15.08.18.

Please cite as:
Brinker TJ, Brieske CM, Esser S, Klode J, Mons U, Batra A, Rüther T, Seeger W, Enk AH, von Kalle C, Berking C, Heppt MV, Gatzka MV, Bernardes-Souza B, Schlenk RF, Schadendorf D
A Face-Aging App for Smoking Cessation in a Waiting Room Setting: Pilot Study in an HIV Outpatient Clinic
J Med Internet Res 2018;20(8):e10976
URL: http://www.jmir.org/2018/8/e10976/
doi:10.2196/10976
PMID:30111525

©Titus Josef Brinker, Christian Martin Brieske, Stefan Esser, Joachim Klode, Ute Mons, Anil Batra, Tobias Rüther, Werner Seeger, Alexander H Enk, Christof von Kalle, Carola Berking, Markus V Heppt, Martina V Gatzka, Breno Bernardes-Souza, Richard F Schlenk, Dirk Schadendorf. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 15.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Church-Based Weight Loss Intervention in African American Adults using Text Messages (LEAN Study): Cluster Randomized Controlled Trial

Robert L Newton Jr[1], PhD; Leah A Carter[1], MS; William Johnson[1], PhD; Dachuan Zhang[1], MApStat; Sandra Larrivee[1], MS; Betty M Kennedy[1], PhD; Melissa Harris[1], MS; Daniel S Hsia[1], MD

Pennington Biomedical Research Center, Baton Rouge, LA, United States

Corresponding Author:
Robert L Newton Jr, PhD
Pennington Biomedical Research Center
6400 Perkins Road
Baton Rouge, LA, 70808
United States
Phone: 1 225 763 3034
Fax: 1 225 763 0905
Email: Robert.Newton@pbrc.edu

Abstract

Background: African American adults experience a high prevalence of obesity and its associated comorbidities, including diabetes. Church-based interventions have been shown to be effective in decreasing weight in this population. mHealth interventions can address two needs for obesity treatment in this community, including enhancing weight loss and providing wide dissemination.

Objective: This study aimed to assess the feasibility and efficacy of a church-based weight loss intervention that incorporates mHealth technology.

Methods: In this study, 8 churches (n=97) were randomly assigned to the intervention or delayed intervention condition (control group). We recruited participants through their respective church. Volunteer church members were trained by study staff to deliver the 10-session, 6-month intervention. Participants in the intervention group attended group sessions and received automated short message service (SMS) text messages designed to reinforce behavioral strategies. Conversely, participants in the delayed intervention condition received SMS text messages related to health conditions relevant for African American adults. We obtained measures of body composition, blood pressure, blood glucose, and cholesterol.

Results: We successfully recruited 97 African American adults, with a mean age of 56.0 (SE 10.3) years and a mean body mass index of 38.6 (SE 6.4) kg/m2 (89/97, 91.8% females), who attended the churches that were randomized to the intervention (n=68) or control (n=29) condition. Of these, 74.2% (72/97) of the participants (47/68, 69.1% intervention; 25/29, 86.2% delayed intervention) completed the 6-month assessment. The average intervention group attendance was 55%. There was a significant difference in weight loss (*P*=.04) between participants in the intervention (–1.5 (SE 0.5) kg) and control (0.11 (SE 0.6) kg) groups. Among participants in the intervention group, the correlation between the number of SMS text messages sent and the percent body fat loss was r=.3 with *P*=.04. The participants reported high satisfaction with the automated SMS text messages.

Conclusions: Automated SMS text messages were well-received by participants, suggesting that more enhanced mHealth technologies are a viable option for interventions targeting African American adults.

Trial Registration: ClinicalTrials.gov NCT02863887; https://clinicaltrials.gov/ct2/show/NCT02863887 (Archived by WebCite at http://www.webcitation.org/71JiYzizO)

(J Med Internet Res 2018;20(8):e256) doi:10.2196/jmir.9816

KEYWORDS

African Americans; behavioral strategies; community health; mHealth; mobile phone; obesity; text messages; weight loss

XSL•FO
RenderX

Introduction

Obesity represents a health inequity for African American adults [1,2]. African American adults with obesity are at greater risk for developing chronic diseases such as diabetes. A major risk factor for developing obesity and its comorbidities is an unhealthy lifestyle comprising a lack of regular physical activity and poor nutrition [3,4]. Intensive behavioral programs have resulted in improved dietary intake, increased physical activity, and weight loss [5-7]. However, weight loss results for African American adults have been less than those for other ethnic groups [8]. Therefore, innovative strategies addressing the obesity epidemic in African American adults are warranted.

One novel strategy could be the utilization of mHealth technology (ie, mobile phones, wearables, etc). Mobile phones are an emerging intervention tool because they are compact, portable, normally "on," readily available to individuals, affordable, user-friendly, offer advanced functionality, and allow Web-based access 24 hours a day for 7 days per week [9-11]. mHealth technologies can affect the amount of treatment delivered by increasing communication between study staff and participants, providing behavioral strategies, and promoting behavioral change [12]. Several reviews [13-16] have shown that mHealth interventions can be effective in promoting weight loss; however, few of these studies have included substantial numbers of African American adults despite this ethnic group's well-documented high prevalence of obesity.

mHealth interventions utilizing mobile phones may be particularly well suited for African American adults. Recent estimates have shown that 94% of African American adults own a mobile phone [17], 72% own a smartphone [17], and nearly 70% utilize mobile phones to access health information [18]. Short message service (SMS) text messaging is one form of mHealth, and African American adults report that SMS text messaging is an appropriate form of intervention delivery [19,20]. This realization has led to the development of SMS text message-based pilot and feasibility studies targeting various behaviors such as the medication adherence, HIV prevention, healthy eating, and physical activity [21-25]. Although these studies demonstrate the acceptability, feasibility, and potential effectiveness of SMS text message interventions designed for African American adults, few have utilized randomized controlled designs [22], and fewer were specifically designed to promote weight loss [26]. Therefore, the purpose of the Lifestyle Changes Through Exercise and Nutrition (LEAN) study was to test the feasibility (SMS text message tolerance and satisfaction) and effectiveness (compared with a control group) of utilizing SMS text messages to promote weight loss in African American adults enrolled in a church-based weight loss program.

Methods

Design

The study utilized a cluster-randomized trial design (NCT02863887). Originally, 9 churches agreed to participate and were randomized into the study. While 5 churches were randomized to the intervention group, 4 churches were randomized to the control group. Notably, the churches were matched by their membership size between the conditions so that the potential participant sample sizes would be similar.

Participants

Individuals were considered eligible for this study if they (1) were aged 18-75 years, (2) self-identified as African American, (3) were a member of a participating church, (4) had a body mass index (BMI) of >30 kg/m^2, (5) owned a mobile phone with SMS text messaging capabilities, (6) were at risk for diabetes (ie, had a history of gestational diabetes or at least one diagnosed nuclear family member) or had been diagnosed with type 2 diabetes, (7) were willing to change their diet and physical activity to promote weight loss, (8) were able to participate in face-to-face counseling sessions as scheduled, and (9) had low to moderate cardiovascular disease risk. Participants were paid up to US $60 for completing the study.

Procedures

Churches were recruited utilizing a community outreach coordinator who attended pastoral fellowship meetings and held an initial meeting with the pastor or health ministry leader. Then, the participating churches were randomized to either the intervention or control group. Churches were randomized to the intervention or control group using a 1:1 ratio; two churches with the largest and smallest memberships (n=8000 and 75) were clustered together. The study statistician used a computerized pseudorandom number generator to determine the randomization order in advance. The study coordinator revealed the randomization to the churches prior to the start of the study. Following randomization, participants were recruited utilizing standard strategies, including presentations during church service, small groups, and health fairs; flyers placed on informational boards, tables, etc; advertisements placed in the church bulletins; and utilizing the community health coaches from each church who volunteered to deliver the intervention (refer to the Community Health Coaches section below). Next, interested individuals were phone-screened to determine their initial eligibility. Those eligible were invited to the baseline visit. During the baseline visit, potential participants were provided with information related to the purpose, setting, and duration of the program; the nature of the behavioral change program; information on the number, duration, timing, and content of the automated SMS text messages; participant and community health coach expectations; and the assessments. Written informed consent was obtained from each potential participant before they completed their baseline assessments. Baseline assessments and community health coach training were completed approximately 1 week prior to the start of the intervention. This study protocol was reviewed and approved by the Pennington Biomedical Research Center's Institutional Review Board.

Study Groups

Intervention

Community Health Coaches

Each church was required to select 2 volunteers from their congregation to lead the group sessions. Although these

XSL•FO
RenderX

individuals are often referred to as "lay health providers," the volunteers in the LEAN study collectively decided to be referred to as "community health coaches." All community health coaches were women who were employed as nurses, pharmacists, or health education or physical education teachers. Community health coaches received an 8-hour training conducted by the study staff on how to deliver a behaviorally based, empirically proven, lifestyle change and weight loss intervention (ie, Diabetes Prevention Program). Several topics were covered during the training, including leading effective groups, behavior modification, physical activity demonstrations, and a detailed review of each session's content.

Group Sessions

At each designated church, 10 group sessions (Textbox 1) were conducted over a 6-month period. The session frequency was weekly (n=4) for the first month, bimonthly for 2 months (n=4), and monthly for 2 months (n=2). Attendance and participant weight were recorded at each session by the community health coach.

Short Message Service Text Messages

Participants received 5 automated SMS text messages each week pertaining to lesson content (n=2), motivation (n=1), and behavioral prompts (n=2). Lesson content SMS text messages contained information derived from the lesson materials (ie, behavioral strategies, information, reminders, etc). Motivational SMS text messages contained verbiage that encouraged and reinforced participant behavioral change. Behavioral prompt SMS text messages were designed to remind participants to engage in physical activity, check weight daily, and consume healthy meals. A library of SMS text messages was created for each component, and each participant received a random selection of these messages. These SMS text messages were adapted from a previous physical activity intervention conducted by the primary author [27] and were based on Social Cognitive Theory. The text message content did not change during the study.

Control

Participants randomized to the delayed intervention were encouraged to maintain their normal eating and exercise habits for 6 months. During this 6-month period, they received 2-3 automated SMS text messages per week, which contained information on health topics specific to African American adults, including stroke prevention, lupus, cardiovascular disease, etc.

Measures

Participants attended a baseline and 6-month visit and were required to start fasting at 10 pm the evening before each visit.

Outcome Measures

Anthropometrics

The participants' height and weight were measured with them wearing clothes, not wearing shoes and socks, and after removing heavy pocket items. The height was measured to the nearest 0.1 cm using a wall-mounted stadiometer (Holtain Ltd., Crymych, Dyfed, United Kingdom). The weight was measured to the nearest 0.1 kg using the portable Tanita Body Composition Analyzer (SC-240).

Body Fat

The percent body fat was estimated using the portable Tanita Body Composition Analyzer (SC-240) [28]. Participants stood on the analyzer with their bare feet placed on the electrodes. Measurements were recorded to the nearest 0.1%.

Vital Signs

Blood pressure and pulse were measured using an automatic blood pressure cuff (Omron, Model BP710). Participants rested for approximately 5 minutes prior to the measurement.

Blood Glucose or Cholesterol

Fingerstick blood sampling (40 µL) was collected using the Alere Cholestech LDX Analyzer. Following the successful completion of quality control procedures, whole blood was collected from a fingerstick in a lithium heparin-coated capillary tube.

Food Intake

We used the National Cancer Institute (NCI) fat screener [29] to estimate the percentage of energy from fat over the past 12 months. The standard 7-item fruit and vegetable screener developed by the NCI and National 5 a Day Program [30,31] assessed how often fruit and vegetables were consumed in the past month.

Textbox 1. Session topics.

Session 1: Obesity and chronic disease, goal setting, and self-monitoring
Session 2: Portion control, energy balance, and physical activity intensity levels
Session 3: Reading food labels, estimating calorie needs, and calorie expenditure
Session 4: Diabetes, carbohydrates, physical activity and diabetes
Session 5: Proteins, fats, physical activity and diabetes
Session 6: Cues for eating and activity behavior, and eating out
Session 7: Social support and meal planning for healthy eating
Session 8: Emotional eating and stress management
Session 9: Problem solving
Session 10: Relapse and prevention of relapse

XSL•FO
RenderX

Physical Activity

The International Physical Activity Questionnaire (IPAQ) [32] was used to assess self-reported physical activity. The metabolic equivalent of task (MET)-minutes/week of physical activity in each domain of leisure time activity (sedentary, light, moderate, and vigorous intensity) was calculated. The IPAQ has been shown to be a reliable and valid measure of physical activity and has been sensitive to changes in physical activity resulting from intervention programs [33].

Quality of Life

We used the 31-item Impact of Weight on Quality of Life Questionnaire-lite (IWQOL-lite) [34] to assess quality of life. The IWQOL-lite measures 5 domains of functioning, physical function, self-esteem, sexual life, public distress, and work. High scores indicate poor functioning.

Feasibility Measures

Text Message Tolerance

SMS text messages were deemed feasible to deliver as an intervention if <25% of participants requested a cessation of SMS text messages.

Satisfaction Questionnaire

Each participant's satisfaction with the different elements of the study, including their community health coach, group sessions, and SMS text messages, was measured at the 6-month period. The questionnaire contains 15 items and is rated along a scale of Agree (1) to Disagree (5), with lower scores indicating higher levels of satisfaction. Satisfaction scores ≤2.0 ("Agree" or "Slightly Agree") indicated the feasibility of the intervention component. The questionnaire was adapted from another satisfaction questionnaire developed by the primary author [35].

Data Entry

All data from the measurements were entered into REDCap [36], which is a secure, Web-based application designed to support data capture for research studies, providing (1) an intuitive interface for validated data entry, (2) audit trails for tracking data manipulation and export procedures, (3) automated export procedures for seamless data downloads to common statistical packages, and (4) procedures for importing data from external sources. It improves data entry through real-time validation rules (with the automated data type and range checks) and provides easy data manipulation (with audit trails for reporting, monitoring, and querying patient records). Pennington Biomedical has a license agreement with Vanderbilt University, and the database and software are installed on our local servers and hosted by Pennington Biomedical, and not by Vanderbilt. The Vanderbilt University Office of Research was not used as a central location for data processing and management.

Treatment Fidelity

In this study, 2 sessions from each church were audiorecorded to allow a study staff member to monitor the intervention delivery. While the first recorded session was selected from the first 5 group meetings, the second recorded session was selected from the final 5 group meetings. The audiorecordings were reviewed and compared with a checklist of topics to be covered for each session. The topics included the introduction and closing of sessions (6 items), engaging participants (2 items), session content (7 items), and the physical activity component (5 items). Items were scored Yes (1) or No (2) for the greeting, engaging, and physical activity items, and was scored Poorly (1), Adequately (2), and Extremely Well (3) for delivery of the session content.

Statistical Analysis

The study was a repeated measures cluster-randomized trial. Participants were nested within churches, and churches were nested within study conditions; thus, outcome data were analyzed accordingly. We used linear mixed models, with churches and participants having random effects and interventions having fixed effects, to compare the mean differences in outcome variables at baseline between the two intervention groups. The primary outcome variable was weight loss after 6 months. A second linear mixed effects model with churches having random effects and baseline outcome used as a covariate was used to analyze the outcome change between baseline and 6 months. The estimation of model parameters used restricted maximum likelihood procedures using all available data. Each secondary outcome (eg, waist circumference and questionnaires) was analyzed similarly. Multiple imputations were used to replace missing data. For each analytical mixed model, both completer's analysis and intent-to-treat analysis were conducted to assess the sensitivity of statistical findings to our choice of statistical approach. In each analytic model, we performed a two-step procedure to test the following hypotheses: (1) churches do not have a significant effect on the outcome variable; and (2) no difference is observed in the effects of the intervention and delayed intervention. If (1) is significant, the degrees of freedom for the F test used in (2) depends on the number of churches, in addition to the number of participants in the churches in the intervention and control groups. If (1) is not significant, then the degrees of freedom for the F test used in (2) depends solely on the total number of participants. In the first case, the churches play an important role in determining the power of the test for hypothesis (1); whereas, in the latter case, the power is determined by the total number of participants. The churches vary in active membership, and although attempts were made to balance the number of participants between intervention and control groups, inequities persist between the groups because of large differences in church sizes and a small number of churches. The differential variability in the weight loss showed no evidence of a component of variability being attributable to churches, suggesting that randomization of churches effectively randomized participants to study conditions. Hence, it is feasible to assume that, except for receiving different conditions, participants in the two groups, although unbalanced, are representative samples from a common underlining population. Therefore, the validity of inferences pertaining to weight loss differences in the intervention and control groups are justified. All results are reported based on the completer's analysis, and the results are consistent, in terms of significance, between completer's analysis and intent-to-treat analysis, unless otherwise noted. Furthermore, Spearman correlations were calculated to assess the association between

XSL•FO
RenderX

the SMS text message use and weight loss parameters. All data were analyzed utilizing SAS Version 9.4 (SAS Institute Inc., Cary, NC, USA).

Power Analysis

The power analysis was based on recruiting 9 churches in total, with 5 churches randomized to the intervention group and 4 to the control group. The primary outcome variable was weight loss after 6 months, and the anticipated minimum difference in the average weight loss was 2 kg (~2% of the expected baseline weight). The anticipated SD for the weight loss was not expected to exceed 3.5 kg. With an alpha level set at.05, the study was designed to have at least 80% power to detect at least a 2% difference between interventions with a sample size of 12 participants in each church (total N=108).

There were 72 participants (25 from control group, 47 from intervention group) who completed the study. Fortunately, the power of the study was not compromised because the churches proved to have no differential effect on the weight loss between the study groups (see Statistical Analysis section above for the rationale). The poststudy mean weight loss was 1.6 (SD 3.1) kg, and the percent weight loss was 1.5% (SD 2.8%). These poststudy findings are more favorable than estimates used in determining power and sample size. Consequently, the poststudy power was 79%. Furthermore, secondary outcomes were viewed as possibly providing exploratory findings and therefore were not considered in power or sample size calculations.

Results

Demographics

Churches were recruited from February to May 2016. Initially, 9 churches were randomized. As 1 church in the delayed intervention condition was unresponsive to attempts to designate community health coaches or set a baseline date, it was subsequently dropped from the study. Therefore, 8 churches participated in this study, and the randomization resulted in 5 intervention churches and 3 delayed intervention churches. Of these, 2 churches were Catholic, 4 were Baptist, and 2 were Christian nondenominational. The churches varied in size from 75 to ~8000 members.

Participants were recruited from May to September 2016. We successfully recruited 97 African American adults, with a mean age of 56.0 (SE 10.3) years and a mean BMI of 38.6 (SE 6.4) kg/m^2 (89/97, 91.8% females; 52.6% having a college or postgraduate degree), who attended churches that were randomized to the intervention (n=68) or delayed intervention (n=29) condition. In addition, 47.4% (46/97) of the sample indicated prediabetes according to the Cholestech fingerstick analysis (glucose, 100-125 mg/dL). No significant differences were noted between the study groups on baseline variables (*P*>.09; Tables 1 and 2). Notably, 74.2% (72/97) of these individuals (47/68, 69.1% intervention; 25/29, 86.2% delayed intervention) completed the 6-month assessment (Figure 1).

Short Message Service Text Messages

Participants in both groups reported receiving SMS text messages. Although participants were not required to respond to the automated SMS text messages, 41 participants sent at least 1 text message (30 from intervention, 11 from delayed intervention). While participants in the intervention churches sent 84 SMS text messages, those in the control churches sent 25 SMS text messages, totaling 109 SMS text messages sent during the study. In addition, 24 individuals sent 80 messages that were related to the study (eg, attendance, asking for assistance, appreciation for the text, etc). However, 11 participants sent 22 messages that were determined to be unrelated to the study or were ambiguous (eg, voting, unintelligible symbols, and sleep). Furthermore, 6 individuals (4 from control, 2 from intervention) sent 7 messages requesting that the SMS text messages be stopped; this level of request (6/97, 6.1%) met our <25% feasibility criteria. Among participants in the intervention group, the correlation between the number of SMS text messages sent and change in percent body fat was statistically significant (*r*=.3 with *P*=.04), whereas the correlation between the number of SMS text messages sent and change in weight loss was not statistically significant.

Attendance

Participants in the intervention group attended a median of 6 sessions. The average attendance rate was 55%, with attendance decreasing over the course of the intervention (*P*<.001). While session 1 had the highest attendance (70.6%), session 10 had the lowest (27.9%).

Body Composition

A significant between-group difference was noted in weight loss (*P*=.03, effect size=0.55). Participants in the intervention lost –1.4 (SE 0.4) kg, whereas those in the control group gained 0.2 (SE 0.6) kg. No difference was observed in the percent body fat as measured by the Tanita scale (*P*=.30, effect size=0.5). While participants in the intervention lost –0.3% (SE 0.9%), those in the control group gained 1.2% (SE 1.2%; Table 3).

Cardiometabolic Outcomes

We observed no significant between-group differences in the systolic or diastolic blood pressure, glucose, or cholesterol levels (*P*>.356).

Quality of Life

We observed significant between-group differences on the IWQOL-lite. The total score changed significantly (*P*=.006) between the intervention and delayed intervention groups. Significant differences were observed in the physical function (*P*=.04), self-esteem (*P*=.03), and public distress (*P*=.03) subscales. A marginally significant (*P*=.08) difference was noted in the sexual life subscale. In every situation, changes indicated that quality of life improved in the intervention group and diminished in the delayed intervention group.

Physical Activity

No significant between-group differences were observed in the change in MET-minutes/week (*P*=.99) or total time sitting (*P*=.76).

XSL•FO
RenderX

Fruit and Vegetable Intake

We observed no significant between-group differences in the change in fruit and vegetable (P=.23) or percent fat intake (P=.41).

Table 1. Participants' characteristics at baseline.

Characteristics	Control (n=29)	Intervention (n=68)	Total (N=97)
Age in years, mean (SD)	58.6 (8.7)	54.9 (10.7)	56.0 (10.3)
Gender, n (%)			
Female	25 (86)	64 (94)	89 (92)
Male	4 (14)	4 (6)	8 (8)
Race, n (%)			
African American	29 (100)	68 (100)	97 (100)
Marital status, n (%)			
Single	5 (17)	24 (35)	29 (30)
Married	16 (55)	30 (44)	46 (47)
Divorced	6 (21)	11 (16)	17 (18)
Widowed	2 (7)	1 (2)	3 (3)
Education, n (%)			
Some High School	0 (0)	2 (3)	2 (2)
High School Diploma/General Equivalency Diploma	5 (17)	16 (24)	21 (22)
1-3 years college	5 (17)	17 (25)	22 (23)
College degree	7 (24)	15 (22)	22 (23)
Post graduate degree	12 (41)	17 (25)	29 (30)
Income, n (%)			
<US $50,000	11 (38)	35 (51)	46 (47)
US $50,000- US $100,000	7 (24)	21 (31)	28 (29)
>US $100,000	7 (24)	7 (11)	14 (14)
Did not answer	4 (14)	5 (7)	9 (9)
Weight (kg), mean (SD)	101.6 (16.8)	103.8 (18.0)	103.2 (17.5)
Body mass index (kg/m^2), mean (SD)	37.7 (5.6)	38.9 (6.7)	38.6 (6.4)
Body percent fat (%), mean (SD)	47.1 (6.1)	49.0 (4.7)	48.5 (5.2)
Systolic blood pressure (mmHg), mean (SD)	127.7 (14.1)	126.4 (14.3)	126.8 (14.2)
Diastolic blood pressure (mmHg), mean (SD)	75.9 (9.2)	79.9 (9.1)	78.7 (9.3)
Glucose (mg/dL), mean (SD)	112.7 (23.6)	105.2 (22.4)	107.5 (22.9)
Suspected prediabetes, n (%)	13 (50)	32 (51)	45 (51)
Total cholesterol (mg/dL), mean (SD)	177.1 (41.5)	177.7 (34.5)	177.5 (36.5)

Medication

In this study cohort, 24 participants were taking medications to control their diabetes. One participant started taking medications after beginning the program and another decreased their dose.

Adverse Events

No adverse events occurred during this study.

Satisfaction Questionnaire

In this study, >80% of participants either "agreed" or "slightly agreed" with the satisfaction questions. The average satisfaction item score for the group sessions, SMS text messages, and the community health coach was 1.4 (0.53), 1.4 (0.67), and 1.4 (0.72), respectively, indicating high levels of satisfaction. These average scores of 1.4 met our feasibility criteria of <2.0.

Treatment Fidelity

A random session from each half of the intervention was audiotaped. The average score for session content was 2.0, which is equivalent to an "adequate" rating. We conducted 40% of the actual physical activity sessions and implemented 28% of the physical activity components during the physical activity sessions. Furthermore, the community health coaches introduced

XSL•FO
RenderX

and closed sessions 80% of the time and engaged participants 100% of the time.

Table 2. Questionnaire data at baseline.

Baseline variables	Control, mean (SE)	Intervention, mean (SE)
Quality of Life[a]		
Physical Function Scale	43.0 (1.6)	42.5 (1.0)
Self-Esteem Scale	25.3 (1.3)	25.7 (0.9)
Sexual Life Scale[b]	17.2 (0.8)	16.3 (0.6)
Public Distress Scale[b]	23.0 (0.8)	22.6 (0.4)
Work Scale[b]	17.8 (0.7)	17.7 (0.4)
Total score	125.1 (4.3)	124.0 (2.8)
Physical activity[c]		
Sitting total weekday	326.7 (51.1)	380.0 (31.9)
Metabolic equivalent of task-min/week[b]	899.5 (249.5)	1325.3 (294.4)
Food intake		
Fruits and veggies[b]	4.0 (0.5)	4.8 (0.4)

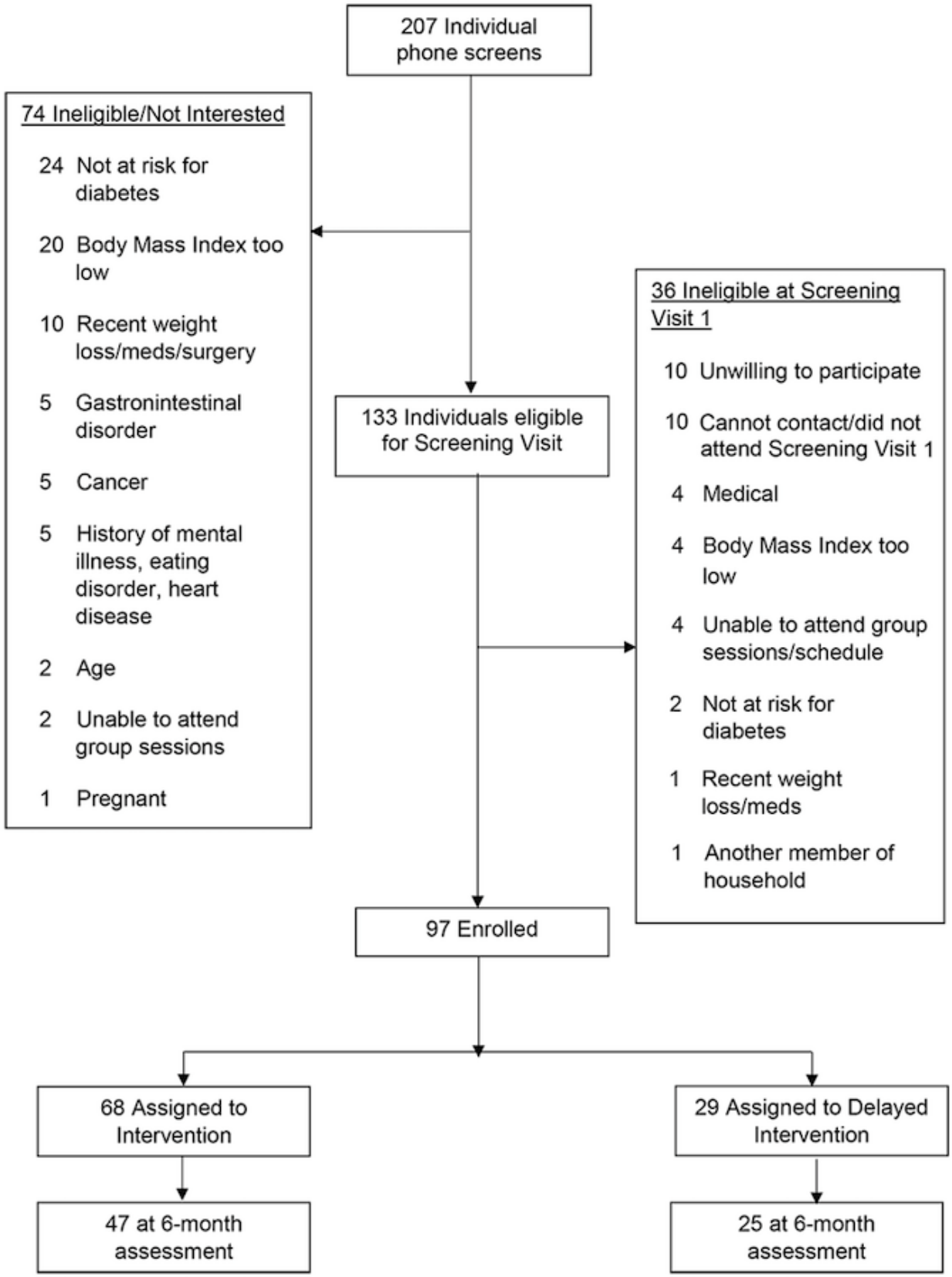
[a]Control (n=18-28), intervention (n=50-64); sample sizes differed between the different measures.

[b]Wilcoxon test, mean (SE).

[c]Control (n=27-29), intervention (n=64-68); sample sizes differed between the different measures.

XSL•FO
RenderX

Figure 1. Study flowchart.



XSL•FO
RenderX

Table 3. Six-month change scores in physiological variables.

Physiological variables	Completers analyses			Intent-to-treat analyses		
	Control (n=25), mean (SE)	Intervention (n=47), mean (SE)	*P* value	Control (n=29), mean (SE)	Intervention (n=68), mean (SE)	*P* value
Anthropometric or physiological						
Percent weight loss	0.2 (0.6)	–1.4 (0.4)	.03	0.3 (0.7)	–1.6 (0.5)	.04
Percent body fat	1.2 (1.2)	–0.3 (0.9)	.30	1.2 (1.1)	–0.4 (0.8)	.26
Systolic	–0.3 (3.1)	0.3 (2.3)	.89	–0.4 (3.9)	0.2 (2.9)	.90
Diastolic	0.4 (1.8)	2.4 (1.3)	.36	0.3 (2.4)	2.3 (1.8)	.51
Glucose	–2.2 (3.9)	–6.5 (2.9)	.38	–1.8 (4.7)	–4.6 (3.4)	.64
Cholesterol	–5.9 (6.2)	–2.4 (4.6)	.66	–6.7 (7.2)	–2.8 (5.3)	.67
Quality of Life						
Physical Function	1.8 (1.2)	–1.4 (0.9)	.04	1.5 (1.3)	–1.8 (0.9)	.03
Self-Esteem	–0.6 (1.0)	–3.4 (0.7)	.03	–0.4 (1.2)	–3.5 (0.8)	.03
Sexual Life	0.4 (0.7)	–1.1 (0.5)	.08	0.8 (0.9)	–1.0 (0.6)	.11
Public Distress	0.2 (0.4)	–0.8 (0.3)	.03	0.2 (0.4)	–0.8 (0.3)	.04
Work	–0.2 (0.4)	–0.3 (0.3)	.80	–0.03 (0.5)	–0.6 (0.4)	.38
Total	1.8 (2.5)	–6.9 (1.8)	.01	2.9 (3.1)	–7.4 (2.4)	.01
Physical activity						
Sitting total weekday	5.7 (50.2)	–13.1 (33.7)	.76	–1.7 (41.1)	–12.5 (28.9)	.82
Metabolic equivalent of task-min/week	586.4 (503.3)	587.2 (362.8)	.99	373.5 (492.9)	295.8 (335.5)	.90
Food intake						
Fruits and veggies	0.8 (0.7)	–0.3 (0.5)	.23	0.8 (0.7)	–0.2 (0.5)	.27
National Cancer Institute percent fat	1.4 (0.6)	2.1 (0.4)	.41	1.67 (0.6)	2.0 (0.5)	.63

Discussion

Principal Findings

The LEAN study showed that church-affiliated African American participants were receptive to receiving SMS text messages related to behavioral strategies as part of a weight loss intervention. Participants engaged with SMS text messages, indicated satisfaction with the SMS text messages, and few requested the messages to be stopped, demonstrating the feasibility of the intervention. This church-based study resulted in statistically significant weight loss and improved quality of life for participants in the intervention group compared with the control group. Importantly, engagement with SMS text messages was associated with body fat loss. However, other variables assessed, including self-reported physical activity and dietary intake, and clinical outcomes, were unchanged.

The unique component of the LEAN study was the use of SMS text messages in a randomized controlled trial targeting African American adults. Three recent studies of African American adults have utilized SMS text messages to assist in the initiation of behavioral change in a church-based setting [25,26,37]; the SMS text messaging system in these studies included personalization, support, and delivering recipes and meal plans. They also showed within-group changes in weight loss or physical activity outcomes across 12 weeks. The LEAN study builds upon these investigations by utilizing a randomized controlled trial design and intervening over a 6-month period. The use of SMS text messages in the LEAN study did not appear to result in greater weight loss compared with most other church-based studies targeting African American adults [38]. However, our satisfaction data showed that SMS text messages were well-received, similar to recent studies [25,26,37]. Of note, the majority of LEAN participants responded to SMS text messages even though they were not required to do so, and higher text message response was associated with higher percent body fat loss. It is difficult to explain the correlation between text message response and body fat loss in the absence of a significant correlation with BMI. Although BMI and percent body fat are highly correlated, these do not measure the same component of adiposity. Participants responding to SMS text messages might have made behavioral changes that affected bodily tissues (eg, adipose tissue size and water) that are estimated through body impedance but not BMI. Alternatively, this could also be a spurious finding, and therefore, inferential interpretation should be made with caution until this finding can be replicated. Nonetheless, the LEAN study demonstrates that it is feasible and potentially efficacious to incorporate a form of mHealth (SMS text messages) into a translational behavioral change study targeting African American adults. It is suggested that future health disparities research targeting African American adults should explore the use of other

XSL•FO
RenderX

mHealth technologies, including apps, social networking, and wearables [39,40]. These mHealth technologies could provide objective assessments of physical activity, diet, and weight; immediate feedback on participant behavior; automated or personalized messages (eg, from the church-leadership or lay health provider); and more frequent and effective communication between lay health providers and participants, which may enhance efficacy. Future work should also utilize innovative methodologies such as optimization, microrandomized trials, and adaptive designs [41-43].

Comparison With Prior Work

The LEAN study is one of the numerous church-based translational weight loss programs targeting African American adults. Lancaster et al [38] recently conducted a systematic review of these programs showing that the majority of studies reported statistically significant weight loss, similar to the LEAN results. The LEAN study incorporated elements common among the studies reviewed, including enrolling overweight or obese African American adults, embedding religious principles within the intervention, and utilizing lay health advisors (known as "community health coaches" in the LEAN study) to deliver the intervention. However, few studies achieved clinically significant weight loss of >3% [26,38,44], including the LEAN study. Our treatment fidelity data indicated that of the subset of sessions observed, the physical activity component was infrequently implemented, and although the actual weight loss content was delivered adequately, there was room for improvement. Enhancing the community health coach training or increasing contact with the community health coaches during the implementation of the project through booster sessions might improve the program implementation adherence. Another contributor to the small weight loss might be the lack of attendance. Greater weight loss in the LEAN study could have been obtained through a combination of enhanced implementation and greater attendance.

Few church-based weight loss interventions have assessed quality of life [26,45]. Some similar church-based health promotion interventions involving individuals with overweight or obesity have measured overall quality of life but did not measure the impact of participants' weight on their quality of life [26,44]; this is an important aspect for individuals with overweight or obesity. Kennedy et al [45] utilized the IWQOL scale but did not find changes in their church-based weight loss intervention. The LEAN findings suggest that the intervention decreased the negative impact that weight had on the participants' quality of life. As quality of life is an important outcome from a patient-centered perspective, it is necessary to continue to investigate the effect of church-based weight loss studies on this outcome.

Strengths and Limitations

There are several strengths of the LEAN study. The LEAN study is one of the few church-based programs to incorporate text messaging, utilize a cluster-randomized trial design, and include a delayed intervention control group. It was delivered in 8 different churches, included assessment of treatment fidelity and was adequately powered to detect changes in the primary outcome variable. In addition, the LEAN study included a number of measures beyond weight loss, including an estimated percent body fat, cardiometabolic outcomes, impact of weight on quality of life, and treatment satisfaction. However, this study also has several limitations. First, the sample sizes were unbalanced between the intervention and control groups, which could be attributed to the fact that we were able to recruit ~14 participants per church in the intervention condition and only ~9 per church in the control condition. Nevertheless, sufficient power remained to detect significant differences. In addition, the LEAN study provides important findings related to the feasibility of mHealth interventions in this practical setting, and efficacy can be confirmed in a larger clinical trial. Second, the LEAN study was only 6 months in duration. However, the duration was twice as long as similar studies [25,26,37] and was sufficient to determine the study feasibility. Third, in terms of mHealth technology, automated, one-way SMS text messages are rather rudimentary. The fact that SMS text messages were not personalized, or that SMS text messages did not provide feedback on participants' behavior, might have contributed to the small weight loss findings. However, one of the purposes of this study was to determine whether automated SMS text messages were feasible to deliver to this population, and we were able to achieve this aim. In addition, even these simple SMS text messages were associated with percent body fat loss. Fourth, while the LEAN study comprised participants with some demographic diversity (education, income, marital status, etc), it will be important to show that the study can be effective in different regions of the United States and with a greater proportion of male participants. Finally, we utilized self-report measures of diet and physical activity.

Conclusions

Overall, the LEAN study demonstrated that mHealth technology (ie, SMS text messages) is feasible to deliver and well-received by African American adults in a church-based setting. The study also showed that a church-based intervention resulted in significant weight loss and improvements in quality of life. A key to future investigations will be to determine whether more advanced mHealth technologies (eg, mobile apps, activity monitors, and automated scales) can be implemented successfully, without increased burden on the lay health provider or participants. Furthermore, it will be important to determine whether these technologies can increase weight loss to clinically significant levels and to show that mHealth technology can be utilized over an extended period.

Acknowledgments

We would like to thank Belfair Baptist Church, Greater Mount Carmel Baptist Church, Living Faith Christian Center, New Light Missionary Baptist Church, Oasis Christian Church, St. Francis Xavier Catholic Church, St. Paul Catholic Church, and Star Hill

XSL•FO
RenderX

Baptist Church for participating in the program. Emily Foret (study coordinator), Sheletta Donatto (Registered Dietician), Patti Boyd (study support), and the Intervention Resources assessment team are appreciated for their diligent work on this project. A special thank you goes to all of the community health coaches who volunteered their time, energy, and resources to the program and participants. This project was funded by an award from the Humana Foundation (24423711), and this work was also supported, in part, by 2 U54 GM104940 from the National Institute of General Medical Sciences of the National Institutes of Health, which funds the Louisiana Clinical and Translational Science Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors have no financial disclosures. The sponsors did not play a role in the study design, analysis, administration, or write-up of the study.

Authors' Contributions

RLN Jr designed the study. LAC and MH collected the data. WJ, DZ and SL performed the data analysis. RLN Jr, WJ, and DSH interpreted the data. RLN Jr, WJ, DZ, and SL drafted the manuscript. LAC performed the literature search. RLN Jr, WJ, BMK, MH, and DSH performed the critical review of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 1MB - jmir_v20i8e256_app1.pdf]

References

1. Ogden CL, Fakhouri TH, Carroll MD, Hales CM, Fryar CD, Li X, et al. Prevalence of Obesity Among Adults, by Household Income and Education - United States, 2011-2014. MMWR Morb Mortal Wkly Rep 2017 Dec 22;66(50):1369-1373 [FREE Full text] [doi: 10.15585/mmwr.mm6650a1] [Medline: 29267260]
2. Menke A, Casagrande S, Geiss L, Cowie CC. Prevalence of and Trends in Diabetes Among Adults in the United States, 1988-2012. JAMA 2015 Sep 08;314(10):1021-1029. [doi: 10.1001/jama.2015.10029] [Medline: 26348752]
3. Sieverdes JC, Sui X, Lee D, Church TS, McClain A, Hand GA, et al. Physical activity, cardiorespiratory fitness and the incidence of type 2 diabetes in a prospective study of men. Br J Sports Med 2010 Mar;44(4):238-244. [doi: 10.1136/bjsm.2009.062117] [Medline: 19656767]
4. Lee I, Djoussé L, Sesso HD, Wang L, Buring JE. Physical activity and weight gain prevention. JAMA 2010 Mar 24;303(12):1173-1179 [FREE Full text] [doi: 10.1001/jama.2010.312] [Medline: 20332403]
5. Love R, Adams J, van Sluijs EMF, Foster C, Humphreys D. A cumulative meta-analysis of the effects of individual physical activity interventions targeting healthy adults. Obes Rev 2018 Apr 27;19(8):1164-1172. [doi: 10.1111/obr.12690] [Medline: 29701299]
6. Møller G, Andersen HK, Snorgaard O. A systematic review and meta-analysis of nutrition therapy compared with dietary advice in patients with type 2 diabetes. Am J Clin Nutr 2017 Dec;106(6):1394-1400. [doi: 10.3945/ajcn.116.139626] [Medline: 29092883]
7. Haw JS, Galaviz KI, Straus AN, Kowalski AJ, Magee MJ, Weber MB, et al. Long-term Sustainability of Diabetes Prevention Approaches: A Systematic Review and Meta-analysis of Randomized Clinical Trials. JAMA Intern Med 2017 Dec 01;177(12):1808-1817. [doi: 10.1001/jamainternmed.2017.6040] [Medline: 29114778]
8. Davis KK, Tate DF, Lang W, Neiberg RH, Polzien K, Rickman AD, et al. Racial Differences in Weight Loss Among Adults in a Behavioral Weight Loss Intervention: Role of Diet and Physical Activity. J Phys Act Health 2015 Dec;12(12):1558-1566. [doi: 10.1123/jpah.2014-0243] [Medline: 25742122]
9. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. Epidemiol Rev 2010;32:56-69 [FREE Full text] [doi: 10.1093/epirev/mxq004] [Medline: 20354039]
10. Burner ER, Menchine MD, Kubicek K, Robles M, Arora S. Perceptions of successful cues to action and opportunities to augment behavioral triggers in diabetes self-management: qualitative analysis of a mobile intervention for low-income Latinos with diabetes. J Med Internet Res 2014;16(1):e25 [FREE Full text] [doi: 10.2196/jmir.2881] [Medline: 24476784]
11. Lyzwinski LN, Caffery LJ, Bambling M, Edirippulige S. Consumer perspectives on mHealth for weight loss: a review of qualitative studies. J Telemed Telecare 2017 Jan 01:1357633X17692722. [doi: 10.1177/1357633X17692722] [Medline: 28181859]
12. McCarroll R, Eyles H, Ni MC. Effectiveness of mobile health (mHealth) interventions for promoting healthy eating in adults: A systematic review. Prev Med 2017 Dec;105:156-168. [doi: 10.1016/j.ypmed.2017.08.022] [Medline: 28882743]
13. Fukuoka Y, Gay CL, Joiner KL, Vittinghoff E. A Novel Diabetes Prevention Intervention Using a Mobile App: A Randomized Controlled Trial With Overweight Adults at Risk. Am J Prev Med 2015 Aug;49(2):223-237 [FREE Full text] [doi: 10.1016/j.amepre.2015.01.003] [Medline: 26033349]

XSL•FO
RenderX

14. Flores MG, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile Phone Apps to Promote Weight Loss and Increase Physical Activity: A Systematic Review and Meta-Analysis. J Med Internet Res 2015;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
15. Schippers M, Adam PC, Smolenski DJ, Wong HT, de Wit JB. A meta-analysis of overall effects of weight loss interventions delivered via mobile phones and effect size differences according to delivery mode, personal contact, and intervention intensity and duration. Obes Rev 2017 Dec;18(4):450-459. [doi: 10.1111/obr.12492] [Medline: 28187246]
16. Semper HM, Povey R, Clark-Carter D. A systematic review of the effectiveness of smartphone applications that encourage dietary self-regulatory strategies for weight loss in overweight and obese adults. Obes Rev 2016 Sep;17(9):895-906. [doi: 10.1111/obr.12428] [Medline: 27192162]
17. Pew Research Center. Mobile Fact Sheet URL: http://www.pewinternet.org/fact-sheet/mobile/ [WebCite Cache ID 6vqvx7Efi]
18. Anderson M. Pew Research Center. Racial and ethnic differences in how people use mobile technology URL: http://www.pewresearch.org/fact-tank/2015/04/30/racial-and-ethnic-differences-in-how-people-use-mobile-technology/ [accessed 2017-12-20] [WebCite Cache ID 6vqwFKr6v]
19. Bonar EE, Cunningham RM, Collins RL, Cranford JA, Chermack ST, Zimmerman MA, et al. Feasibility and Acceptability of Text Messaging to Assess Daily Substance Use and Sexual Behaviors among Urban Emerging Adults. Addict Res Theory 2018;26(2):103-113. [doi: 10.1080/16066359.2017.1310205] [Medline: 29632458]
20. Callender C, Thompson D. Text Messaging Based Obesity Prevention Program for Parents of Pre-Adolescent African American Girls. Children (Basel) 2017 Dec 04;4(12):- [FREE Full text] [doi: 10.3390/children4120105] [Medline: 29207536]
21. Reese JM, Joseph RP, Cherrington A, Allison J, Kim Y, Spear B, et al. Development of Participant-Informed Text Messages to Promote Physical Activity Among African American Women Attending College: A Qualitative Mixed-Methods Inquiry. J Transcult Nurs 2017 Dec;28(3):236-242 [FREE Full text] [doi: 10.1177/1043659616644959] [Medline: 27093904]
22. Kolmodin MK, Naar S, Gibson-Scipio W, Lam P, Secord E. The Detroit Young Adult Asthma Project: Pilot of a Technology-Based Medication Adherence Intervention for African-American Emerging Adults. J Adolesc Health 2016 Oct;59(4):465-471 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.05.016] [Medline: 27475032]
23. Rodgers RF, Franko DL, Shiyko M, Intille S, Wilson K, O'Carroll D, et al. Exploring healthy eating among ethnic minority students using mobile technology: Feasibility and adherence. Health Informatics J 2016 Sep;22(3):440-450. [doi: 10.1177/1460458214565950] [Medline: 25609082]
24. Evans C, Turner K, Suggs LS, Occa A, Juma A, Blake H. Developing a mHealth intervention to promote uptake of HIV testing among African communities in the UK: a qualitative study. BMC Public Health 2016 Dec 28;16:656 [FREE Full text] [doi: 10.1186/s12889-016-3278-4] [Medline: 27465586]
25. McCoy P, Leggett S, Bhuiyan A, Brown D, Frye P, Williams B. Text Messaging: An Intervention to Increase Physical Activity among African American Participants in a Faith-Based, Competitive Weight Loss Program. Int J Environ Res Public Health 2017 Dec 29;14(4):- [FREE Full text] [doi: 10.3390/ijerph14040326] [Medline: 28353650]
26. Vaughn NA, Brown D, Reyes BO, Wyatt C, Arnold KT, Dalianis E, et al. A 40-Day Journey to Better Health: Utilizing the DanielFast to Improve Health Outcomes in Urban Church-Based Settings. Healthcare (Basel) 2018 Mar 05;6(1):E25 [FREE Full text] [doi: 10.3390/healthcare6010025] [Medline: 29510546]
27. Newton RL, Marker AM, Allen HR, Machtmes R, Han H, Johnson WD, et al. Parent-targeted mobile phone intervention to increase physical activity in sedentary children: randomized pilot trial. JMIR Mhealth Uhealth 2014 Nov 10;2(4):e48 [FREE Full text] [doi: 10.2196/mhealth.3420] [Medline: 25386899]
28. Barreira TV, Staiano AE, Katzmarzyk PT. Validity assessment of a portable bioimpedance scale to estimate body fat percentage in white and African-American children and adolescents. Pediatr Obes 2013 Apr;8(2):e29-e32 [FREE Full text] [doi: 10.1111/j.2047-6310.2012.00122.x] [Medline: 23239610]
29. Thompson FE, Midthune D, Subar AF, Kipnis V, Kahle LL, Schatzkin A. Development and evaluation of a short instrument to estimate usual dietary intake of percentage energy from fat. J Am Diet Assoc 2007 May;107(5):760-767. [doi: 10.1016/j.jada.2007.02.006] [Medline: 17467371]
30. Thompson FE, Kipnis V, Subar AF, Krebs-Smith SM, Kahle LL, Midthune D, et al. Evaluation of 2 brief instruments and a food-frequency questionnaire to estimate daily number of servings of fruit and vegetables. Am J Clin Nutr 2000 Jun;71(6):1503-1510 [FREE Full text] [Medline: 10837291]
31. Havas S, Heimendinger J, Damron D, Nicklas TA, Cowan A, Beresford SA, et al. 5 A Day for better health--nine community research projects to increase fruit and vegetable consumption. Public Health Rep 1995;110(1):68-79 [FREE Full text] [Medline: 7838947]
32. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
33. Frederix I, Hansen D, Coninx K, Vandervoort P, Vandijck D, Hens N, et al. Medium-Term Effectiveness of a Comprehensive Internet-Based and Patient-Specific Telerehabilitation Program With Text Messaging Support for Cardiac Patients: Randomized Controlled Trial. J Med Internet Res 2015;17(7):e185 [FREE Full text] [doi: 10.2196/jmir.4799] [Medline: 26206311]

XSL•FO
RenderX

34. Kolotkin RL, Head S, Brookhart A. Construct validity of the Impact of Weight on Quality of Life Questionnaire. Obes Res 1997 Sep;5(5):434-441. [Medline: 9385618]
35. Newton RL, Perri MG. A randomized pilot trial of exercise promotion in sedentary African-American adults. Ethn Dis 2004;14(4):548-557. [Medline: 15724775]
36. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
37. Tkatch R, Musich S, Draklellis J, Hetzel M, Banks J, Dugan J, et al. Claim More™: Empowering African American Women to Make Healthy Choices. J Holist Nurs 2018 Mar;36(1):91-100 [FREE Full text] [doi: 10.1177/0898010117691167] [Medline: 29172987]
38. Lancaster KJ, Carter-Edwards L, Grilo S, Shen C, Schoenthaler AM. Obesity interventions in African American faith-based organizations: a systematic review. Obes Rev 2014 Oct;15 Suppl 4:159-176. [doi: 10.1111/obr.12207] [Medline: 25196412]
39. Turner-McGrievy GM, Wilcox S, Boutté A, Hutto BE, Singletary C, Muth ER, et al. The Dietary Intervention to Enhance Tracking with Mobile Devices (DIET Mobile) Study: A 6-Month Randomized Weight Loss Trial. Obesity (Silver Spring) 2017 Dec;25(8):1336-1342 [FREE Full text] [doi: 10.1002/oby.21889] [Medline: 28600833]
40. Spring B, Pellegrini CA, Pfammatter A, Duncan JM, Pictor A, McFadden HG, et al. Effects of an abbreviated obesity intervention supported by mobile technology: The ENGAGED randomized clinical trial. Obesity (Silver Spring) 2017 Jul;25(7):1191-1198 [FREE Full text] [doi: 10.1002/oby.21842] [Medline: 28494136]
41. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. Ann Behav Med 2016 Sep 23:446-462. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
42. Collins LM, Kugler KC, Gwadz MV. Optimization of Multicomponent Behavioral and Biobehavioral Interventions for the Prevention and Treatment of HIV/AIDS. AIDS Behav 2016 Jan;20 Suppl 1:S197-S214 [FREE Full text] [doi: 10.1007/s10461-015-1145-4] [Medline: 26238037]
43. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. Health Psychol 2015 Dec;34 Suppl:1220-1228. [doi: 10.1037/hea0000305] [Medline: 26651463]
44. Sattin RW, Williams LB, Dias J, Garvin JT, Marion L, Joshua TV, et al. Community Trial of a Faith-Based Lifestyle Intervention to Prevent Diabetes Among African-Americans. J Community Health 2016 Feb;41(1):87-96 [FREE Full text] [doi: 10.1007/s10900-015-0071-8] [Medline: 26215167]
45. Kennedy BM, Paeratakul S, Champagne CM, Ryan DH, Harsha DW, McGee B, Lower Mississipi Delta Nutrition Intervention Research Initiative. A pilot church-based weight loss program for African-American adults using church members as health educators: a comparison of individual and group intervention. Ethn Dis 2005;15(3):373-378. [Medline: 16108295]

Abbreviations

BMI: body mass index
IPAQ: International Physical Activity Questionnaire
IWQOL-lite: Impact of Weight on Quality of Life Questionnaire-lite
LEAN: Lifestyle Changes Through Exercise and Nutrition
MET: metabolic equivalent of task
NCI: National Cancer Institute
SMS: short message service

Edited by G Eysenbach; submitted 25.01.18; peer-reviewed by C Schnelle, L Song, F Modave; comments to author 08.03.18; revised version received 17.05.18; accepted 18.06.18; published 24.08.18.

Please cite as:
Newton Jr RL, Carter LA, Johnson W, Zhang D, Larrivee S, Kennedy BM, Harris M, Hsia DS
A Church-Based Weight Loss Intervention in African American Adults using Text Messages (LEAN Study): Cluster Randomized Controlled Trial
J Med Internet Res 2018;20(8):e256
URL: http://www.jmir.org/2018/8/e256/
doi:10.2196/jmir.9816
PMID:30143478

XSL•FO
RenderX

©Robert L Newton Jr, Leah A Carter, William Johnson, Dachuan Zhang, Sandra Larrivee, Betty M Kennedy, Melissa Harris, Daniel S Hsia. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 24.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Social Media Use in Interventions for Diabetes: Rapid Evidence-Based Review

Elia Gabarron[1], PhD; Eirik Årsand[1,2], PhD; Rolf Wynn[2,3], MD, PhD

[1]Norwegian Centre for E-health Research, University Hospital of North Norway, Tromso, Norway

[2]Department of Clinical Medicine, Faculty of Health Sciences, The Arctic University of Norway, Tromso, Norway

[3]Division of Mental Health and Addictions, University Hospital of North Norway, Tromso, Norway

Corresponding Author:
Elia Gabarron, PhD
Norwegian Centre for E-health Research
University Hospital of North Norway
Forskningsparken i Breivika 3 etg Sykehusvn 23
Tromso, 9019
Norway
Phone: 47 77754030
Fax: 47 77754099
Email: elia.gabarron@ehealthresearch.no

Abstract

Background: Health authorities recommend educating diabetic patients and their families and initiating measures aimed at improving self-management, promoting a positive behavior change, and reducing the risk of complications. Social media could provide valid channel to intervene in and deliver diabetes education. However, it is not well known whether the use of these channels in such interventions can help improve the patients' outcomes.

Objective: The objective of our study was to review and describe the current existing evidence on the use of social media in interventions targeting people affected with diabetes.

Methods: A search was conducted across 4 databases (PubMed, Scopus, EMBASE, and Cochrane Library).The quality of the evidence of the included primary studies was graded according to the Grading of Recommendations Assessment, Development and Evaluation criteria, and the risk of bias of systematic reviews was assessed by drawing on AMSTAR (A MeaSurement Tool to Assess systematic Reviews) guidelines. The outcomes reported by these studies were extracted and analyzed.

Results: We included 20 moderate- and high-quality studies in the review: 17 primary studies and 3 systematic reviews. Of the 16 publications evaluating the effect on glycated hemoglobin (HbA_{1c}) of the interventions using social media, 13 reported significant reductions in HbA_{1c} values. The 5 studies that measured satisfaction with the interventions using social media found positive effects. We found mixed evidence regarding the effect of interventions using social media on health-related quality of life (2 publications found positive effects and 3 found no differences) and on diabetes knowledge or empowerment (2 studies reported improvements and 2 reported no significant changes).

Conclusions: There is very little good-quality evidence on the use of social media in interventions aimed at helping people with diabetes. However, the use of these channels is mostly linked to benefits on patients' outcomes. Public health institutions, clinicians, and other stakeholders who aim at improving the knowledge of diabetic patients could consider the use of social media in their interventions.

(J Med Internet Res 2018;20(8):e10303) doi:10.2196/10303

KEYWORDS

social media; social networking; health promotion; intervention studies; diabetes

Introduction

The prevalence of diabetes has been growing worldwide for the last few decades [1], and it has become one of the four priority noncommunicable diseases targeted by world leaders, together with cardiovascular disease, cancer, and chronic respiratory disease [1]. All types of diabetes can lead to complications, reduce the quality of life, and increase the risk of premature

XSL•FO
RenderX

death [1,2]. To support clinical practice, health authorities recommend educating diabetic patients and their families and initiating prevention measures aimed at improving self-management and promoting a positive behavior change, thereby reducing the risk of complications [1-7].

The use of social media has increased dramatically in the recent years [8], and social media channels could be effective in supporting clinical practice and delivering education to improve self-management and to promote a positive behavior change among people affected with chronic diseases [9]. However, it is not well known whether the use of these channels in interventions can help improve diabetic patients' outcomes, and the evidence of using social media in interventions for people with diabetes needs to be updated.

Evidence on positive effects of social media interventions on health behavior-related outcomes (ie, weight loss and physical activity) exists in 2 meta-analyses focusing on several health conditions [10,11]. However, 2 other meta-analyses have reported mixed results regarding the use of social media in health interventions [12,13]. Furthermore, a third meta-analysis concluded that using social media did not contribute to reducing risk factors in patients with noncommunicable diseases [14].

Norway is one of the most connected countries in the world, and most of the Norwegian population uses social media [15,16]. Due to its ubiquity, a health promotion intervention using social media, aimed at people affected by diabetes and their relatives, is now being initiated by our research team [17]. An updated status on the evidence that exists regarding the use and usefulness of social media in diabetes is essential. Hence, the objective of this paper was to review and describe the current evidence on the use of social media in interventions targeting people with diabetes.

Methods

A Rapid Review

We performed a rapid review to quickly capture the current evidence on the use of social media in interventions on diabetes. We had two research questions: (1) Is there evidence on the use of social media in interventions aimed at improving, maintaining, or promoting health among people affected with diabetes? and (2) What are the reported outcomes, for example, the effects on clinical parameters, effects on behavior, or other effects?

The rapid review method was chosen as it typically provides similar conclusions as systematic reviews, and it allows to quickly and efficiently access the current evidence on the topic [18-21]. In this rapid review, we followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [22] and the MeaSurement Tool to Assess systematic Reviews (AMSTAR) [23,24] guidelines. This review has been registered in PROSPERO (registration number: CRD42018088206).

Search Strategy

To answer the research questions, we performed an electronic search on February 13, 2018. It covered published studies comprising the terms "Social media," "Social networking," "Facebook," "Twitter," or "YouTube" in combination with the term "Diabetes" included in the title or abstract and indexed in the following databases: PubMed (Medical Subject Heading terms and text word), Scopus, EMBASE, and the Cochrane Library. The search strategy was limited to studies published in English. The full search strategy is summarized in Multimedia Appendix 1.

Inclusion and Exclusion Criteria

Publications were included in the review if they (1) focused on diabetes or involved participants affected by diabetes; (2) described interventions aimed at improving, maintaining, or promoting health; (3) reported results from the intervention; and (4) used social media in the intervention. Both primary studies and reviews were considered to be of interest and were, therefore, included in this review. Papers that did not meet all four criteria were excluded from the review.

Eligibility and Data Extraction

All references captured by the search engine were uploaded to EndNote X7 (Clarivate Analytics; Philadelphia, PA, USA). Duplicates were identified and removed. To assess the eligibility of the papers, two passes were done. In the first pass, all titles and abstracts were examined by one reviewer (EG). In the second pass, the full text of the studies selected on the first pass was extracted and carefully analyzed to confirm their eligibility. When it was unclear whether the studies should be included, they were discussed and agreed with a second reviewer (EÅ). The agreed upon studies were included in the quality assessment. A single reviewer (EG) extracted the data from the included studies. The following data were extracted: interventions (duration and participants), social media use (channels, use as main tool for the intervention or as supporting tool), and outcomes (effects on clinical parameters, on behavior, or other effects).

Quality Evidence Assessment and Risk of Bias

The quality of evidence and risk of bias of the studies included in this review were classified by one reviewer (EG). The quality of evidence of primary studies was assessed following the Grading of Recommendations Assessment, Development and Evaluation guidelines [25]. A second reviewer (RW) verified the assigned quality of a random sample of primary studies. The risk of systematic bias was assessed by drawing on the AMSTAR criteria [19,23,24].

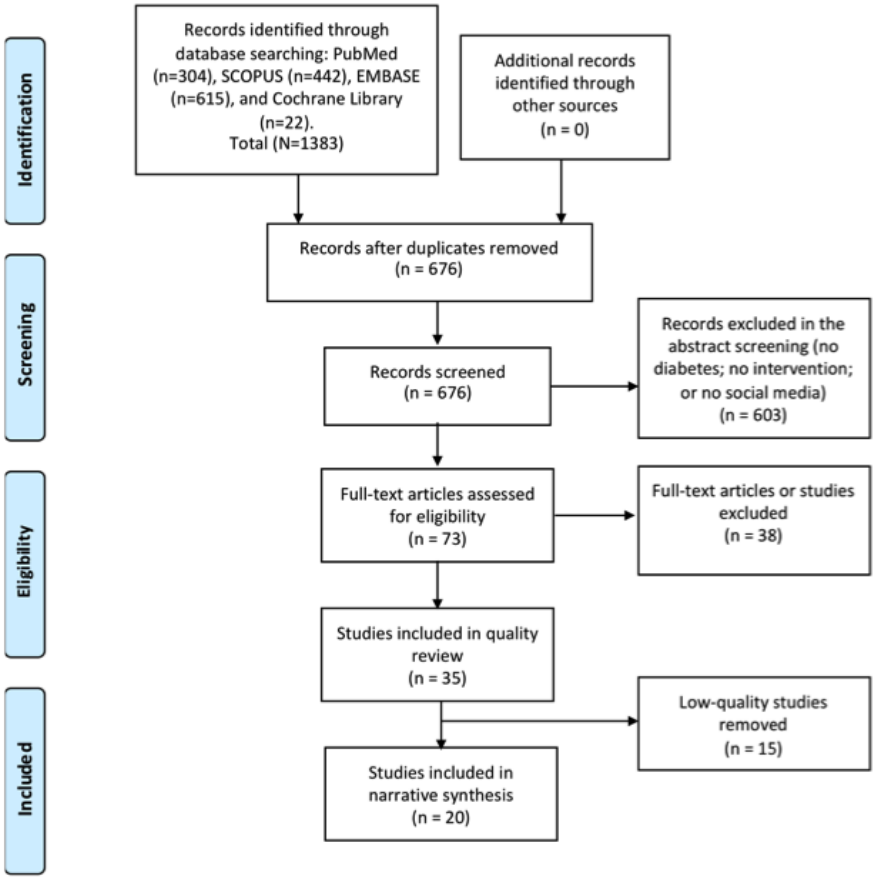
Results

Sample

A total of 1383 publications were identified, and after removing duplicates, 676 titles and abstracts were screened. The full search strategy and its results are summarized in Multimedia Appendix 1. The list of all potentially relevant studies that were read in full-text form but were excluded from the review can be found in Multimedia Appendix 2. Of these publications, 35 met the inclusion criteria [26-60]; of them, 32 were primary studies [26-32,34-38,40-59] and 3 were systematic reviews [33,39,60] (Figure 1).

XSL•FO
RenderX

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



Quality of The Evidence and Risk of Bias Assessment

Only 1 of the 35 included studies was considered to be of high quality [31]; 19 studies were considered to be of moderate quality: the 3 systematic reviews [33,39,60,61] (Multimedia Appendix 3), all 15 randomized controlled trials (RCTs) [26-30,32,34,35,42,44,45,48,54,55,58], and 1 nonrandomized intervention [40]. The remaining 15 nonrandomized studies were weighted as being of low or very low quality, and therefore, they were removed from the narrative synthesis (these studies are listed in Multimedia Appendix 4). Hence, 20 studies (1 of high quality and 19 of moderate quality) were included in this review. Multimedia Appendix 5 summarizes these 20 studies.

The PRISMA checklist of this study can be found in Multimedia Appendix 6.

Evidence: Clinical Effects

Of the 20 included studies, 19 reported clinical outcomes and only 1 study did not refer to any clinical effect [55].

Glycated Hemoglobin

The main reported clinical outcome was glycated hemoglobin (HbA_{1c}). Eighteen publications evaluated HbA_{1c} values and reported heterogeneous effects. On one hand, 4 publications reported significant improvements in HbA_{1c} values, favoring the groups that used social media comparison with control groups. This effect has been reported in 2 systematic reviews summarizing the evidence from studies focusing on both type one diabetes (T1D) and type two diabetes (T2D) and in 2 RCTs focusing specifically on T1D [27,33,39,44]. In one of these systematic reviews two analyses were performed: one of them was a meta-analysis including RCTs only; in the second analysis, nonrandomized studies were also included. In the latter, a significant mean reduction in HbA_{1c} values was found, favoring the social media group (0.49%, 95% CI −0.64 to −0.34, I^2=86%) [39]. On the other hand, 13 RCTs reported significant improvements in HbA_{1c} values among all the study participants, independently of whether they were allocated to the group where social media was used or not. These studies mostly targeted young people affected with T1D [26,28,30,32,34,35,45,54,60]. Furthermore, another study reported mixed results [58]. The data analysis of all the participants included in this study (208 adults with T2D) did not show significant decreases in HbA_{1c} values. However, in a second analysis, it was found that patients with HbA_{1c}≥10% at baseline had a significant decrease at 6 months [58]. Two additional studies focusing on young people with T1D did not find any differences in HbA_{1c} values [31,42].

Blood Pressure

Effects on blood pressure were reported only by 2 systematic reviews, both finding improvements associated with social media use. One of the systematic reviews referred to 5 studies with a total of 2580 patients (1317 in the intervention groups and 1263 in the control groups), where there was a significant mean difference in systolic blood pressure (3.47 mmHg, 95% CI 5.01 to 1.94, P<.001, I^2=0%) and diastolic blood pressure

XSL•FO
RenderX

(1.84 mmHg, 95% CI 2.98 to 0.70, *P*=.002, I^2=29%), favoring the intervention groups using social media [39]. The second review also found reductions in blood pressure associated with Web 2.0 participation, reported in 2 studies [33].

Other Reported Clinical Outcomes

Additional evidence on clinical effects has been reported for triglycerides and severe hypoglycemia episodes. Only 1 systematic review referred to the effect on triglycerides. This review reported the effect drawing on 10 studies with a total of 989 patients. A significant reduction of 11.05% (95% CI: 20.92 to 1.18, *P*=.03, I^2=0) was found among the study participants where social media had been used [39]. An RCT carrying out a 1-year intervention did not find any differences in severe hypoglycemia episodes for any of the study participants [31].

Evidence: Effects on Behavior

Of the 20 included studies, 16 referred to different effects on behavior.

Satisfaction With the Intervention

Five publications had comparable findings related to patients' high satisfaction with the interventions where social media were used or to internet visits being preferred by patients [26,28,30,39,60].

Health-Related Quality of Life

Five studies reported on this item, reaching different conclusions. Two studies, an RCT and a systematic review, reported improvements in health-related quality of life (HRQoL) among T1D and T2D patients linked to social media use [33,44]. Three RCTs with young T1D patients did not find any differences regarding HRQoL between the groups after the intervention [27,31,48].

Diabetes Knowledge and Diabetes Empowerment

This effect was reported in 4 studies and with mixed conclusions. One study involving adolescents with T1D found significantly increased diabetes knowledge on treatment or testing in the intervention group at 4, 8, and 12 months [42]. In another study on young people with T1D, it was found that social media use allowed patients to gain knowledge and information about diabetes and interact when making their daily insulin adjustments [45]. On the other hand, 2 studies, both referring to a 1-year intervention carried out in Sweden with T1D patients, showed no improvement in diabetes empowerment [27,31].

Medication Adherence

Two publications reported on treatment adherence. No differences were found regarding adherence in a systematic review [33], while the social media intervention group of an RCT had significantly better medication adherence (*P*=.01) [29].

Healthier Life-Styles

There were mixed results on this outcome as well. A systematic review described reductions in dietary fat intake, but the same review also found no effects on physical activity [33]. However, a recently published RCT found a significant improvement in self-reported physical activity (est 0.28, *P*=.046) for those with the highest engagement in the site and a decreased sedentary time (Fitbit data) for the intervention group compared with that for a control group (est −12.17, *P*=.048) [58].

Other Reported Behavioral Outcomes

Additional reported effects on behavior were an increase in diabetes communication (*P*<.05) and medical self-efficacy (*P*<.01) [48], reductions in depressive symptoms [33], and no significant differences in perceived quality of care [31]. Table 1 summarizes the evidence identified in this review on clinical and behavioral outcomes of interventions using social media and addressing people with diabetes.

Targeted Population

Most (14/20) of the publications focused only on young people affected with T1D, with different age groups ranging from 0 to 23 years (as specified in the studies) [26-32,34,35,42,44,45,48,54]. Five publications referred to both T1D and T2D patients, and therefore, participants had a broader age range [33,39,40,55,60]. Only one study specifically targeted adults diagnosed with T2D [58]. The included primary studies were performed in Macedonia [26,28,30,32,34,35,44,45,54], Sweden [27,31], the United States [48,55], and Ireland [42], and the location was unspecified in 3 of the publications [29,40,58].

Social Media Use

The main use of social media was as a supporting tool for the main intervention (14/20), primarily used for reinforcing regular visits and with the purpose of engaging patients in treatment and improving self-management and diabetes control. In these cases, the chosen social media channels were Facebook (group), Facebook (chat), Skype, specific social networking sites, or social media in general [26-28,30-32,34,35,40,44,45,48,55,58]. Of these studies, 7 (all of them belonging to the same research group) reported improvements in HbA_{1c} values for all study participants [26,28,30,32,34,35,45]. Two studies found improvements in HbA_{1c} values only in the participants allocated to the social media groups [27,44]. Two studies did not find any differences in HbA_{1c} values when social media was used as a supporting tool [31,48]. Moreover, one study found mixed results, with no improvements in HbA_{1c} values for the whole sample but improvements for the participants in whom the values were higher at baseline [58]. Two publications did not report on HbA_{1c} values linked to the use of social media as a supporting tool [40,55].

XSL•FO
RenderX

Table 1. Summary of the evidence on reported outcomes (n=20).

Outcomes	Supported by number of publications			
	Significant positive effects	Mixed results	No significant differences	Outcome not reported
Clinical effects				
HbA_{1c}[b]	13[a] [26-28,30,32-35,39,44,45,54,60]	1 [58]	4 [29,31,42,48]	2 [40,55]
Blood pressure	2 [33,39]	0	0	18 [26-32,34,35,40,42,44,45,48,54,55,58,60]
Triglycerides	1 [39]	0	0	19 [26-35,40,42,44,45,48,54,55,58,60]
Severe hypoglycemia	0	0	1 [31]	19 [26-30,32-35,39,40,42,44,45,48,54,55,58,60]
Effects on behavior				
HRQoL[c]	2 [33,44]	0	3 [27,31,48]	15 [26,28-30,32,34,35,39,40,42,44,54,55,58,60]
Knowledge or empowerment	2 [42,45]	0	2 [27,31]	16 [26,28-30,32-35,39,40,44,48,54,55,58,60]
Medication adherence	1 [29]	0	1 [33]	18 [26-28,30-32,34,35,39,40,42,44,45,48,54,55,58,60]
Healthier self-reported lifestyles	1 [58]	1 [33]	0	18 [26-32,34,35,39,40,42,44,45,48,54,55,60]
Self-efficacy	1 [40]	1 [48]	0	18 [26-35,39,42,44,45,54, 55,58,60]
Depressive symptoms	1 [33]	0	0	19 [26-32,34,35,39,40,42,44,45,48,54,55,58,60]
Perceived quality of care	0	0	1 [31]	19 [26-30,32-35,39,40,42,44,45,48,54,55,58,60]

[a]13 studies reported improvements in HbA_{1c} values in all study participants; 4 of these studies reported improvements only in the intervention groups (comparison with control groups) [27,33,39,44].

[b]HbA_{1c}: Glycated hemoglobin.

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

Three RCTs studied social media as the main channel for delivering the intervention. These 3 studies used peers in educational and behavioral interventions aimed at young people affected with T1D. The purpose of these interventions was to increase diabetes knowledge and to improve clinical outcomes. These 3 studies used Facebook closed groups and Viber [29,42,54]. Two of them did not find any significant differences in HbA_{1c} values [29,42], while the third study reported improvements in HbA_{1c} values among all the participants [54]. The use of social media as a main channel for delivering the intervention or as a supporting tool was not clearly stated in any of the 3 systematic reviews [33,39,60].

Discussion

Summary of the Evidence

A rapid review method was used to quickly capture the current evidence on the use of social media in interventions on diabetes. Following a search in 4 databases, only 20 publications considered of adequate quality were included in this review: 3 systematic reviews and 17 primary studies.

The research evidence shows that the most commonly reported outcome in intervention studies using social media is HbA_{1c}, followed by satisfaction with the intervention, HRQoL, and diabetes knowledge or empowerment. Most of the intervention studies using social media that evaluated HbA_{1c} values reported significant improvements (13/16 publications) [26-28,30,32-35,39,44,45,54,60]. Four of these publications, 2 systematic reviews and 2 RCTs, reported improvements only in intervention groups compared with that in control groups [27,33,39,44]. However, due to a heterogeneity in the methods that were used in the studies, including differences in the characteristics of participants, sample sizes, and study lengths, comparing them is difficult.

The 5 studies that measured satisfaction with the interventions where social media were used unanimously reported positive effects [26,28,30,39,60]. Two publications reported positive effects on HRQoL [33,44], and 2 others found improvements in knowledge [42,45], while 3 publications did not report any significant differences in HRQoL [27,31,48] or any improvements in knowledge [27,31].

Should We Use Social Media in Interventions for People With Diabetes?

Health authorities have recommended educating diabetic patients and their families with the aim of improving self-management,

XSL•FO
RenderX

promoting a positive behavior change, and reducing the risk of complications [1-7]. Although the use of social media has not been linked to clear improvements in one meta-analysis focusing on patients with noncommunicable diseases [14], there are several other meta-analyses where some favorable effects have been found among people affected with chronic diseases [10-13,39]. The findings of the present review suggest that the use of social media in interventions for diabetes in many cases has been beneficial, and we did not find any studies that suggested worsened outcomes with this type of intervention. Studies using social media in their interventions have mostly showed superior results linked to the use of social media. Only one of the publications included in this review did not report any benefit on clinical or behavioral outcomes [30]. In this case, the researchers used social media in a 1-year intervention, and they used their own social media channel [31].

It is interesting to note that more than half of the studies used social media as a tool or resource to enhance the main intervention, and in these cases, the interventions resulted in improvements in HbA_{1c} values. Participants in these studies who were allocated to receive education either through Facebook chat or Skype as reinforcement of the main intervention had significant decreases in HbA_{1c} values [26,28,30,32,34,35,45]. Furthermore, compared with patients in the control group, improvements in HbA_{1c} levels were found among patients receiving health education through closed groups on Facebook and were found in one study using its own social media network [27,44]. In only 2 studies where social media was used as a supporting tool no differences in HbA_{1c} values were found; these 2 studies used their own social media channel [31,48].On the other hand, only 1 of the 3 studies that used social media as a main channel for delivering the intervention and measured HbA_{1c} values [29,42,54] reported significant improvements. This study used Viber communication for delivering doctor or peer support [54].

These findings suggest that using social media as a supporting tool for the main intervention is beneficial for improving health outcomes in T1D patients. Furthermore, it seems that the clinical improvement is most likely to happen when the chosen social media is one of the most popular networking sites.

Our review has mainly identified studies conducted with young T1D populations; however, our conclusions are comparable to those reported in a meta-analysis published in 2014 that predominantly analyzed studies involving T2D patients [39]. Therefore, we think that public health institutions, clinicians, and other stakeholders should consider the use of social media in their interventions targeting people affected with diabetes. However, special attention should be paid to the risk of misinformation or harmful health material that can coexist when carrying out interventions in open social media channels as it could lead to undesirable or unexpected effects [62-67].

Knowledge Gaps and Next Directions

Most of the included studies focused only on young people affected with T1D, probably because it is believed that these media are typically used by young people. Certainly, since the origin of social media, younger people have been the most frequent users of these channels. However, the presence of older generations on social media has increased in recent years, and about 80% of North Americans and Norwegians in their 40s report being social media users [8,68]. Hence, including older populations through social media would also make sense. This could be an especially valuable way of targeting people affected with T2D, a disease that is mostly diagnosed in adulthood and whose prevalence has dramatically increased in the last few decades [1]. Intervention studies using social media seem to improve health outcomes in T1D patients, and they could be beneficial for people with T2D as well. However, more research, using social media, on diabetes types is needed to answer this question.

In this review, we identified many abstracts presented at conferences, but there were fewer full papers reporting methods and results in detail. Knowing further details of the method used and the interventions could help identify the mechanisms or behavior techniques that work better for improving patient outcomes. So far, it seems that studies that use social media as supporting tool and where the social media is used for delivering health education report better outcomes. However, there are not enough studies where social media was used as the main channel for delivering the intervention. In future research, one should consider using different social media channels as main sources for delivering the intervention.

In research projects, it is more common to use restricted-access social media (ie, Facebook closed groups, Facebook chat, Skype, etc), which allows the researcher to have a better control of the environment and the contents and also protect the patients' privacy. However, the use of open social media channels offers the possibility of a large-scale impact. Providing high-quality contents on diabetes through the most commonly used open social media channels and interacting with the social media users could potentially help people with diabetes. By having access to free-of-charge quality information, they could improve their knowledge, an important prerequisite for improving self-management and health behaviors. Further research should explore how to best use open social media channels for health promotion interventions in diabetes.

Strengths and Limitations

Our results and conclusions might be susceptible to bias as a consequence of streamlining the systematic review process. There might be a selection bias (failure to search in additional potentially relevant databases, only 1 reviewer selecting the studies) and a publication bias (we only searched in 4 databases; we did not search for gray literature; and our search was limited to the English language). Eight of the included studies conducted in Macedonia could be based on the same study, although we treated the reported results independently, as they provided different sample sizes, different age ranges of the included participants, and different intervention periods. Because many of the included publications were abstracts presented at conferences and because we did not have access to complete data, a quantitative synthesis was not possible.

XSL•FO
RenderX

Conclusion

There is little evidence on the use of social media in interventions aimed at people affected with diabetes. However, after weighing the existing evidence, it seems that the use of these channels is predominantly linked to beneficial patient outcomes. Public health institutions, clinicians, and other stakeholders who aim at improving diabetes patient education should consider the use of social media in their interventions.

Acknowledgments

This work was supported by the Northern Norway Regional Health Authority (Helse Nord RHF), grant number HNF1370-17.

Authors' Contributions

EG designed the study, performed the searches, extracted the data, assessed the quality of the studies, analyzed the data, drafted and revised the manuscript, and approved the final manuscript. RW performed searches, assessed the quality of some papers, drafted and revised the manuscript, and approved the final manuscript. EÅ helped refining the search process and in deciding inclusion or exclusion, drafted and revised the manuscript, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy (search date: February 13, 2018).

[PDF File (Adobe PDF File), 43KB - jmir_v20i8e10303_app1.pdf]

Multimedia Appendix 2

List of all potentially relevant studies that were read in full-text form but excluded from the review.

[PDF File (Adobe PDF File), 52KB - jmir_v20i8e10303_app2.pdf]

Multimedia Appendix 3

Risk of bias assessment of the systematic reviews included in the review according to the MeaSurement Tool to Assess systematic Reviews (AMSTAR) criteria (n=3).

[PDF File (Adobe PDF File), 32KB - jmir_v20i8e10303_app3.pdf]

Multimedia Appendix 4

List of excluded studies because of low Grading of Recommendations Assessment, Development and Evaluation (GRADE) scores.

[PDF File (Adobe PDF File), 38KB - jmir_v20i8e10303_app4.pdf]

Multimedia Appendix 5

Summary of publications included in the review (n=20).

[PDF File (Adobe PDF File), 82KB - jmir_v20i8e10303_app5.pdf]

Multimedia Appendix 6

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist [22].

[PDF File (Adobe PDF File), 89KB - jmir_v20i8e10303_app6.pdf]

References

1. World Health Organization. 2016. Global report on diabetes URL: http://apps.who.int/iris/bitstream/10665/204871/1/9789241565257_eng.pdf; [WebCite Cache ID 6xgkLHMyL]
2. International Diabetes Federation. 2012. Global Guideline for type 2 diabetes URL: https://www.idf.org/e-library/guidelines/79-global-guideline-for-type-2-diabetes [accessed 2018-03-05] [WebCite Cache ID 6xgkSYzvG]
3. Lange K, Swift P, Pańkowska E, et al. ISPAD Clinical Practice Consensus Guidelines 2014. Diabetes education in children and adolescents. Pediatr Diabetes 2014 Sep;15 Suppl 20:77-85. [doi: 10.1111/pedi.12187] [Medline: 25182309]

XSL•FO
RenderX

4. Montgomery B. The Pillars of Prevention: Discover, Advocate, and Educate. Diabetes Spectr 2018 Feb;31(1):99-104. [doi: 10.2337/ds17-0078] [Medline: 29456433]
5. National Clinical Guideline Centre (UK). Type 1 Diabetes in Adults: Diagnosis and Management. London: National Institute for Health and Care Excellence (UK); 2015 Aug. National Institute for Health and Care Excellence: Clinical Guidelines 2015;NICE Guideline, No. 17:1-613. [Medline: 26334079]
6. National Collaborating Centre for Women's and Children's Health (UK). Diabetes (Type 1 and Type 2) in Children and Young People Diagnosis and Management. National Collaborating Centre for Women's and Children's Health (UK). London: National Institute for Health and Care Excellence (UK); 2015 Aug 2015;NICE Guideline, No. 18:1-517. [Medline: 26334077]
7. Schaper NC, Van NJJ, Apelqvist J, et al. Prevention and management of foot problems in diabetes: a Summary Guidance for Daily Practice 2015, based on the IWGDF Guidance Documents. Diabetes Metab Res Rev 2016 Jan;32 Suppl 1:7-15. [doi: 10.1002/dmrr.2695] [Medline: 26335366]
8. Pew Research Center. 2018. Social Media Fact Sheet URL: http://www.pewinternet.org/fact-sheet/social-media/ [accessed 2018-03-05] [WebCite Cache ID 6xgkXELo1]
9. Coiera E. Social networks, social media, and social diseases. BMJ 2013 May 22;346(may22 16):f3007-f3007. [doi: 10.1136/bmj.f3007]
10. Laranjo L, Arguel A, Neves AL, 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]
11. An R, Ji M, Zhang S. Effectiveness of Social Media-based Interventions on Weight-related Behaviors and Body Weight Status: Review and Meta-analysis. Am J Health Behav 2017 Nov 01;41(6):670-682. [doi: 10.5993/ajhb.41.6.1]
12. Välimäki M, Athanasopoulou C, Lahti M, Adams CE. Effectiveness of Social Media Interventions for People With Schizophrenia: A Systematic Review and Meta-Analysis. J Med Internet Res 2016 Apr 22;18(4):e92 [FREE Full text] [doi: 10.2196/jmir.5385] [Medline: 27105939]
13. Williams G, Hamm MP, Shulhan J, et al. Social media interventions for diet and exercise behaviours: a systematic review and meta-analysis of randomised controlled trials. BMJ Open 2014 Feb;4(2):e003926 [FREE Full text] [doi: 10.1136/bmjopen-2013-003926] [Medline: 24525388]
14. Mita G, Ni MC, Jull A. Effectiveness of social media in reducing risk factors for noncommunicable diseases: a systematic review and meta-analysis of randomized controlled trials. Nutr Rev 2016 Apr;74(4):237-247 [FREE Full text] [doi: 10.1093/nutrit/nuv106] [Medline: 26946250]
15. Translate Media. 2018. Scandinavia Social Media URL: https://www.translatemedia.com/translation-services/social-media/scandinavia-social-media/ [accessed 2018-03-05] [WebCite Cache ID 6xgkbsLAo]
16. Statista. 2018. Forecast of social network user numbers in Norway from 2014 to 2021 (in million users) URL: https://www.statista.com/statistics/569019/predicted-number-of-social-network-users-in-norway/ [accessed 2018-03-05] [WebCite Cache ID 6xgkmh9FI]
17. Gabarron E, Bradway M, Fernandez-Luque L, et al. Social media for health promotion in diabetes: study protocol for a participatory public health intervention design. BMC Health Serv Res 2018 Jun 05;18(1):414 [FREE Full text] [doi: 10.1186/s12913-018-3178-7] [Medline: 29871675]
18. Garritty C, Stevens A, Gartlehner G, et al. Cochrane Rapid Reviews Methods Group to play a leading role in guiding the production of informed high-quality, timely research evidence syntheses. Syst Rev 2016 Dec 28;5(1):184 [FREE Full text] [doi: 10.1186/s13643-016-0360-z] [Medline: 27793186]
19. Kelly SE, Moher D, Clifford TJ. Quality of conduct and reporting in rapid reviews: an exploration of compliance with PRISMA and AMSTAR guidelines. Syst Rev 2016 May 10;5:79 [FREE Full text] [doi: 10.1186/s13643-016-0258-9] [Medline: 27160255]
20. Khangura S, Konnyu K, Cushman R, et al. Evidence summaries: the evolution of a rapid review approach. Syst Rev 2012 Feb 10;1:10 [FREE Full text] [doi: 10.1186/2046-4053-1-10] [Medline: 22587960]
21. Tricco AC, Antony J, Zarin W, et al. A scoping review of rapid review methods. BMC Med 2015 Sep 16;13:224 [FREE Full text] [doi: 10.1186/s12916-015-0465-6] [Medline: 26377409]
22. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009 Oct;62(10):1006-1012. [doi: 10.1016/j.jclinepi.2009.06.005] [Medline: 19631508]
23. Shea BJ, Hamel C, Wells GA, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol 2009 Oct;62(10):1013-1020. [doi: 10.1016/j.jclinepi.2008.10.009] [Medline: 19230606]
24. Shea B, Reeves B, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017 Sep 21;358:j4008. [doi: 10.1136/bmj.j4008]
25. Meerpohl JJ, Langer G, Perleth M, et al. [GRADE guidelines: 3. Rating the quality of evidence (confidence in the estimates of effect)]. Z Evid Fortbild Qual Gesundhwes 2012 Jan;106(6):449-456. [doi: 10.1016/j.zefq.2012.06.013] [Medline: 22857733]
26. Petrovski G, Zivkovic M, Stratrova SS, et al. Carelink, Skype and Facebook improve diabetes control in adolescents on pump therapy. Diabetologia 2010 Sep;53(Suppl 1):S416-S416. [doi: 10.1007/s00125-010-1872-z] [Medline: 20690007]

XSL•FO
RenderX

27. Hanberger L, Ludvigsson J, Nordfeldt S. Use of a web 2.0 portal to improve education and communication in young patients with families: randomized controlled trial. J Med Internet Res 2013 Aug 23;15(8):e175 [FREE Full text] [doi: 10.2196/jmir.2425] [Medline: 23973555]
28. Petrovski G, Dimirovski C, Bogoev M, et al. Internet visits improve diabetes control in adolescents on pump therapy. Diabetes Technology & Therapeutics 2011 Feb;13(2):261-262. [doi: 10.1089/dia.2010.1219]
29. Lapp J, White NH. Social networking and peer support (SNAPS) in adolescents with type 1 diabetes mellitus (T1D): A pilot study. Diabetes 2012 Jun 08;61(Supplement_1):A195-A195. [doi: 10.2337/db12-656-835]
30. Petrovski G, Milenkovic T, Petrovska I, et al. Social media and diabetes: can we improve glucose control in adolescents on pump therapy? One year experience. Diabetes 2012;61:A230-A230.
31. Hanberger L, Ludvigsson J, Nordfeldt S. Use of a web 2.0 portal to improve education and communication in young patients with families: randomized controlled trial. J Med Internet Res 2013 Aug 23;15(8):e175 [FREE Full text] [doi: 10.2196/jmir.2425] [Medline: 23973555]
32. Petrovski G, Milenkovic T, Jovanovska B, et al. Can we improve glucose control in type 1 diabetics on insulin pump using social media and diabetes ? One year experience. Diabetes Technol Ther 2013;15:A111-A111.
33. Stellefson M, Chaney B, Barry AE, et al. Web 2.0 chronic disease self-management for older adults: a systematic review. J Med Internet Res 2013 Feb;15(2):e35 [FREE Full text] [doi: 10.2196/jmir.2439] [Medline: 23410671]
34. Petrovski G, Milenkovic MZT, Jovanovska B, et al. One-year experience of using social media as tool to improve glucose control in adolescents with type 1 diabetes : a crossover study. Diabetes 2014;63:A172-A172.
35. Petrovski G, Milenkovic T, Subeska S, et al. Social media and diabetes: a tool to improve glucose control in type 1 diabetic adolescents on insulin pump: cross-over study. Diabetes Technol Ther 2014;16:A41-A41.
36. Saboo B, Chandarana HK, Talaviya P, et al. Impact of use of social media in patients with type 1 diabetes for management of diabetes. Pediatric Diabetes 2014 Sep 03;15:67-67. [doi: 10.1111/pedi.12194_2]
37. Scaramuzza A, Bosetti A, Redaelli F, et al. To whatsapp or not to whatsapp? What could be done with new social media to manage type 1 diabetes in adolescents. Diabetes Technology & Therapeutics 2014 Feb;16(S1):A161-A162. [doi: 10.1089/dia.2014.1515]
38. Thomas J, Donaldson BJL. Sugar Free with Justin T.: Diabetes education through community partnerships. Journal of Extension 2014;52(6):6-6.
39. Toma T, Athanasiou T, Harling L, et al. Online social networking services in the management of patients with diabetes mellitus: systematic review and meta-analysis of randomised controlled trials. Diabetes Res Clin Pract 2014 Nov;106(2):200-211. [doi: 10.1016/j.diabres.2014.06.008] [Medline: 25043399]
40. Gómez-Zúñiga B, Pousada M, Hernandez MM, et al. The Online Big Blue Test for Promoting Exercise: Health, Self-Efficacy, and Social Support. Telemed J E Health 2015 Oct;21(10):852-859. [doi: 10.1089/tmj.2014.0158] [Medline: 26121419]
41. Lim PK, Cheng TS, Hui YCA, et al. D-buddy peer support for better health outcomes in adolescents with diabetes mellitus. International Journal of Pediatric Endocrinology. 2015 Presented at: Conferenceth Biennial Scientific Meeting of the Asia Pacific Paediatric Endocrine Society, APPES; 29 October - 1 November 2014; Darwin, Australia p. P15-P15.
42. McDarby V, Hevey D, Cody D. Evaluation of a social media site to improve diabetes glycaemic control, knowledge and self-efficacy: the ASSIST study. Pediatr Diabetes 2015;16:35.
43. Ng SM. Improving patient outcomes with technology and social media in paediatric diabetes. BMJ Qual Improv Rep 2015;4(1):u209396.w3846 [FREE Full text] [doi: 10.1136/bmjquality.u209396.w3846] [Medline: 26734405]
44. Petrovski G, Zivkovic M. Facebook as a useful tool to improve glucose control in patients with type 1 diabetes: One year follow-up study. Diabetes Technology & Therapeutics 2015 Feb;17(S1):A23-A23. [doi: 10.1089/dia.2015.1525]
45. Petrovski G, Zivkovic M, Stratrova SS. Social media and diabetes: can Facebook and Skype improve glucose control in patients with type 1 diabetes on pump therapy? One-year experience. Diabetes Care 2015 Apr;38(4):e51-e52. [doi: 10.2337/dc14-2487] [Medline: 25805869]
46. Rothenberg R, Zetelski M, Sivitz J, et al. Use of smartphone, a cellular glucometer and social media app in the management of type 1 DM in the adolescent population: The future of diabetes care. In: Horm Res Paediatr. 2015 Presented at: 4th Annual Meeting of the European Society for Paediatric Endocrinology (ESPE); October 1-3, 2015; Barcelona, Spain p. 374-375 URL: https://www.karger.com?DOI=10.1159/000437032 [doi: 10.1159/000437032]
47. Spehar UA, Bogdanic A, Krnic N, et al. Diabetes-management empowerment intervention 'Youth for adolescents with type 1 diabetes'. Pediatric Diabetes 2015 Oct 06;16:136-137. [doi: 10.1111/pedi.12309]
48. Whittemore R, Jeon S, Liberti L, et al. Implementation and efficacy of two internet interventions for teens with T1D. Diabetes 2015;1:A215-A215.
49. Wilson L, Allardice B, Brillante N, et al. The role of My Diabetes My Way (MDMW) social media sites in promoting diabetes education and self-management. Diabet. Med 2015 Mar 11;32:134-135. [doi: 10.1111/dme.12668_1]
50. Yi-Frazier JP, Cochrane K, Mitrovich C, et al. Using Instagram as a Modified Application of Photovoice for Storytelling and Sharing in Adolescents With Type 1 Diabetes. Qual Health Res 2015 Oct;25(10):1372-1382 [FREE Full text] [doi: 10.1177/1049732315583282] [Medline: 25904674]
51. Yi-Frazier JP, Mitrovich C, Pascual M, et al. Using instagram to improve outcomes in adolescents with type 1 diabetes (T1D): A feasibility study. Diabetes 2015;1:A671-A671.

XSL•FO
RenderX

52. Blackstock S, Solomon S, Watson M, et al. The use of a WhatsAppTM broadcast group to improve knowledge and engagement of adolescents with type 1 diabetes. Archives of Disease in Childhood 2016;101(Supplement 1):A315-A316.
53. Marsland N, Twenefour D, Kelly T. Impact of 'Enjoy Food': Diabetes UK's programme to promote healthy eating. Diabetic Medicine 2016;1:117-117.
54. Petrovski G, Zivkovik M, Ahmeti I, et al. Mobile social media and diabetes: Viber communication improves glucose control in type 1 diabetes patients on insulin pump. Diabetes Technology and Therapeutics 2016;1:A97-A97.
55. Pyatak E, Carandang K, Diaz J, et al. Resilient, empowered, active living (real) with diabetes: Implementing an occupational therapy (OT) intervention for diabetes management. Diabetes 2016;65(Supplement 1):A561-A561.
56. Kariyawasam D, Pender S, Jones M, et al. An evaluation of a novel programme for empowering young people with Type 1 diabetes: YES -Youth Empowerment Skills. Diabetic Medicine 2017;34(Supplement 1):121-121.
57. Marsland N, Twenefour D, Elvin E. Impact of 'Enjoy Food': Diabetes UK's programme to promote healthy eating. Diabetic Medicine 2017;34(Supplement 1):128-128.
58. Vorderstrasse A, Melkus GD, Feinglos M, et al. Virtual environment for diabetes self-management education and support: Preliminary RCT outcomes. Circulation. Conference: Resuscitation Science Symposium, ReSS 2017;136(Supplement 1):A17519-A17519.
59. Wilson L, Cunningham SG, Allardice B, et al. My diabetes my way: Delivering innovative diabetes care. Diabetic Medicine 2017;34(Supplement 1):116-117.
60. Alanzi T. Role of Social Media in Diabetes Management in the Middle East Region: Systematic Review. J Med Internet Res 2018 Feb 13;20(2):e58 [FREE Full text] [doi: 10.2196/jmir.9190] [Medline: 29439941]
61. Seo H, Kim KU. Quality assessment of systematic reviews or meta-analyses of nursing interventions conducted by Korean reviewers. BMC Med Res Methodol 2012 Aug 28;12:129-129 [FREE Full text] [doi: 10.1186/1471-2288-12-129] [Medline: 22928687]
62. Gabarron E, Serrano JA, Wynn R, Lau AYS. Tweet content related to sexually transmitted diseases: no joking matter. J Med Internet Res 2014 Oct 06;16(10):e228 [FREE Full text] [doi: 10.2196/jmir.3259] [Medline: 25289463]
63. Lau AYS, Gabarron E, Fernandez-Luque L, Armayones M. Social media in health--what are the safety concerns for health consumers? Health Inf Manag 2012;41(2):30-35. [Medline: 23705132]
64. Oyeyemi SO, Gabarron E, Wynn R. Ebola, Twitter, and misinformation: a dangerous combination? BMJ 2014;349:g6178. [Medline: 25315514]
65. Pal BR. Social media for diabetes health education - inclusive or exclusive? Curr Diabetes Rev 2014;10(5):284-290. [Medline: 25316149]
66. Weitzman ER, Kelemen S, Garvey KC. Social networking for care improvement and panel management. Diabetes 2013;1:A349-A349.
67. Wynn R, Oyeyemi SO, Johnsen JA, et al. Tweets are not always supportive of patients with mental disorders. International J Integrated Care 2017;17(3):A149-A149 [FREE Full text]
68. Statista. 2018. Share of social media users in Norway in 2017, by age group URL: https://www.statista.com/statistics/679689/share-of-social-media-users-in-norway-by-gender/ [accessed 2018-03-05] [WebCite Cache ID 6xgkynVMp]

Abbreviations

HbA_{1c}: glycated hemoglobin
HRQoL: health-related quality of life
RCT: randomized controlled trial
T1D: type 1 diabetes
T2D: type 2 diabetes

Edited by G Eysenbach; submitted 06.03.18; peer-reviewed by R Fensli, E Da Silva, D Greenwood; comments to author 19.04.18; revised version received 15.05.18; accepted 18.06.18; published 10.08.18.

Please cite as:
Gabarron E, Årsand E, Wynn R
Social Media Use in Interventions for Diabetes: Rapid Evidence-Based Review
J Med Internet Res 2018;20(8):e10303
URL: http://www.jmir.org/2018/8/e10303/
doi:10.2196/10303
PMID:30097421

XSL•FO
RenderX

©Elia Gabarron, Eirik Årsand, Rolf Wynn. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 10.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Social Media Landscape of the Tertiary Referral Hospitals in China: Observational Descriptive Study

Wei Zhang[1], PhD; Zhaohua Deng[1], PhD; Richard Evans[2], PhD; Fei Xiang[1], PhD; Qing Ye[3], MA; Runxi Zeng[4], PhD

[1]Institute of Smart Health, School of Medicine and Health Management, Huazhong University of Science and Technology, Wuhan, China

[2]Business Information Management and Operations, University of Westminster, London, United Kingdom

[3]Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[4]School of Journalism and Communication, Chongqing University, Chongqing, China

Corresponding Author:
Zhaohua Deng, PhD
Institute of Smart Health
School of Medicine and Health Management
Huazhong University of Science and Technology
No. 13 Hangkong Road, Qiaokou District
Wuhan, 430030
China
Phone: 86 15926318828
Email: zh-deng@hust.edu.cn

Abstract

Background: Social media has penetrated all walks of life. Chinese health care institutions are increasingly utilizing social media to connect with their patients for better health service delivery. Current research has focused heavily on the use of social media in developed countries, with few studies exploring its usage in the context of developing countries, such as China. Tertiary hospitals in China are usually located in city centers, and they serve as medical hubs for multiple regions, with comprehensive and specialized medical care being provided. These hospitals are assumed to be the pioneers in creating official social media accounts to connect with their patients due to the fact that they appear to have more resources to support this innovative approach to communication and health care education.

Objective: The objective of our study was to examine China's best tertiary hospitals, as recognized by The National Health Commission of the People's Republic of China (NHCPRC), and to map out the landscape of current social media usage by hospitals when engaging with patients.

Methods: We examined the best 705 tertiary hospitals in China by collecting and analyzing data regarding their usage of popular Chinese social media apps Sina Weibo and WeChat. The specific data included (1) hospital characteristics (ie, time since established, number of beds, hospital type, and regions or localities) and (2) status of social media usage regarding two of the most popular local social media platforms in China (ie, time of initiation, number of followers, and number of tweets or posts). We further used a logistic regression model to test the association between hospital characteristics and social media adoption.

Results: Of all, 76.2% (537/705) tertiary referral hospitals have created official accounts on either Sina Weibo or WeChat, with the latter being more popular among the two. In addition, our study suggests that larger and newer hospitals with greater resources are more likely to adopt social media, while hospital type and affiliation with universities are not significant predictors of social media adoption among hospitals.

Conclusions: Our study demonstrated that hospitals are more inclined to use WeChat. The move by hospitals from Sina Weibo to WeChat indicates that patients are not satisfied by mere communication and that they now place more value on health service delivery. Meanwhile, utilizing social media requires comprehensive thinking from the hospital side. Once adopted, hospitals are encouraged to implement specific rules regarding social media usage. In the future, a long journey still lies ahead for hospitals in terms of operating their official social media accounts.

(J Med Internet Res 2018;20(8):e249) doi:10.2196/jmir.9607

KEYWORDS

social media usage; best tertiary hospitals; China; Sina Weibo; WeChat

XSL•FO
RenderX

Introduction

Social media is now an indispensable part of human life. The number of globally active social media users reached a new high of 2.8 billion in 2017, representing 37% worldwide penetration. Facebook has the highest number of active users (1871 million), followed by Facebook Messenger, WhatsApp, and YouTube [1]. The ease of use of current social media websites and apps, including their uncomplicated operation and simplicity in generating user content and instant messages, allows people to connect in a more convenient manner across space and over time. It is undeniable that social media has enabled the reshaping of how we identify, get to know, and maintain relationships with others.

In recent years, health care professionals and institutes have begun to realize the benefits of using social media for building patient-physician relationship and for health care service delivery [2,3]. Evidence suggests that social media usage by hospitals and health care professionals contributes to the increase in hospital website visitors, brand establishments, and patient recruitment for research projects [4-6]. In the United States, 70% of hospitals use Facebook [7] while most hospitals utilize at least one social media platform [8]. Wong et al [9] have investigated the adoption of social media by children's hospitals in the United States and concluded that social media can serve as a channel for providing health care education and community engagement. In Western Europe, hospitals in the Netherlands and the United Kingdom are seen as pioneers in health care social media adoption, and the most widely adopted social media platforms by hospitals are YouTube, LinkedIn, and Facebook [10]. Studies have also discovered certain patterns of social media usage among hospitals; for example, large, urban, not-for-profit hospitals and hospitals that are affiliated with universities or health systems are more likely to operate official social media accounts [7,8]. Despite the recognized rise in social media usage by hospitals, a survey conducted in 2014, which explored the official websites of Italian local health authorities and public hospitals, found a low social media usage [11]. Researchers [12-14] have also established that the effects of social media on patient engagement vary between hospitals and that substantial differences exist between social media adoption and its effects. Further content analysis has revealed that the current social media usage in public health agencies has enabled the centralization of information distribution, rather than interactive communication, and a strategic communication plan has been deemed necessary for expanding the reach of social media and fostering interactivity and engagement among patients [12]. A study on local health care departments in the United States also indicated that although Twitter has been widely adopted, its primary use has been a one-way communication by hospitals on personal health care topics and organization-related information [13].

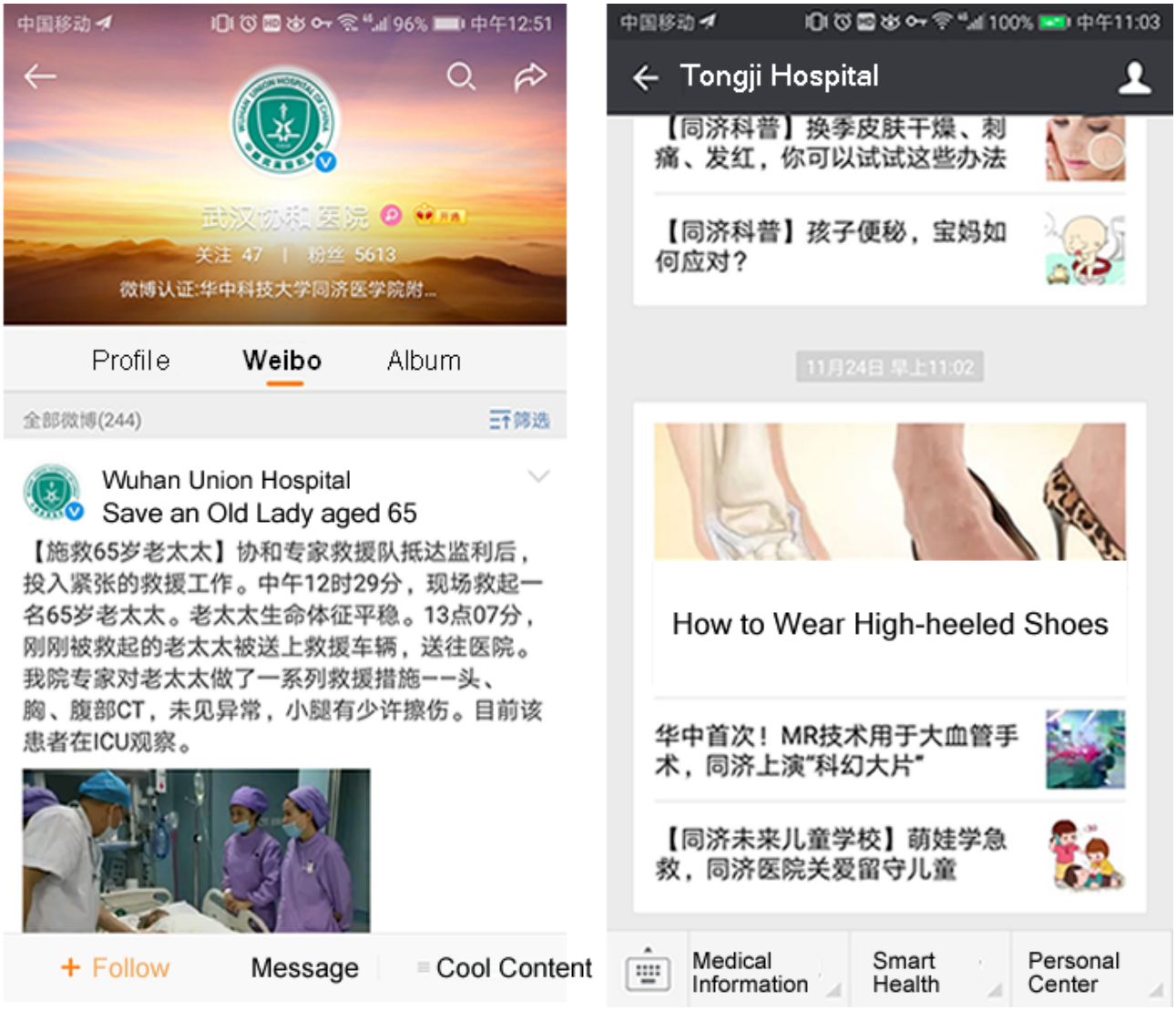
All these studies have contributed to an understanding of how hospitals use social media and its impact. However, almost all research conducted has been in the context of developed countries, with less attention being paid to developing countries, such as China. Due to differences in health care systems and local culture, social media usage by hospitals in developing countries may vary greatly from similar hospitals in developed countries.

Over the past several years, social media has been gradually adopted by China's hospitals for various reasons, such as building harmonious physician-patient relationships, reducing a hospital's burden on health service delivery, and improving public health literacy. In China, social media usage offers at least the following two benefits to the public: (1) It is a reliable source of health information. Enabled by information communication technology, everyone effortlessly becomes a self-publisher. Web-based health information can be provided by anyone who has access to the internet, although the trustworthiness of the information is not always guaranteed. With the increase in social media accounts operated by recognized health institutions, the public can now increasingly trust health information posted online and not be misled by unrecognized self-publishers. (2) It is a convenient and precise match for medical consultations in hospitals. For patients paying a first-time visit to a hospital, the interaction with the hospital's staff via social media is of great help in finding the right physician within a short period of time. Otherwise, patients may waste lots of time in, for example, figuring out the right department and the appropriate physician who specializes in their illness.

This study aimed to provide an empirical analysis of social media usage by the best hospitals in China, one of the largest developing countries. Although China has 29,719 hospitals, which were being used by more than 1.3 billion individuals by the end of July 2017 [15], the vast majority of individuals in China want to go to the best hospitals and spare no efforts when seeing a doctor; this is deemed to increase the pressure on China's health care system, especially on the best tertiary hospitals in the country. As a result, the best tertiary hospitals are typically overcrowded and have a strong desire to improve their patient-centered communication and health care services. Meanwhile, 57% of individuals in China are active on social media, spending an average of 1 hour and 50 minutes on social media platforms per day [1]. The most popular social media platforms in China are those tailored toward Chinese nationals, which include WeChat (67%), Youku, Sina Weibo (45%), and Tencent Weibo (31%); all of these platforms are different from Facebook and Twitter in terms of their functionality and bespoke features for user cultures. For example, WeChat is a closed social networking platform that allows users to connect with everyone in their circle, whereas strangers and the general public are not allowed to review or comment on their personal pages. In addition, WeChat embraces instant messaging with text, image, voice, and video chat; these features make the social media usage by China's best hospitals unique [16]. Hospitals implement WeChat to provide services to patients, including setting up appointments with physicians and providing health education in general, while their usage of Sina Weibo focuses on instant communication with patients and health education. Figure 1 presents an example of WeChat and Sina Weibo adoption by 2 hospitals in China.

XSL•FO
RenderX

Figure 1. Social media usage by two hospitals (Sina Weibo on the left and WeChat on the right).



The following research questions (RQs) were answered in this paper:

RQ 1: How has social media been adopted and utilized by China's best tertiary hospitals over time?

RQ 2: What factors are associated with the impact of social media on China's best tertiary hospitals?

Methods

Study Sample

In total, 705 of China's best tertiary hospitals, as recognized by the NHCPRC, were selected for this study. These hospitals were assumed to be the pioneers in creating official social media accounts to connect with their patients due to the fact that they appear to have more resources to support this innovative approach to communication and health care education. In China, hospitals are classified into three categories according to their ability to provide medical care and education and to conduct medical research: (1) primary, (2) secondary, and (3) tertiary. For example, tertiary hospitals are usually located in city centers, and they serve as medical hubs for multiple regions, with comprehensive and specialized medical care being provided [17]. Furthermore, each hospital in China is graded as either A, B, or C based on an overall evaluation of medical technology adoption, medical equipment, patient safety, hospital management, etc. Grade "A" indicates the best quality, whereas grades "B" and "C" indicate lower qualities. In our study, the best tertiary hospitals are defined as those categorized as tertiary and graded A. We followed the most recent updates on the NHCPRC's official website using the correct data as of June 1, 2017. The sample of 705 hospitals is spread across 31 provincial governments in Mainland China, and the number of hospitals in each provincial government ranges from 1 to 66.

Data Collection

Data collection consisted of two parts: (1) collection of hospital attributes and (2) collection of data on social media usage. For part 1, we collected data pertinent to the characteristics of the hospitals, including number of beds, geographical location, year of establishment, and hospital type (comprehensive or specialized hospitals), from their official websites. For part 2, we focused on the hospitals' official presence on Sina Weibo and WeChat, the most popular social media platforms in China. Specifically, the following data were included: date of account creation, number of followers, number of tweets or posts, the

XSL•FO
RenderX

most recent update time, and the major functionality offered on their social media account.

To ensure the comprehensiveness of data on official social media usage, we explored the required information using two approaches [7]: (1) searching the hospitals' official websites to check for social media accounts and (2) searching verified account information on Sina Weibo and WeChat using various keywords, including full name and abbreviation(s). We developed an initial data collection framework with detailed instructions. To collect data, three research assistants were employed and trained for the work. To ensure that the social media accounts were measured consistently, we removed the verified accounts of the separate departments or sections of the same hospital, for example, Department of Medicine of Hospital A or Nursing Station of Hospital B; this was done because these verified accounts only partially represented each hospital and not the hospital as a whole.

In addition, we further assessed the top 10 most popular hospitals on Sina Weibo as a case study for the availability of their data to explore the contents initiated by these hospitals on their social media accounts and their interaction with the general public.

Data Analysis

To analyze the collected data, a unique coding scheme was developed. Specifically, the coding consisted of two parts: the first was for hospital characteristics and the second for hospital social media account. In the first part, the Year of Establishment included 4 codes (1=before 1900, 2=1900-1949, 3=1950-1999, and 4=2000 onwards); Number of Beds included 5 codes (1=below 500, 2=501-1000, 3=1001-1500, 4=1501-2000, and 5=over 2000); and Status of Affiliation included 2 codes (0=no affiliation with a university and 1=affiliation with a university). Furthermore, we divided Mainland China into three regions according to the National Bureau of Statistics of China. The three regions were coded as Eastern, Central, and Western, with Eastern having the highest economic development level and Western having the lowest economic development level. In addition, hospital types were divided into "specialized" and "comprehensive." Regarding social media usage, our assessment was based on whether the hospital had an active account on Sina Weibo or WeChat. Meanwhile, Year of Verification, Number of Follows, and Number of Tweets or Posts were also coded.

To answer RQ1, we mapped out social media usage among the 705 hospitals across time using ArcGIS. To answer RQ2, we used the logistic regression model to analyze the correlation between hospital characteristics and their adoption of social media platforms.

Results

Descriptive Study

The distribution of the studied 705 hospitals is presented in Figure 2. The best medical resource in Mainland China is unevenly distributed across the three regions. It is evident that the Eastern region has more of the best tertiary hospitals than the Central and Western regions. Among these hospitals, 76.2% (537/705) have official accounts on either Sina Weibo or WeChat (Figure 3).

Social Media Adoption Is Prevalent Among the Best Tertiary Hospitals in China, With WeChat Being More Popular Than Sina Weibo

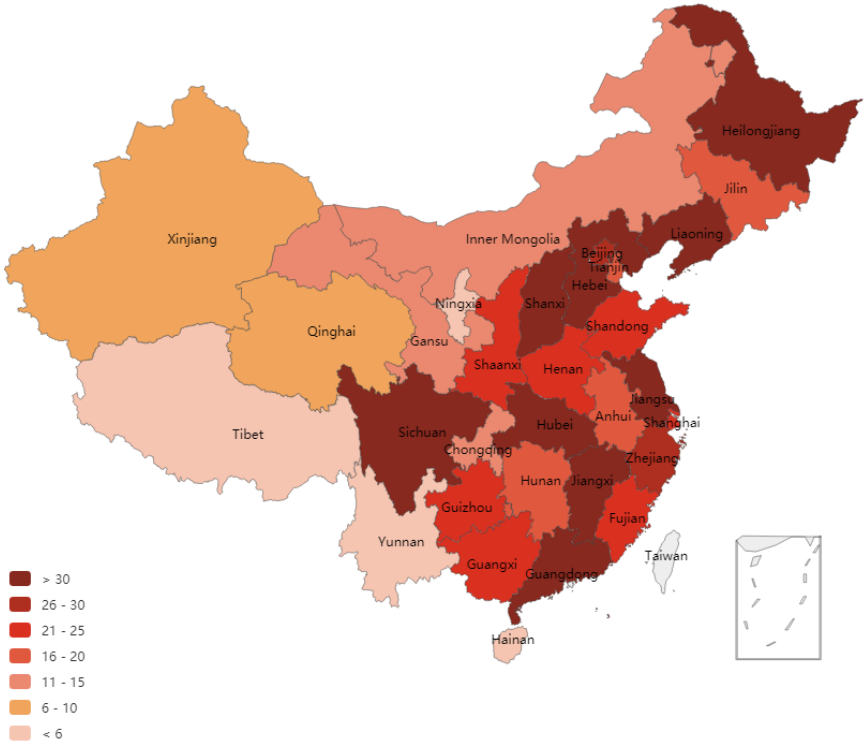
Among the 537 hospitals that have opened at least one official social media account, 267 use Sina Weibo and 446 use WeChat. Table 1 provides the number of social media accounts by province. In total, 176 hospitals have opened accounts on both Sina Weibo and WeChat. The use of Sina Weibo greatly varies across the three regions. Over half of the accounts (142/267, 53.2%) were identified in the Eastern region, whereas only 20.9% (56/267) were found in the Central region and 25.8% (69/267) in the Western region. Meanwhile, the earliest hospitals using Sina Weibo appear in the Eastern region, in Beijing and Tianjin in around 2009. Regarding the use of WeChat, the diffusion of hospitals is distributed evenly across the three regions. Specifically, the Eastern and Central regions have 182 and 173 hospitals using WeChat, respectively. The earliest WeChat accounts were identified in 2014, with 20 hospitals from 7 provinces across the three regions being identified. Surprisingly, Heilongjiang, located in the Central region of China, outnumbered other provinces with its 12 WeChat accounts. We also found that the total number of best hospitals that adopted WeChat outnumbered those that adopted Sina Weibo. In total, 23 out of 31 provinces had more hospital WeChat accounts than Sina Weibo accounts, excluding 3 provinces from the Eastern (Beijing, Shanghai, and Shandong) and 4 provinces from the Western (Chongqing, Inner Mongolia, Yunnan, and Ningxia) regions.

Social Media Penetration Rates Vary Across Provinces, With the Best Tertiary Hospitals From the Central and Western Regions Catching Up

The best tertiary hospitals from almost all provinces showed interest in adopting Sina Weibo or WeChat. The best hospitals from 29 of the 31 provinces have opened their account on Sina Weibo, except for Hainan from the Eastern Region and Tibet from the Western region. All of the best hospitals in Beijing and Ningxia Hui autonomous region have opened accounts on Sina Weibo, while Hebei, in the Central region, has the lowest penetration rate on Sina Weibo (6%). Regarding the total number of official accounts, Guangdong and Beijing have the highest numbers with 30 accounts, followed by Shanghai (17 accounts). Regarding WeChat, 30 out of the 31 provinces have at least one best hospital with a WeChat account, except for the best hospitals in Tibet. Hubei province has the highest penetration rate at 97.2%. On the other hand, Yunnan has the lowest penetration rate at 20%. Regarding the total number of official accounts on WeChat, Guangdong has the highest number (41), followed by Hubei (35).

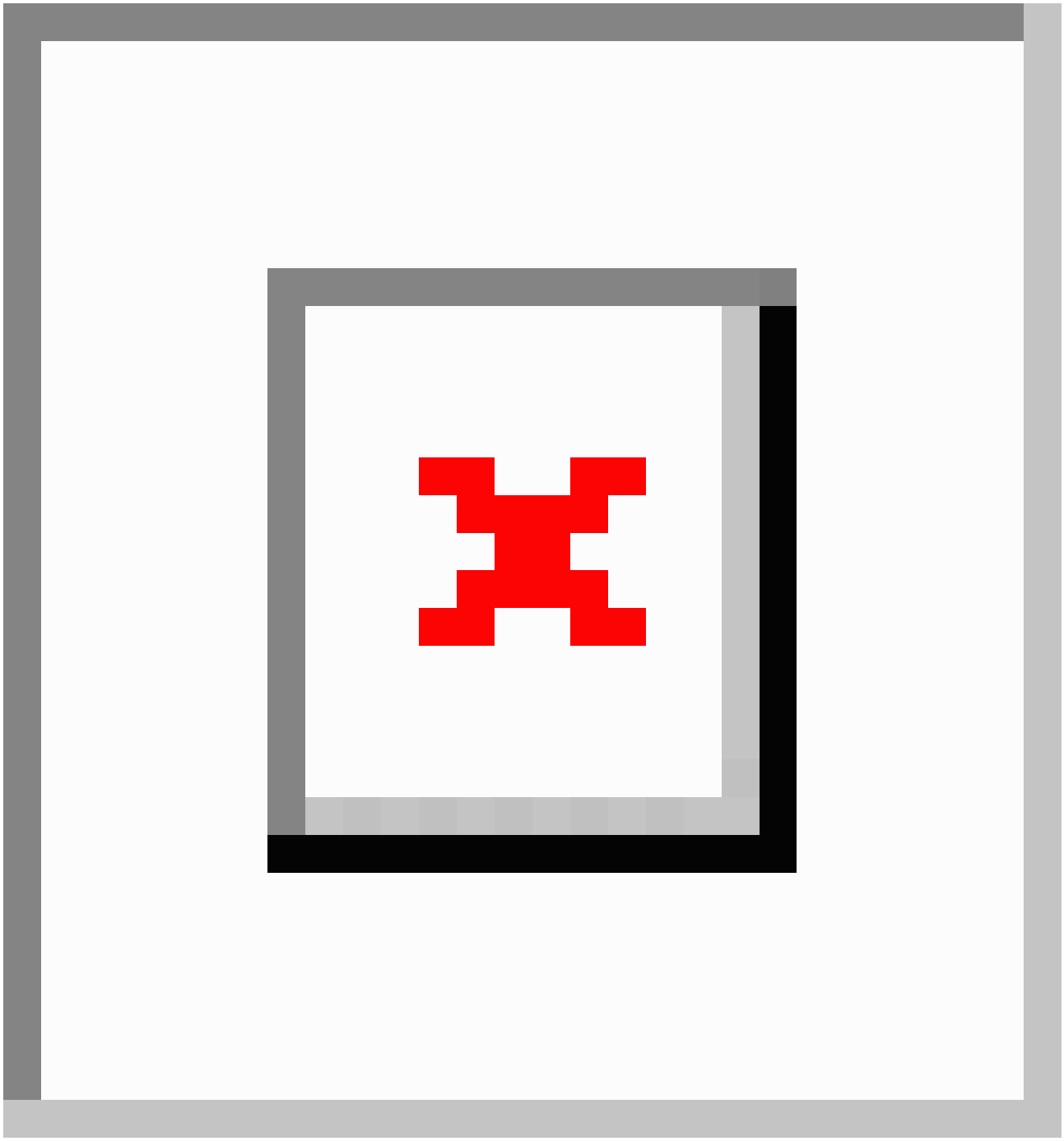
XSL•FO
RenderX

Figure 2. Distribution of the best tertiary hospitals across Mainland China.



XSL•FO
RenderX

Figure 3. Diffusion of Sina Weibo and WeChat among the best tertiary hospitals in Mainland China.



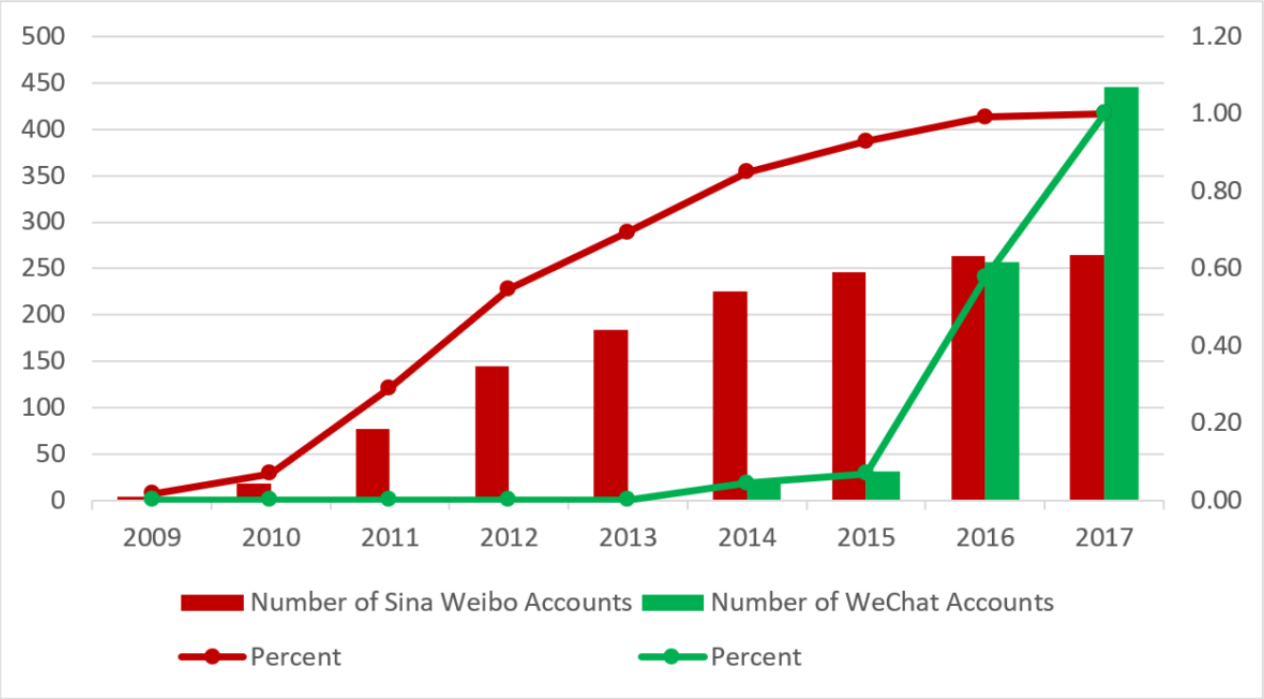
XSL•FO
RenderX

Table 1. Sina Weibo and WeChat diffusion among the 705 hospitals in Mainland China.

Province	No. of Best Hospitals	No. of Sina Weibo and WeChat accounts	No. of Sina Weibo accounts	No. of WeChat accounts
Eastern Region				
Beijing	30	45	30	15
Tianjin	17	18	8	10
Hebei	32	13	2	11
Liaoning	36	17	5	12
Shanghai	24	28	17	11
Jiangsu	38	36	15	21
Zhejiang	26	29	13	16
Shandong	21	27	14	13
Guangdong	66	71	30	41
Fujian	24	29	8	21
Hainan	5	3	0	3
Central Region				
Hubei	36	47	12	35
Hunan	20	28	10	18
Henan	24	24	7	17
Anhui	20	18	2	16
Jiangxi	33	34	8	26
Shanxi	32	31	8	23
Jilin	20	14	3	11
Heilongjiang	31	30	5	25
Western Region				
Guangxi	25	23	8	15
Chongqing	11	14	9	5
Sichuan	36	43	16	27
Guizhou	23	16	5	11
Inner Mongolia	13	11	9	2
Yunnan	5	3	2	1
Tibet	1	0	0	0
Shaanxi	25	28	11	17
Gansu	12	15	5	10
Qinghai	8	6	1	5
Ningxia	3	4	3	1
Xinjiang	8	8	1	7
Total	705	713	267	446

XSL•FO
RenderX

Figure 4. Social media usage among the best tertiary hospitals over time (2009-2017).



New Adopters Have Emerged, With Inactive Members Also Appearing

Over the past decade, the social media landscape of China's best hospitals has dramatically changed (Figure 4). Regarding Sina Weibo, the best hospitals have already started to explore this innovative communication tool since its establishment in 2009. Adoption peaked in 2012, with 68 best hospitals using the platform; after that, new adopters began to decline. Two best hospitals, one from Shandong and the other from Anhui province, have recently adopted Sina Weibo in the first half of 2017. Figure 5 provides the diffusion of Sina Weibo use across provinces from 2009 until mid-2017. Meanwhile, we also found that 25.5% (68/267) hospitals across 24 provinces had inactive Sina Weibo accounts during the past 6 months.

We identified a similar diffusion trend for WeChat. However, WeChat has fewer inactive hospital accounts. WeChat introduced an account verification service in the late 2012, with the first verified account among the best hospitals appearing in 2014. In total, 20 early WeChat adopters covered three regions, including 12 adopters from Heilongjiang, 3 from Beijing, and the remaining from Shanghai, Guangdong, Guangxi, Sichuan, and Shaanxi. In 2015, 11 hospitals joined WeChat, while in 2016, 227 did so. During the first half of 2017, 197 new best hospitals started using WeChat to connect with the public, indicating large-scale and continuous adoption. Figure 6 shows the diffusion of WeChat across provinces from 2009 until mid-2017. However, we also noted that WeChat accounts of 6.5% (29/446) best hospitals across 13 provinces had been inactive for the past 6 months.

Social Media Adoption and Hospital Characteristics

To examine the association between social media adoption and hospital characteristics, we performed logistic regression analysis. We found that Sina Weibo and WeChat adoptions were differently predicted using hospital characteristics. Regarding Sina Weibo, a hospital's year of establishment (P=.09), hospital's affiliation with a university (P=.07), and hospital type (P=.007) were significant, whereas, regarding WeChat, number of beds (P=.003) and regions (P=.02) were significant. We used Sina Weibo as an example to unravel the popularity of social media in the best hospitals, exploring visibility for their followers. Of these, the average number of followers was 37,512.26 (min 12, max 1,576,347, SD 151,038.36), whereas the average number of tweets posted was 1446.16 (min 0, max 30,047, SD 279,674). We further treated the number of tweets posted as a dependent variable, the number of followers as an independent variable, and hospital characteristics as control variables. The linear regression indicated a strong correlation between the number of tweets and the number of followers (F_1=38.21; P<.001). The list of the top 10 most popular verified hospital accounts on Sina Weibo is presented in Table 2. Among the top 10 hospital accounts, 9 are from the Eastern region of China, with 1 being from the Western region. In the Eastern region, Beijing, China's capital city, has 7 such hospital accounts. Meanwhile, 7 of the 10 were specialized hospitals, with children's hospitals receiving greater attention from the public on social media.

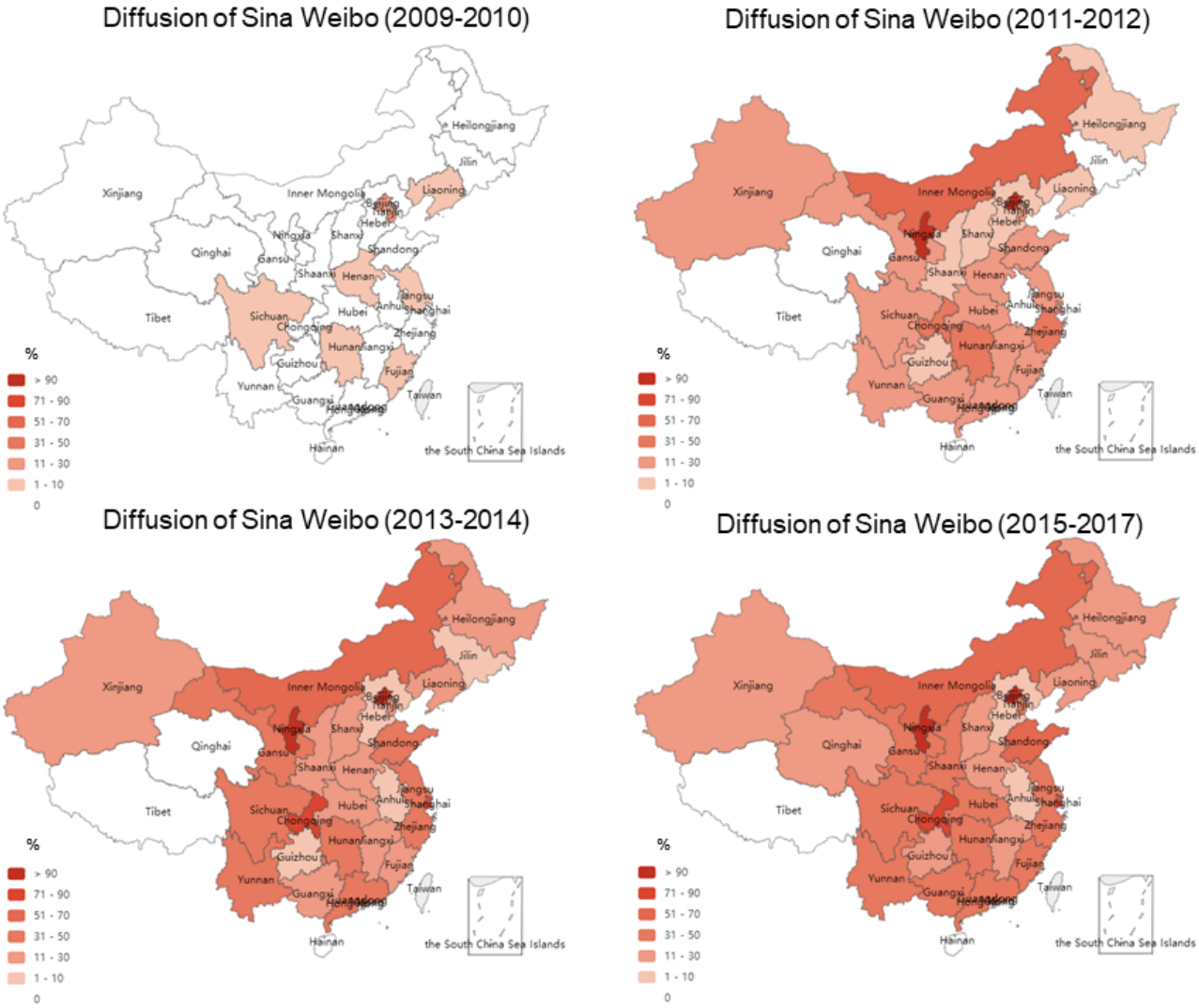
We further explored the content of the social media posts and the interactions through these by collecting the top 10 commented tweets or blogs from each of the 10 hospitals. Generally, the 100 posts can be divided into 5 categories on the basis of their content: (1) Health education, (2) Hospital news, (3) Medical consultation information, (4) Patient engagement, and (5) Official declaration. Health education posts are primarily associated with common health advice or misunderstood health facts. Hospital news is about any important event that is happening or has happened in the hospital, which serves as a self-promotion mechanism. Medical consultation information

XSL•FO
RenderX

concerns special arrangements for outpatient visits, for example, those during holidays or temporary changes such as closures. Patient engagement collects public opinions about health service preferences, such as any suggestions in relation to hospital operations. Official declaration clarifies recent misinformed or misunderstood information about hospital administration or health staff. Of the 100 most commented posts, health education and hospital news had 44 posts and 33 posts, respectively. Table 3 presents more details on these.

Additionally, we examined the most commented posts on the 10 most popular accounts; 70% (7/10) of the most commented posts were hospital news, and a significant difference was also identified on their likes (min 0, max 7554, SD 2335.89), reposts (min 3, max 22,464, SD 7802.73), and comments (min 23, max 3428, SD 1245.36). Among them, three posts from the Children's Hospital, Beijing Cancer Hospital, received more than 1000 comments. Results are presented in Table 4. Interestingly, a post related to hospital news on a general surgeon's relief work in Kunming attracted nationwide attention because the public mistook the general surgeon for a measurement of the physician's rank, rather than classification. Most of the comments were focused on why the hospital sent an ordinary doctor to the rescue and provided an explanation for the general surgeon. This post has now become a famous joke on the public's poor knowledge about medical issues. We also found that almost all interactions happened among netizens.

Figure 5. Diffusion of Sina Weibo across the best tertiary hospitals over time (2009-2017).



XSL•FO
RenderX

Figure 6. Diffusion of WeChat across the best tertiary hospitals over time (2009-2017).

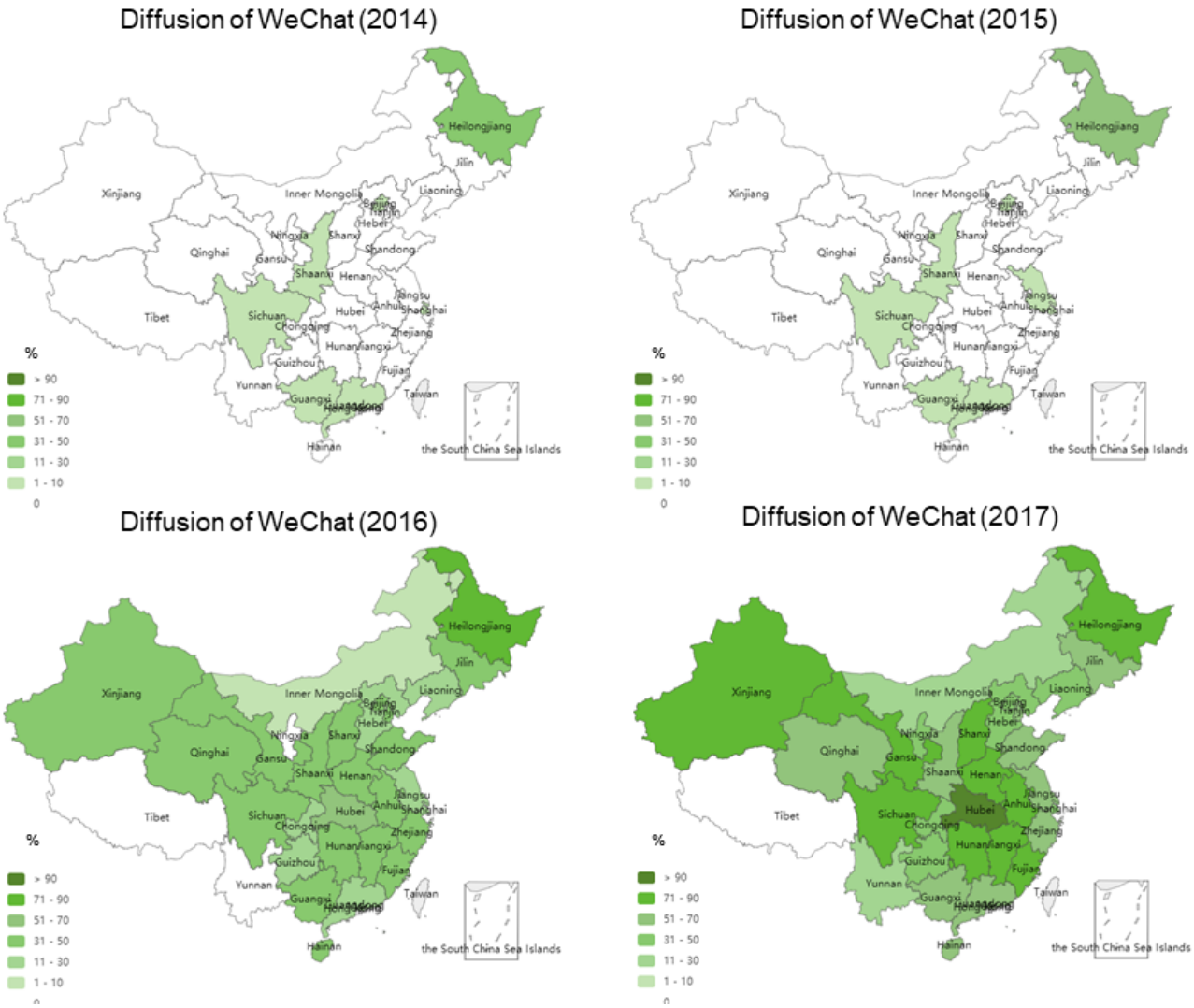


Table 2. List of the top 10 popular social media accounts among the 705 best hospitals.

Sina Weibo account	No. of followers	No. of Tweets
Children's Hospital, Capital Institute of Pediatrics	1,576,347	9631
Peking University, First Hospital	1,075,588	2511
Beijing Cancer Hospital	852,964	5178
Beijing Obstetrics and Gynecology Hospital, Capital Medical University	790,216	2132
Beijing Huilong Guan Hospital	650,954	1800
Beijing Children's Hospitals, Capital Medical Hospital	648,248	4871
Tianjin Stomatological Hospital, Hospital of Stomatology, Nankai University	379,160	30,047
Children's Hospital of Shanghai	268,430	3580
West China Hospital, Sichuan University	238,265	13,113
Peking University People's Hospital	203,612	3,133

XSL•FO
RenderX

Table 3. Category of the 100 most commented posts.

Type of information	No. of posts	Example
Health education	44	Never shake your baby.
Hospital news	31	Peking University People's Hospital has reached a new high! It is now first class, compared to Stanford University Hospital.
Medical consultation information	10	The hours of our outpatient visits are changing from 9-5 to 8-4, during the Spring Festival.
Patient engagement	8	Vote for your ideal type of psychiatrist.
Official declaration	5	An official declaration on the issue of workplace violence encountered by our nurses.

Table 4. Information about the most commented posts.

Posts	Type	No. of Likes	No. of Reposts	No. of Comments
A hard decision for the parents on their son's liver replacement surgery because of the unbreakable cost (Children's Hospital).	HN[a]	44	18	24
Our general surgeon Dr Jiang has flown to Kunming for disaster relief work (Peking University, First Hospital).	HN	1459	22464	2718
A new edition of *How far is cancer for you* has been released, and a brief introduction is provided (Beijing Cancer Hospital).	HE[b]	0	863	1222
For the convenience of the to-be-mothers, our hospital has implemented a 24×7 outpatient service scheme (Beijing Obstetrics and Gynecology Hospital).	HN	14	79	84
The reports on a patient visiting our psychiatry department, who was checked-in un-dressed, is not reliable (Beijing Huilong Guan Hospital).	OD[c]	0	53	42
A picture is more powerful than a thousand words (the hardworking doctors in Beijing Children's Hospitals).	HN	880	1080	182
Women's beauty relies on good sleep (Tianjin Stomatological Hospital).	HE	2	3	23
The music dream classroom, sponsored by a popular singer, in our hospital has opened (Children's Hospital of Shanghai).	HN	1655	2917	736
Using a make-up blogger's approach to promote polymerase chain reaction products by our medical students (West China Hospital, Sichuan University).	HN	7554	14573	3428
Our Prof Mu has participated in an international conference and has featured in a live broadcast medical plastic surgery section (Peking University People's Hospital).	HN	55	329	63

[a]HN: hospital news.

[b]HE: health education.

[c]OD: official declaration.

Discussion

Hospital Social Media Adoption Varies Across Regions, With Regional Centralization Being Most Important

Hospitals in the Eastern region were the first to adopt social media, with the total number of hospitals adopting social media being larger than that of the other two regions. This may be attributable to the abundant resources available in the hospitals in the Eastern region [18,19], such as technicians and financial support. In addition, compared with the inland regions, they are more open-minded and responsive to changes in general [20,21]. From the perspective of diffusion, neighboring provinces are likely to present similar adoption patterns [22]; this is due to their proximity and exchange of information between hospital managers and staff being more likely. With this informal sharing, the number of hospitals that have appreciated the importance of social media in connecting with patients has increased, and social media has been identified as a common strategy to improve patient-physician communication. In addition, other factors may also play a critical role, such as the adoption of social media by public health agencies, the local popularity of social media platforms, and hospital manager characteristics (ie, gender, age, and openness toward social media).

Transition of Usage From Communication to Service Orientation Makes WeChat More Attractive Among Hospitals

Sina Weibo initially attracted users because of its instant communication functionality. With the introduction of WeChat, Sina Weibo has gradually lost that advantage. As a new social media platform, WeChat offers greater functionality for engaging patients and offers nearly all the functions that Sina Weibo offers. More hospitals now turn to WeChat instead of Sina Weibo simply because WeChat reaches almost all ages, while users of Sina Weibo are decreasing and mostly comprise the younger age group. In addition, operating a microblogging account requires additional staff members to handle communication with internet citizens (netizens); this is time

XSL•FO
RenderX

consuming and less rewarding. For example, a hospital's Sina Weibo account is seen negatively as it distributes too much hospital information and offers less interaction with netizens. On the other hand, in the case of WeChat, individuals are more likely to review or provide a "thumbs up" or "like" and less likely to make comments because of its design. This makes operating a public WeChat account easier because it requires less attention compared with Sina Weibo. At the same time, the extra functionality that WeChat provides, such as appointment scheduling with the physicians and the checking of medical reports, makes it more attractive to users. Considering the long waiting times and difficulty in seeing expert physicians in China, WeChat has been very much welcomed by the public in China.

Adoption of Social Media is Faster in Larger, Newer Hospitals With More Resources

During the early adoption of social media, Sina Weibo attracted newer and comprehensive hospitals. This may be attributed to the target group of Sina Weibo (primarily youth) and to the fact that recently built hospitals are more open-minded toward adopting new technology than the long established ones. Hospitals with a long history are seldom regarded as early social media users because more concerns may arise, for example, whether Sina Weibo is aligned with their traditional image. The best traditional hospitals have already established their reputation among patients. As an innovative method of engaging customers, social media application in hospitals is still unpredictable. Instead of enhancing communication abilities, they prefer to invest in patient safety, such as medical error reduction. In addition, comprehensive hospitals are larger in size than the specialized ones, and it is not surprising that they have more resources to support the use of Sina Weibo. Interestingly, specialized hospitals have more supporters than comprehensive ones [23]. A possible reason for this is that a specialized hospital can provide more tailored health information to its patients.

Opening an Account is Never the End, and a Long Journey is Still Ahead for Hospitals

The verification of an official hospital account on a social media platform is only the first step. In fact, some of the studied hospitals discontinued their social media usage after only a few posts. This could be explained by the lack of staff, insufficient financial support, or simply little engagement from users. For hospitals, operating a social media account is not necessarily a task that they are judged on. Our observational study reveals that the content of hospital social media accounts is dominated by hospital news, and quite a few mutual communications between the public and the hospital are presented. Most of the interactions are generated by netizens, and the hospital seldom responds. Having official accounts means that hospitals need to interact with the public at all times. If a hospital fails to provide the public with the desired information, it may face intense criticism online. Therefore, operating social media accounts may put a hospital's reputation in danger. Even though some pioneers have started to produce specific internal guidelines for hospitals' social media usage [24,25] (eg, how to respond to difficult questions from the public, such as patient consultation for specific medical advice), most hospitals have no strict rules regarding its operation. For hospitals, it is definitely worth a second thought before creating official social media accounts. The process of monitoring, reviewing, and replying to patient concerns online may seem daunting to under-staffed information technology departments; however, it is crucial for developing a healthy online perception of the hospital and for enhancing relationships with current and prospective patients. In addition, although several hospitals have started realizing that patient comments posted online allow them to build and establish rapport and obtain feedback, rarely hospitals have successfully utilized them to further develop and improve their health care service delivery.

With the evolution of social media, it is quite possible that new forms of social media may replace once popular forms. Accordingly, another noteworthy problem—*A tale of two hospitals* on social media—arises for the hospitals that have both Sina Weibo and WeChat accounts. How do they operate their Sina Weibo and WeChat accounts—by simply synchronizing their content or by distinguishing it in some way?

Limitations

This is a descriptive study of social media usage among the best tertiary hospitals in China. Our survey sample is limited to tertiary hospitals, with secondary or primary hospitals being excluded due to their extremely low social media adoption rates. Future studies must include a greater number of hospitals and compare variations in their social media adoption. Furthermore, our survey featured two predominant social media platforms, Sina Weibo and WeChat. Other social media platforms also deserve attention. For example, YouTube has been widely adopted by hospitals in the United States [26], while Youku (similar to YouTube) is seldom used by China's hospitals [27], and it is worth investigating. In addition, this study is quantitative in nature; future studies, using a qualitative approach, should be conducted to explore specific strategies used by hospitals for social media adoption, such as who operates hospital accounts, what are the barriers and facilitators for them in engaging their patients on social media platforms [28], how hospital managers perceive the use of social media, and how does the social media account satisfy public needs [2].

Conclusions

This study revealed the landscape of social media usage among China's best tertiary hospitals. We found that social media adoption is prevalent among China's best hospitals and that WeChat is more popular than Sina Weibo. Although the number of adopters across the three regions of China is increasing, early adopters are more cautious in terms of their social media strategies. In general, larger and newer hospitals with more resources are more likely to adopt social media. The transition of users from Sina Weibo to WeChat suggests that patients were not satisfied with mere communication functionality and they now place greater value on health services A long journey still lies ahead for hospitals in terms of operating their official social media accounts. However, if they actively monitor, manage, and engage with patients online, they should reap the benefits of building lasting relationships, enhancing satisfaction, and reassuring patients that their voices are being heard.

XSL•FO
RenderX

Acknowledgments

This work has been partially supported by the National Natural Science Foundation of China (Award #71671073), National Social Science Foundation of China (Award #18BZZ078), and the Fundamental Research Funds for the Central Universities (HUST: 2018KFYYXJJ062).

Conflicts of Interest

None declared.

References

1. Kemp S. Digital in 2017 Global Overview: Wearesocial. Digital in 2017 Global Overview: Wearesocial; 2017. URL: https://www.slideshare.net/wearesocialsg/digital-in-2017-global-overview [accessed 2018-05-05] [WebCite Cache ID 717x8jMu4]
2. Payton F, Pinto N. Hospital Social Media Strategies: Patient or Organization Centric? Handbook of Research on Healthcare Administration and Management. In: Health Care Delivery and Clinical Science: Concepts, Methodologies, Tools, and Applications (3 Volumes). USA: Information Resources Management Association; Dec 2017:2016-2458.
3. Gallagher L, McCuaig J, Benoit L, Davies C. It's Time for the Genetic Counseling Profession to Embrace Social Media. J Genet Couns 2016 Dec;25(6):1338-1341. [doi: 10.1007/s10897-016-9950-0] [Medline: 27026235]
4. Thielst CB. Using social media to engage patients: many tools exist to connect, communicate and build loyalty. Healthc Exec 2011 May;26(3):66, 68-66, 70. [Medline: 21612166]
5. 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]
6. Trivedi R, Shah G, Bangar A. Use of Facebook by local health departments: usage, public health themes, and maintenance. Frontiers in Public Health Services and Systems Research 2015;4(5):1-6.
7. Richter JP, Muhlestein DB, Wilks CE. Social media: how hospitals use it, and opportunities for future use. J Healthc Manag 2014;59(6):447-460. [Medline: 25647968]
8. Griffis HM, Kilaru AS, Werner RM, Asch DA, Hershey JC, Hill S, et al. Use of social media across US hospitals: descriptive analysis of adoption and utilization. J Med Internet Res 2014 Nov 27;16(11):e264 [FREE Full text] [doi: 10.2196/jmir.3758] [Medline: 25431831]
9. Wong C, Ostapovich G, Kramer-Golinkoff E, Griffis H, Asch D, Merchant R. How U.S. children's hospitals use social media: A mixed methods study. Healthc (Amst) 2016 Mar;4(1):15-21. [doi: 10.1016/j.hjdsi.2015.12.004] [Medline: 27001094]
10. Van DBTH, Berben S, Samsom M, Engelen L, Schoonhoven L. Use of social media by Western European hospitals: longitudinal study. J Med Internet Res 2012 May 01;14(3):e61 [FREE Full text] [doi: 10.2196/jmir.1992] [Medline: 22549016]
11. Vanzetta M, Vellone E, Dal M, Rocco G, De MM, Rosaria A. Communication with the public in the health-care system: a descriptive study of the use of social media in local health authorities and public hospitals in Italy. Annali Dellistituto Superiore Di Sanita 2014;50(2):163-170. [doi: 10.4415/ANN_14_02_10]
12. Thackeray R, Neiger BL, Smith AK, Van Wagenen SB. Adoption and use of social media among public health departments. BMC Public Health 2012 Mar 26;12:242 [FREE Full text] [doi: 10.1186/1471-2458-12-242] [Medline: 22449137]
13. Neiger B, Thackeray R, Burton S, Thackeray C, Reese J. Use of twitter among local health departments: an analysis of information sharing, engagement, and action. J Med Internet Res 2013 Aug 19;15(8):e177 [FREE Full text] [doi: 10.2196/jmir.2775] [Medline: 23958635]
14. Tursunbayeva A, Franco M, Pagliari C. Use of social media for e-Government in the public health sector: A systematic review of published studies. Government Information Quarterly 2017 Apr;34(2):270-282. [doi: 10.1016/j.giq.2017.04.001]
15. NHFPC. 2017 Aug 30. China's National Healthcare Institutes by the end of July in 2017: National Health and Family Planning Committee of the People Republic of China 2017 URL: http://www.nhfpc.gov.cn/mohwsbwstjxxzx/s7967/201708/d05b1b2964ac4780a2ba75b5aaa84b94.shtml [accessed 2018-07-23] [WebCite Cache ID 718a8sx5V]
16. Gan C, Wang W. Uses and gratifications of social media: a comparison of microblog and WeChat. J of Systems and Info Tech 2015 Nov 09;17(4):351-363. [doi: 10.1108/JSIT-06-2015-0052]
17. Li X, Huang J, Zhang H. An analysis of hospital preparedness capacity for public health emergency in four regions of China: Beijing, Shandong, Guangxi, and Hainan. BMC Public Health 2008 Sep 20;8:319 [FREE Full text] [doi: 10.1186/1471-2458-8-319] [Medline: 18803860]
18. Zhang T, Xu Y, Ren J, Sun L, Liu C. Inequality in the distribution of health resources and health services in China: hospitals versus primary care institutions. Int J Equity Health 2017 Dec 03;16(1):42 [FREE Full text] [doi: 10.1186/s12939-017-0543-9] [Medline: 28253876]
19. Ma L, Xie Q, Shi S, Ye X, Zhao A. Regional Maldistribution of China's Hospitals Based on Their Structural System. Sustainability 2017 Jun 16;9(6):1046. [doi: 10.3390/su9061046]

XSL•FO
RenderX

20. Démurger S, Sachs J, Woo W, Bao S, Chang G, Mellinger A. Geography, Economic Policy, and Regional Development in China. Asian Economic Papers 2002 Jan;1(1):146-197. [doi: 10.1162/153535102320264512]
21. Hua X, Johansson A, Wang X. National and regional financial openness in China. Journal of Chinese Economic and Business Studies 2016 Nov 28;15(2):127-140. [doi: 10.1080/14765284.2016.1261490]
22. Rogers E. Diffusion of Innovation. In: Diffusion of Innovations. New York: The Free Press; 1995.
23. Al-Amin M, Zinn J, Rosko MD, Aaronson W. Specialty hospital market proliferation: Strategic implications for general hospitals. Health Care Manage Rev 2010;35(4):294-300. [doi: 10.1097/HMR.0b013e3181e04a06] [Medline: 20844355]
24. Effing R, Spil T. The social strategy cone: Towards a framework for evaluating social media strategies. International Journal of Information Management 2016 Feb;36(1):1-8. [doi: 10.1016/j.ijinfomgt.2015.07.009]
25. Bertot J, Jaeger P, Hansen D. The impact of polices on government social media usage: Issues, challenges, and recommendations. Government Information Quarterly 2012 Jan;29(1):30-40. [doi: 10.1016/j.giq.2011.04.004]
26. Huang E. US hospitals on YouTube™: a test to the Altruistic marketing approach. Journal of Communication in Healthcare 2013 Nov 15;6(2):128-134. [doi: 10.1179/1753807613Y.0000000035]
27. Huang E, Liu T, Wang J. E-health videos on Chinese hospitals' websites. International Journal of Healthcare Management 2014 Jun 24;7(4):273-280. [doi: 10.1179/2047971914Y.0000000081]
28. Househ M, Borycki E, Kushniruk A. Empowering patients through social media: the benefits and challenges. Health Informatics J 2014 Mar;20(1):50-58. [doi: 10.1177/1460458213476969] [Medline: 24550564]

Abbreviations

HE: health education
HN: hospital news
NHCPRC: National Health Commission of the People's Republic of China
OD: official declaration
RQ: research questions

Edited by G Eysenbach; submitted 08.12.17; peer-reviewed by IN Gomez, J Richter, A Benis; comments to author 18.03.18; revised version received 26.05.18; accepted 18.06.18; published 09.08.18.

Please cite as:
Zhang W, Deng Z, Evans R, Xiang F, Ye Q, Zeng R
Social Media Landscape of the Tertiary Referral Hospitals in China: Observational Descriptive Study
J Med Internet Res 2018;20(8):e249
URL: http://www.jmir.org/2018/8/e249/
doi:10.2196/jmir.9607
PMID:30093370

©Wei Zhang, Zhaohua Deng, Richard Evans, Fei Xiang, Qing Ye, Runxi Zeng. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 09.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Therapist-Assisted Internet-Based Cognitive Behavioral Therapy Versus Progressive Relaxation in Obsessive-Compulsive Disorder: Randomized Controlled Trial

Michael Kyrios[1,2,3], PhD; Claire Ahern[3], PhD; Daniel B Fassnacht[2], PhD; Maja Nedeljkovic[3], PhD; Richard Moulding[4,5], PhD; Denny Meyer[6], PhD

[1]College of Education, Psychology & Social Work, Flinders University, Adelaide, Australia

[2]Research School of Psychology, The Australian National University, Canberra, Australia

[3]Department of Psychological Sciences, School of Health Sciences, Swinburne University of Technology, Melbourne, Australia

[4]School of Psychology, Faculty of Health, Deakin University, Geelong, Australia

[5]Centre for Drug Use, Addictive and Anti-Social Behaviour Research, Deakin University, Geelong, Australia

[6]Department of Statistics Data Science and Epidemiology, School of Health Sciences, Swinburne University of Technology, Melbourne, Australia

Corresponding Author:
Michael Kyrios, PhD
College of Education, Psychology & Social Work
Flinders University
Sturt Roadd, Bedford Park
Adelaide,
Australia
Phone: 61 8 8201 3157
Email: michael.kyrios@flinders.edu.au

Abstract

Background: Obsessive-compulsive disorder (OCD) is a highly disabling psychological disorder with a chronic course if left untreated. Cognitive behavioral therapy (CBT) has been shown to be an effective treatment, but access to face-to-face CBT is not always possible. Internet-based CBT (iCBT) has become an increasingly viable option. However, no study has compared iCBT to an analogous control condition using a randomized controlled trial (RCT).

Objective: A 2-armed RCT was used to compare a therapist-assisted 12-module iCBT to an analogous active attention control condition (therapist-assisted internet-based standard progressive relaxation training, iPRT) in adult OCD. This paper reports pre-post findings for OCD symptom severity.

Method: In total, 179 participants (117 females, 65.7%) were randomized (stratified by gender) into iCBT or iPRT. The iCBT intervention included psychoeducation, mood and behavioral management, exposure and response prevention (ERP), cognitive therapy, and relapse prevention; the iPRT intervention included psychoeducation and relaxation techniques as a way of managing OCD-related anxiety but did not incorporate ERP or other CBT elements. Both treatments included audiovisual content, case stories, demonstrations of techniques, downloadable audio content and worksheets, and expert commentary. All participants received 1 weekly email, with a maximum 15-minute preparation time per client from a remote therapist trained in e-therapy. Emails aimed to monitor progress, provide support and encouragement, and assist in individualizing the treatment. Participants were assessed for baseline and posttreatment OCD severity with the telephone-administered clinician-rated Yale-Brown Obsessive-Compulsive Scale and other measures by assessors who were blinded to treatment allocation.

Results: No pretreatment differences were found between the 2 conditions. Intention-to-treat analysis revealed significant pre-post improvements in OCD symptom severity for both conditions (P<.001). However, relative to iPRT, iCBT showed significantly greater symptom severity improvement (P=.001); Cohen d for iCBT was 1.05 (95% CI 0.72-1.37), whereas for iPRT it was 0.48 (95% CI 0.22-0.73). The iCBT condition was superior in regard to reliable improvement (25/51, 49% vs 16/55, 29%; P=.04) and clinically significant pre-post-treatment changes (17/51, 33% vs 6/55, 11%; P=.005). Those undertaking iCBT post completion of iPRT showed further significant symptom amelioration (P<.001), although the sequential treatment was no more efficacious than iCBT alone (P=.63).

XSL•FO
RenderX

Conclusion: This study is the first to compare a therapist-assisted iCBT program for OCD to an analogous active attention control condition using iPRT. Our findings demonstrate the large magnitude effect of iCBT for OCD; interestingly, iPRT was also moderately efficacious, albeit significantly less so than the iCBT intervention. The findings are compared to previous internet-based and face-to-face CBT treatment programs for OCD. Future directions for technology-enhanced programs for the treatment of OCD are outlined.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12611000321943; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=336704 (Archived by WebCite at http://www.webcitation.org/70ovUiOmd)

(J Med Internet Res 2018;20(8):e242) doi:10.2196/jmir.9566

KEYWORDS

obsessive-compulsive disorder; mental health; cognitive behavioral therapy; CBT; online intervention

Introduction

Obsessive-compulsive disorder (OCD) is a common and highly debilitating condition that is considered to be among the most disabling of the psychological disorders [1]. OCD is characterized by intrusive and persistent thoughts, images, or urges (obsessions) that cause distress and lead to repetitive and often ritualistic behaviors (compulsions) intended to reduce threat or discomfort [2]. Cultural and geographically diverse clinical and epidemiological data have shown remarkable consistency with respect to both the presence and characteristics of OCD symptoms [3]. Without appropriate treatment, the course of the illness is understood to be chronic and lifelong [4].

Face-to-face individual and group-based cognitive behavioral therapy (CBT) has been shown to be a highly effective treatment for OCD [5,6], leading to significant improvements in functioning and quality of life [7,8]. Olatunji et al [9] assert that CBT is effective regardless of gender, baseline severity or symptom subtype, comorbidities, or treatment length. Greater pretreatment OCD severity has been associated with larger posttreatment effect sizes of face-to-face CBT, although not consistently across all studies [9-12]. While other psychological treatments have been applied to OCD, inclusive of progressive relaxation [13-16] and acceptance and commitment therapy [17], these interventions lack an extensive evidence base for their efficacy. Pharmacological treatments have also been found to be effective in treating OCD, particularly serotonergic agents, although cognitive behavioral treatments are more effective among outpatients with OCD [18].

Despite the existence of effective interventions, it is reported that there is around a 7-year delay from the individual's first experiences of symptoms of OCD to their presenting for treatment [19,20]. People with the disorder may go undiagnosed for many years due to a failure of health professionals to recognize OCD [21] and because the individual does not disclose their experiences due to intense feelings of embarrassment and guilt [22-24]. For those who do present for help, access to treatment is poor. In particular, a shortage of appropriately qualified professionals (especially in geographically remote areas) along with long waitlists and individual financial constraints mean that only a small percentage of individuals with OCD receive CBT [24,25].

As a large proportion of people seek out mental health information from the internet [26,27], and in some cases feel more comfortable using technology than discussing their concerns in person [28-30], internet-based therapy could help bridge these gaps and make evidence-based treatments accessible and acceptable to individuals with OCD. Online treatments allow anonymity and are more accessible (particularly for geographically remote and rural areas), and depending on the model of therapy delivery, can be associated with reduced costs and allow the dissemination of standardized yet individualized treatments, providing the same content and skills as face-to-face equivalents [31].

While it is a relatively new area of investigation, initial findings suggest that iCBT is an effective modality for the treatment of OCD. A recent meta-analysis of remote CBT for OCD included 18 studies from which 7 were internet-administered interventions; however, only 2 studies were randomized controlled trials (RCTs) [32]. The author concluded that low- and high-intensity remote treatments for OCD lead to large magnitude improvements in OCD symptoms, are more effective than control conditions, and are not meaningfully different in efficacy from face-to-face treatment. In the largest RCT to date, Andersson et al [33] randomly allocated 101 participants to 10 weeks of therapist-guided iCBT or 10 weeks of online supportive therapy. The iCBT condition was associated with a significant reduction in OCD symptoms (Cohen d=1.55) as measured by the clinician-administered Yale-Brown Obsessive-Compulsive Scale (YBOCS) [34] compared to a medium within-group effect size (Cohen d=0.47) for the supportive therapy control. The results of iCBT were maintained at 4-month follow-up. Additionally, there was a large between-group effect size (Cohen d=1.12) on the YBOCS in favor of the iCBT condition. Finally, only 6% of participants in the control condition met criteria for clinically significant change compared with 60% of those receiving iCBT [33].

Similar results were demonstrated by Wootton et al [35], who found that 8-week courses of therapist-guided iCBT and therapist-guided bibliotherapy were both effective compared to a waitlist control condition (between-group effect sizes of Cohen d=1.57 and 1.40, respectively). Mahoney et al [36] conducted an RCT comparing clinically supervised technician-assisted iCBT to a treatment as usual control group. They found that iCBT was more efficacious than treatment as usual in reducing OCD severity, with iCBT demonstrating moderate-to-large

XSL•FO
RenderX

effect sizes at posttreatment and 3-month follow-up depending on which severity measure was used and whether completers or the total sample were used in analyses. They reported that 54% of treatment completers no longer met diagnostic criteria for OCD at follow-up. A more recent study from Korea using an internet-based CBT program for OCD reported significant improvement in OCD severity (measured with the clinician-administered YBOCS) from pre- to posttreatment (Cohen *d*=1.64); however, the study did not include a comparison group and only analyzed completers (64% of the sample) [37].

While these iCBT studies reported effect sizes that are somewhat similar to face-to-face therapy [11], therapist contact times differ markedly. A standard course of face-to-face CBT would be around 15 to 30 one-hour sessions [38,39]. In contrast, Wootton et al [35] reported a mean total therapist time of 1.72 hours for therapist-guided bibliotherapy and 1.48 hours for iCBT. Andersson et al [33] reported a mean therapist contact time of 2.15 hours for iCBT, although this was significantly higher than for their control condition (0.28 hours). However, the between-groups difference remained significant after therapist contact was statistically controlled.

While these studies show promise for iCBT, there were limitations in terms of the study designs for the 2 RCTs conducted to date. First, in Andersson et al [33], the control group comprised nondirective supportive therapy and lacked online modules. Similarly, the control condition in the study by Wootton et al [35] included no modules and no therapist time allocations, although a therapist-directed bibliotherapy group was included as an active control condition. As such, neither of the control groups in the 2 RCTs matched the active online treatment components in terms of format (ie, online self-help information), therapist contact, medium (ie, audiovisual content, downloadable worksheets, and other content), or therapeutic processes (eg, homework). Hence, research still needs to establish how iCBT compares to comparator conditions that serve as bona fide controls relative to active treatment.

Building upon these studies, our study aimed to evaluate therapist-assisted iCBT for OCD as compared to an analogous active control (therapist-assisted internet-based progressive relaxation therapy [iPRT]), allowing us to identify the additive effects of iCBT beyond the nonspecific consequences of anxiety management or expectation of change. At the completion of the iPRT, that group was administered the iCBT program. This paper reports pre-post iCBT treatment outcomes. Specifically, we used a new online CBT program for OCD and compared it to iPRT based on the protocol developed by Bernstein et al [40]. As justification, the most recent evaluation using manualized PRT has shown it to be effective and credible in treating OCD [17], although its efficacy has not been a consistent finding in the past [41,42]. More recently, an online applied relaxation program based on Öst [43] and very similar in content and structure to that of Bernstein et al [40] was found to be effective in the management of anxiety in panic disorder [44].

Consistent with our published study protocol [45], this paper reports primary outcomes of the study, namely pre-post change in OCD severity ratings and the proportion of participants experiencing clinically significant change. A second paper will report secondary outcomes. It was hypothesized that both groups would experience significant improvements in symptoms of OCD from pre- to posttreatment using an iPRT, with significantly greater improvements in iCBT. Specifically, we hypothesized that both the iCBT and iPRT groups would experience a reduction in symptom severity from pre- to postintervention with significantly greater improvement for the iCBT group. Participants who completed the control condition were anticipated to experience a further significant drop in OCD symptoms after completing iCBT. It was further expected that the proportion of participants experiencing clinically significant change in symptom severity would be greater for iCBT than iPRT.

Finally, the influence of sociodemographic and clinical variables was used to predict improvements in YBOCS scores. Given mixed results from previous studies and meta-analyses of face-to-face and remote CBT for OCD, we do not offer directional hypotheses regarding predictors of symptom improvement. Furthermore, evidence for predictors and moderators of outcome in internet-based interventions is still very limited [46]. Hence, we merely explored whether treatment gains were predicted by clinical variables (eg, symptom severity, depression scores, medication status) and sociodemographic measures (eg, gender, marital status).

Methods

Design

The study protocol is described elsewhere, inclusive of details about power analyses and measures [45]. There were minimal deviations from the protocol in the conduct of the trial; the major exception was that, due to difficulties in contacting and engaging participants, we were not able to undertake all the intended posttreatment assessments (eg, posttreatment structured diagnostic interviews). Rather, we focused on conducting as many telephone-administered OCD severity (YBOCS) interviews as possible. Subsequently, analyses are based on OCD severity and associated recovery (as defined below). Our evaluation framework was based on the work of Öst [47], who has developed a psychotherapy outcome study methodology rating form. In summary, the study conformed to Consolidated Standards of Reporting Trials (CONSORT) requirements [48]; participants were randomized post baseline interview into the active treatment (iCBT) or control condition (iPRT) using an independent automated computer-generated randomization sequence that could not be forecast or modified by the researchers. Stratified randomization was used to achieve gender balance between groups. The randomization was coordinated by an independent statistician. This paper compares pre- to posttreatment OCD outcomes using between- and within-group results. Posttreatment assessments were blind to treatment condition.

Participants, Recruitment, and Measures

Participants were recruited by referral from primary care physicians and mental health professionals and through self-referral. Information about the study was publicized on a webpage, on an affiliated online mental health treatment

XSL•FO
RenderX

webpage [49], on YouTube and via online advertisements on Facebook, and in mail-outs to Australian mental health professionals. Recruitment for the trial started in early 2013, preintervention assessments commenced in July 2013, and the last postintervention assessment was conducted in July 2014. From 1298 people who registered initial interest either online or via telephone, 238 participants provided consent and completed the pretreatment interview via telephone. Inclusion criteria were: (1) Australian resident; (2) aged 18 years or over; (3) fit the then-current *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision* (DSM-IV-TR) criteria for a primary diagnosis of OCD [50] where hoarding was not the primary symptom as assessed by the Structured Clinical Interview for DSM-IV Axis 1 Disorders, Clinician Version (SCID-CV) [51]; (4) no current psychosis, substance abuse, head injury, or neurological disorder; (5) no current active suicidal ideation or, if high risk (eg, history of suicidal behavior) then had appropriate psychosocial supports during the course of the trial; and (6) access to a computer. Participant flow is shown in Figure 1. Participant numbers were consistent with those anticipated from our initial power analysis [42] that expected a large effect for iCBT, a moderate effect size for iPRT, power at 80%, and a standard deviation of 5 for the distribution of mean change with significance set at 5% (alpha=.05). We did not collect the planned additional 20% of participants as the research funding period was ending. We ceased recruitment of participants after reaching numbers that were anticipated from the power analyses.

This paper focuses on the major outcome measures. Participants were enrolled by the study coordinator who allocated participants to a pretreatment assessor, the treatment condition on the basis of the randomization process, and an e-therapist. Applicants who met inclusion criteria (N=179) were assessed for baseline OCD severity with the telephone-administered clinician-rated YBOCS [34] and comorbid diagnoses were assessed with the SCID-CV [51]. The SCID-CV is the gold standard in structured diagnostic interviews, and the clinician-rated YBOCS has shown excellent reliability (eg, interrater reliability) and convergent validity with other measures of OCD [52]. Internal consistency for the YBOCS in our sample was acceptable (Cronbach alpha=.75). In addition to assess baseline depressive and anxiety symptoms and consistent with our previous research [53], the Hamilton Depression Rating Scale (HAM-D) [54] and the Hamilton Anxiety Rating Scale (HAM-A) [55] were used. Internal consistencies of the HAM-D (Cronbach alpha=.78) and HAM-A (Cronbach alpha=.79) were acceptable. Two questions were used to assess baseline treatment expectancies ("I believe this treatment is likely to be effective" and "I believe this treatment is likely to result in permanent improvement"); the questions were scored on a 5-point Likert scale ranging from strongly disagree (0) to strongly agree (4), and a composite score of the 2 items was calculated (Spearman-Brown coefficient=.80) to assess treatment expectancies in a subsample across the 2 intervention groups (96/179, 53.6% of the total sample).

Assessors (n=26) were either licensed psychologists or supervised students undertaking a masters or professional doctorate in clinical psychology. An experienced clinical psychologist trained all assessors in psychiatric diagnosis and structured interviews and reviewed all case inclusions. All assessors were trained through the use of videos and clients from the University Psychology Clinic until there was 100% consistent agreement with the senior assessor. Assessors were blind to treatment condition allocation. To examine the reliability between assessors, intraclass correlations (ICC) of YBOCS total scores were calculated; there was a high agreement between assessors (ICC=.94, 95% CI 0.68-0.99).

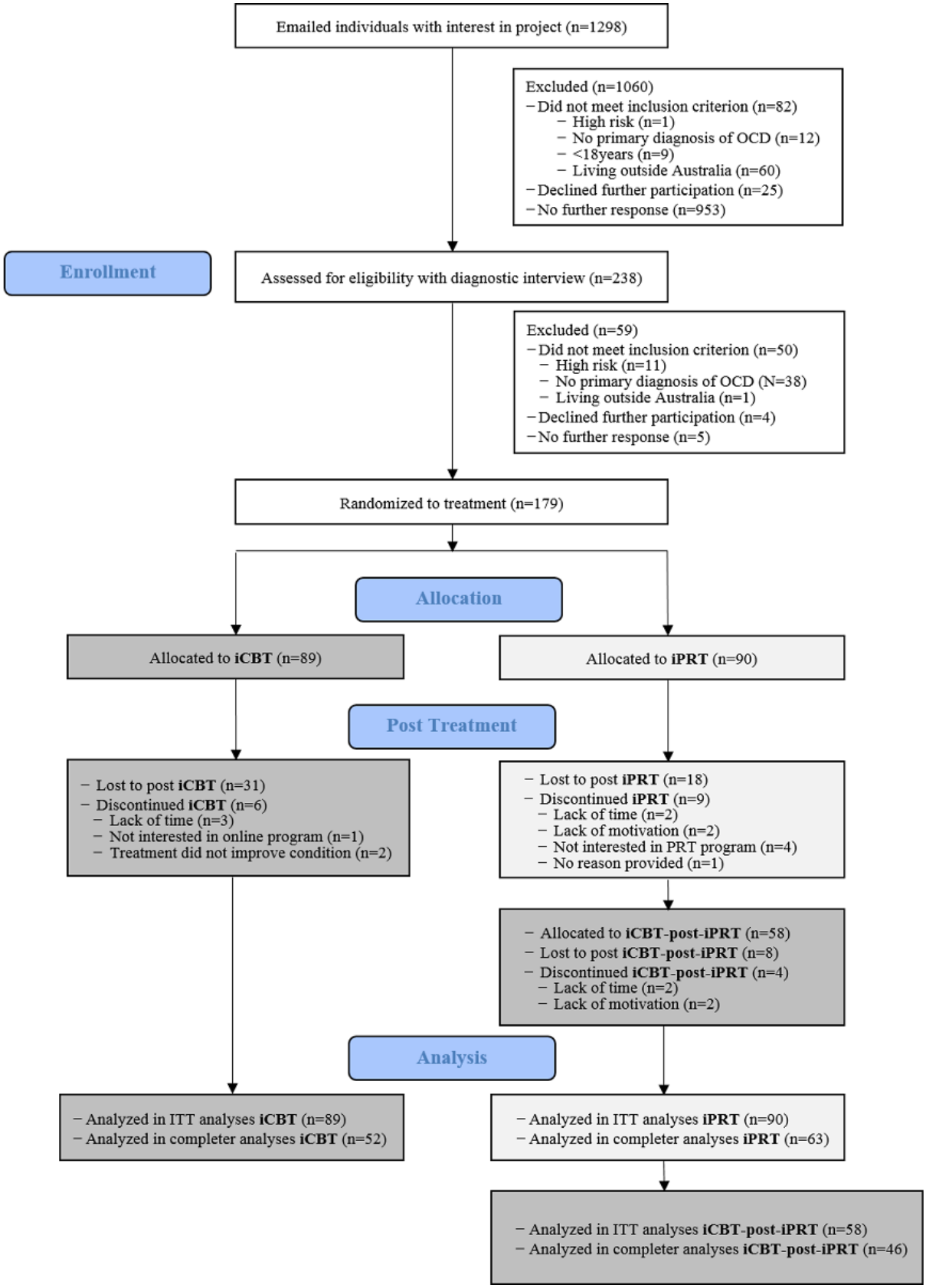
The overall sample was 65.7% female (117/179) and on average aged 33.4 (SD 9.9) years. The majority of participants had at least a bachelor degree (98/179, 54.7%) and were in de facto relationships or were married (87/179, 48.6%). On average, the sample reported having experienced OCD symptoms for almost 14 years (mean 13.7 [SD 0.8] years, range 1-49), with a current YBOCS rating of 21.94 (SD 0.49) and range 7-36. Almost half of the sample (84/179, 46.9%) reported moderate OCD symptoms (YBOCS 16-23), while around another third (62/179, 34.6%) indicated severe symptoms (YBOCS 24-31); the other participants reported either mild (YBOCS 8-15; 21/179, 11.7%) or extreme OCD symptoms (YBOCS 32-40; 9/179, 5.0%). At the start of the trial, approximately one-third (66/179, 36.9%) of participants were receiving another form of psychological treatment and two-thirds (124/179, 69.3%) were on medication, with no changes in these allowed during the trial and randomization meaning there were no differences at baseline between the 2 conditions (see Tables 1 and 2).

Interventions

Both intervention groups used their personal computer or laptop to access the intervention modules; weekly emails by the e-therapist were sent to the email address provided in the baseline assessment. There was no charge for treatment. Participants were not offered any direct incentives but had the opportunity to enter a raffle to win 1 of 3 tablet computers if they completed all assessments. This incentive was only offered midway through the trial as a way to increase questionnaire completion. Note that the incentive was not provided to complete the program but rather to complete the assessments. Previous research and our experience with delivery of treatment programs has indicated that assessment completion does not reflect program completion, with a large portion of individuals completing programs but not the final assessment [56,57].

XSL•FO
RenderX

Figure 1. Consolidated Standards of Reporting Trials flow diagram of study. iCBT: internet-based cognitive behavioral therapy; iPRT: internet-based progressive relaxation training; OCD: obsessive-compulsive disorder.



XSL•FO
RenderX

Table 1. Sociodemographic characteristics of the internet-based cognitive behavioral therapy (iCBT) and internet-based progressive relaxation therapy (iPRT) groups.

Characteristics	iCBT group		iPRT group		Test	*P* value
	n[a]	Value	n[a]	Value		
Age, years, mean (SD)	88	32.59 (9.86)	87	34.23 (9.88)	$F_{1,173}$=1.21	.27
# of children, mean (SD)	81	0.79 (1.17)	84	0.95 (1.33)	$F_{1,163}$=1.09	.41
Gender, n (%)	89		89		χ^2_1=0.03	.88
Male		31 (34.8)		30 (33.7)		
Female		58 (65.2)		59 (66.3)		
Education, n (%)	84		87		χ^2_2=2.02	.37
At most secondary school		38 (45.2)		35 (40.2)		
Bachelor degree		33 (39.3)		31 (35.6)		
Postgrad degree		13 (15.5)		21 (24.1)		
Marital status, n (%)	82		82		χ^2_2=1.73	.42
Married/de facto		43 (52.4)		44 (53.7)		
Never married		35 (42.7)		30 (36.6)		
Other		4 (4.9)		8 (9.8)		
Working, n (%)	67		68		χ^2_2=0.61	.74
No		11 (16.4)		13 (19.1)		
Part-time		18 (26.9)		21 (30.9)		
Full-time		38 (56.7)		34 (50.0)		

[a]Sample size varies due to missing data; N=179 randomized for treatment, n=89 allocated to iCBT, n=90 allocated to iPRT.

Both conditions comprised 12 modules delivered online over a 12-week period. Participants were encouraged to complete 1 module per week for the duration of treatment but all aspects of the program could be accessed from the beginning. Both conditions included online psychoeducational information, weekly homework tasks, downloadable worksheets, and audio files. All participants received a single email per week from a remote therapist irrespective of how many emails participants had sent; participants were free to email as often as they wished. If no emails were received from participants, 2 further reminder emails were sent. Emails to participants aimed to monitor progress, provide support and encouragement, and assist in tailoring the treatment to participants' problems, in line with the online content of their allocated condition. Therapist time spent on emails ranged from 0 to 15 minutes per week per participant during the treatment phase for each of the iCBT and iPRT conditions. If participants did not log into the treatment modules the previous week, they were reminded to do so by their e-therapist in the weekly emails.

Therapists (n=10) were either psychologists or students undertaking a masters or professional doctorate in clinical psychology and underwent an online training module for e-therapists working within the Mental Health Online platform [49]. An experienced clinical psychologist provided biweekly supervision of therapists to maintain adherence to intervention-relevant treatment integrity and ensure clinicians preserved adherence to the time limits expected for responding to emails.

iCBT (see Table 3 for program content) comprised 12 modules with psychoeducation about OCD, anxiety, and an introduction to CBT, along with mood and behavior management strategies. Exposure and response prevention (ERP) strategies were covered along with instructions and examples on how to construct a range of hierarchies and how to conduct ERP. Cognitive therapy techniques (eg, cognitive restructuring, behavioral experiments) targeting OCD-specific cognitive styles (such as inflated responsibility or overestimation of threat, importance/control of thoughts, perfectionism, and uncertainty intolerance beliefs) were also provided. Relapse preventions strategies (eg, problem solving, risk identification) were introduced toward the end of the treatment. The importance of daily practice was emphasized with a focus on enablers of and barriers to maintenance of CBT. Audiovisual content, case stories, demonstrations of techniques, downloadable audio content and worksheets, and expert commentary were provided throughout the program.

The control condition (see Table 3 for program content) was a 12-module iPRT program adapted from Bernstein et al [40]. Participants received basic information about OCD, on the relationship of OCD to anxiety, and about the use of relaxation as a way to manage anxiety. Individuals were taught to relax specific muscle groups while paying attention to sensations associated with both being tense or relaxed. Individuals were

XSL•FO
RenderX

instructed on how to achieve a state of deep relaxation in increasingly shorter periods and control excess tension in stress-inducing situations. Participants sequentially tensed and released muscle groups in order to achieve maximum states of relaxation, with the number of muscle groups decreasing over the 12 modules. Participants had access to downloadable audio and written material to guide their progressive relaxation training. ERP, cognitive therapy, and other CBT elements were not included in the iPRT program. Participants randomized into iPRT were aware that they would have the option of taking up iCBT at the end of their allotted condition (ie, iCBT post-iPRT).

Statistical Analyses

All analyses were conducted with SPSS Statistics version 22 (IBM Corp) using the YBOCS as the primary outcome measure. Group differences in demographic data and pretreatment measures were analyzed with 1-way analyses of variance and chi-square tests. Mixed-models analyses employing an autoregressive covariance structure and restricted maximum likelihood estimation were used to analyze changes in YBOCS scores from pre- to posttreatment in an intention-to-treat (ITT) analysis while controlling for age due to a significant age effect for attrition. Effect sizes (Cohen *d*) with 95% standardized confidence intervals were calculated for both within-group and between-group effects based on observed means and the pooled standard deviations.

The following criteria of clinical significance were used: a person was deemed to have made a reliable improvement if, at pretreatment, a score of 16 or greater was reported on the YBOCS and more than 6 units improvement were observed (ie, approximately 1 standard deviation) during treatment. A YBOCS score of 16 has traditionally been regarded as the cutoff score to indicate clinical significance in OCD trials [23], although some literature regards a cutoff of 14 as more appropriate [58,59]. We took the traditional approach (ie, using a cutoff of 16 on the YBOCS) in our main analysis but also examined the data using the more conservative approach.

Linear regression analyses were used to examine the effect of pretreatment YBOCS levels on the improvement of OCD severity for the 2 treatment groups separately. Additional regression analyses were performed using the total sample to test effects of sociodemographic variables and indicators of disability on YBOCS improvement levels.

Table 2. Baseline clinical characteristics of the internet-based cognitive behavioral therapy (iCBT) and internet-based progressive relaxation therapy (iPRT) groups.

Characteristic	iCBT group		iPRT group		Test	*P* value
	n[a]	Value	n[a]	Value		
YBOCS[b], mean (SD)	89	22.58 (5.53)	88	22.22 (5.76)	$F_{1,175}=0.18$	.67
Years since onset of OCD[c], mean (SD)	67	12.28 (9.16)	71	15.01 (10.49)	$F_{1,136}=2.64$	.11
GAF[d] scale, mean (SD)	80	57.41 (8.00)	74	59.34 (10.74)	$F_{1,152}=1.63$	.20
HAM-D[e] scale, mean (SD)	88	10.78 (6.25)	87	9.98 (5.70)	$F_{1,173}=0.79$	.38
HAM-A[f] scale, mean (SD)	88	15.19 (8.82)	86	14.09 (7.61)	$F_{1,172}=0.78$	.38
# hospitalizations, mean (SD)	85	0.27 (0.79)	86	0.14 (0.62)	$F_{1,170}=2.44$	.12
Treatment expectancy, mean (SD)	50	2.45 (0.71)	46	2.38 (0.76)	$F_{1,94}=0.21$	.65
Other current psychological treatment, n (%)	76		75		$\chi^2_1=0.83$	.36
No		40 (52.6)		45 (60)		
Yes		36 (47.4)		30 (40)		
Current medication, n (%)	85		85		$\chi^2_1=0.48$	.49
No		25 (29.4)		21 (24.7)		
Yes		60 (70.6)		64 (75.3)		
Comorbidity, n (%)	82		78		$\chi^2_1=0.13$	.72
No		19 (23.2)		20 (25.6)		
Yes		63 (76.8)		58 (74.4)		

[a]Sample size varies due to missing data; N=179 randomized for treatment, n=89 allocated to iCBT, n=90 allocated to iPRT.

[b]YBOCS: Yale-Brown Obsessive-Compulsive Scale.

[c]OCD: obsessive-compulsive disorder.

[d]GAF: Global Assessment of Functioning.

[e]HAM-D: Hamilton Depression Rating Scale.

[f]HAM-A: Hamilton Anxiety Rating Scale.

XSL•FO
RenderX

Table 3. Content of the intervention modules for therapist-assisted internet-based cognitive behavioral therapy and progressive relaxation therapy.

Intervention and modules	Contents
iCBT[a]	
1-3	• Psychoeducation about OCD[b] and anxiety • Introduction to CBT • Mood management strategies (eg, activity scheduling)
4-6	• Exposure and response prevention strategies (eg, construction of fear hierarchies)
7-9	• Cognitive therapy techniques (eg, cognitive restructuring) targeting OCD-specific cognitive styles (eg, inflated responsibility or overestimation of threat, importance of control of thoughts)
10-12	• Relapse preventions strategies (eg, problem solving, risk identification, contingency management, and mindfulness techniques)
iPRT[c]	
1-3	• Psychoeducation about OCD and anxiety • Introduction to PRT • Sequential tensing and releasing of 16 muscle groups
4-5	• Sequential tensing and releasing of 7 muscle groups
6-7	• Sequential tensing and releasing of 4 muscle groups
8-11	• Releasing of muscle groups without tension component (relaxation through recall)
12	• Mental summary of previously learned techniques

[a]iCBT: internet-based cognitive behavioral therapy.

[b]OCD: obsessive-compulsive disorder.

[c]iPRT: internet-based progressive relaxation therapy.

Results

Examination of Covariates

All analyses were conducted with and without using pretreatment YBOCS, depression, and anxiety scores as covariates. Inclusion of covariates did not change the pattern of results; thus, results not including covariates are reported. In addition, Pearson correlations were conducted between pretreatment anxiety and depression and posttreatment YBOCS scores, revealing only a marginal influence of the covariates on posttreatment OCD severity; HAM-A r_{119}=.19 and HAM-D r_{120}=.14. Results were compared with those obtained using a completer sample. For the completer analyses, cases were used only if pre- and postintervention OCD severity data (ie, YBOCS) was available. Little's MCAR test supports the assumption of data missing completely at random for pre- and postintervention YBOCS, HAM-D, and HAM-A across the 2 conditions indicated that data were missing at random (χ^2_1=1.02, *P*=.31).

Baseline Differences and Completers

There were no significant differences between the groups on demographic or mental health variables (all *P*>.05; Tables 1 and 2). In addition, there were no significant differences in treatment expectancies ($F_{1,94}$=0.21, *P*=.64; see Tables 1 and 2).

From the iCBT group, 7% of participants (6/89) discontinued the treatment compared to 10% of participants (9/90) in the iPRT group; this difference was not statistically significant (χ^2_1=0.62, *P*=.43). Common reasons offered for ceasing the treatment in both conditions were other life commitments: 2 participants from the iCBT group dropped out because they did not find the treatment helpful, and 4 participants from the iPRT group indicated that the treatment was not specific to OCD, not effective, or that they would prefer to receive the CBT intervention (see Figure 1). Combining participants who discontinued treatment and/or did not complete the posttreatment assessment, there were more such participants in the iCBT group (37/89, 42%) compared to the iPRT group (27/90, 30%); however, this difference was not statistically significant (χ^2_1=2.61, *P*=.11). No significant differences (*P*>.05) were found between participants who did and did not complete posttreatment assessments on gender, education, marital status, number of hospitalizations, pretreatment YBOCS scores, whether participants had received any kind of treatment or medication in the month prior to commencing the study, or baseline treatment expectancies. However, there was a significant difference on age, with younger participants less likely to complete the posttreatment assessment ($F_{1,173}$=4.14, *P*=.04), making an ITT analysis advisable. Posttreatment completion rates increased on average by 3.4% for each year of age (95% CI 0.1%-6.7%).

XSL•FO
RenderX

Pre-Post Treatment Improvements

Means and standard deviations at pre- and posttreatment for the YBOCS are shown in Table 4, while Figure 2 shows the pattern of results. In order to test whether participants in the iCBT group showed greater improvement in OCD severity compared to improvements in the iPRT group, changes in YBOCS scores from pre- to posttreatment across the 2 treatment groups (ie, iCBT vs iPRT) were analyzed while controlling for age.

No negative experiences were reported at posttreatment by participants for the iCBT or iPRT interventions. Results for the ITT sample showed a significant time × group interaction effect ($F_{1,114}$=11.75, P=.001) and a significant main effect for time ($F_{1,148}$=83.52, P<.001); however, there was no significant main effect of group ($F_{1,180}$=3.58, P=.06), suggesting that both treatment groups improved over time but the iCBT group showed greater improvement compared to the iPRT group. These results were replicated in the completer sample: time × group interaction effect ($F_{1,107}$=6.91, P=.01), main effect for time ($F_{1,107}$=110.05, P<.001), and main effect for group ($F_{1,109}$=3.61, P=.06).

Paired t tests were conducted to test within-group improvements on the YBOCS. A t test comparing pre- and post-iCBT scores was statistically significant (t_{56}=7.90, P<.001); on average participants in the intervention group (iCBT) improved by 6.40 units on the YBOCS (95% CI 4.78-8.03). A second t test comparing pre- and post-iPRT scores was also significant (t_{66}= 3.92, P<.001); on average participants in the control group (iPRT) improved by 2.90 units on the YBOCS (95% CI 1.43-4.38). There were large improvements in OCD severity from pre- to posttreatment in the iCBT group and medium-to-large improvements in the iPRT group (see Table 4).

Post-Internet–Based Progressive Relaxation Therapy Improvement

In order to compare overall improvement in OCD severity between the iCBT cohort and the group undertaking iPRT followed by iCBT (iCBT–post-iPRT), changes in YBOCS scores from pretreatment (either condition) to post-iCBT treatment were analyzed while controlling for age: ITT analysis indicated a significant main effect for time ($F_{1,152}$=116.31, P<.001); however, there was no significant interaction effect for time × group ($F_{1,152}$=0.23, P=.63), suggesting that the combined iPRT/iCBT condition was no more effective than the iCBT condition on its own.

A t test comparing pre-iCBT and post-iCBT scores in the group that had previously received the iPRT intervention (iCBT–post-iPRT) was significant (t_{48}=4.03, P<.001); after having received the iPRT treatment, participants who continued with the iCBT treatment improved on average by 3.14 units on the YBOCS (95% CI 1.56-4.72). Moderate improvements were found from pre-iCBT to post-iCBT for the group that had previously received the iPRT treatment (iCBT–post-iPRT) (Table 5). Figure 2 presents the estimated means across the 2 treatments including standard error bars.

Reliable Improvement and Reliable Recovery

Using the previously discussed definition for reliable improvement (ie, at least a 6-unit YBOCS change) and reliable recovery (reliable improvement plus YBOCS below 16 at posttreatment), for the ITT analysis there were statistically significant differences between the 2 treatments for both of these variables (Table 6). Of the people with YBOCS scores of at least 16 prior to the iCBT treatment, 49% (25/51) showed an improvement of at least 6 units and 33% (17/51) made a reliable recovery. For the iPRT treatment, 29% (16/55) of the people with a YBOCS score of at least 16 prior to treatment showed an improvement of more than 6 units, however, only 11% (6/55) made a reliable recovery. Note that only 1 participant with a YBOCS score initially below 16 showed an improvement of at least 6 units. Using more conservative criteria defined by Fisher and Wells [58] (YBOCS cutoff <14 and reliable change of YBOCS >10), 18% (9/51) of the iCBT were considered to have been reliably recovered compared to 6% (3/55) of the iPRT group; this difference was statistically significant (χ^2_1=3.92, P=.048).

We also identified 4 participants who had deteriorated at posttreatment (ie, deterioration of at least 6 units in YBOCS severity scores and an overall posttreatment YBOCS score above 16): 3 participants in the iPRT and 1 participant in the iCBT condition. Demographic variables were comparative to the rest of the sample (eg, age, education, gender); however, pretreatment severity scores tended to be lower compared to the rest of the sample (YBOCS mean 18.75 [SD 6.50], HAM-D mean 8.50 [SD 7.14], HAM-A mean 11.00 [SD 10.03]).

Prediction of Obsessive-Compulsive Disorder Severity Improvement Using Pretreatment Characteristics

Using regression analyses, the effect of pretreatment YBOCS scores on levels of improvement was compared for the iCBT and iPRT treatments. For the iCBT treatment, there was a significant positive relationship between pretreatment severity and improvement levels (t_{55}=2.37, P=.02) with an average improvement of 0.58 units for each additional unit on the pretreatment YBOCS. For the iPRT treatment, there was also a significant positive relationship between pretreatment severity and improvement levels (t_{65}=2.59, P=.01) with an average improvement of 0.43 units for each additional unit for the pretreatment YBOCS. However, this difference in average improvements between the 2 treatments was not significant (t_{121}=0.90, P=.37). These results suggest that both treatments are more effective for individuals with higher initial OCD severity.

XSL•FO
RenderX

Table 4. Changes in the Yale-Brown Obsessive-Compulsive Scale scores from pre- to postintervention.

Analysis	Preintervention		Postintervention		Effect size, Cohen *d* (95% CI)	
	mean (SD)[a]	SE[b]	mean (SD)	SE	Within	Between
ITT[c] analysis						0.55 (0.18-0.91)
iCBT[d] YBOCS[e] score (n=89)	22.44 (5.36)	.61	15.86 (5.65)	.74	1.05 (0.72-1.37)	
iPRT[f] YBOCS score (n=90)	22.13 (5.73)	.62	19.15 (6.45)	.69	0.48 (0.22-0.73)	
Completer analysis						0.57 (0.18-0.95)
iCBT YBOCS score (n=52)	22.18 (5.61)	.81	15.26 (5.01)	.81	1.24 (0.87-1.60)	
iPRT YBOCS score (n=63)	22.64 (6.02)	.62	18.49 (6.35)	.76	0.78 (0.49-1.08)	

[a]Estimated marginal mean and standard deviation based on a mixed model analysis.

[b]SE: standard error.

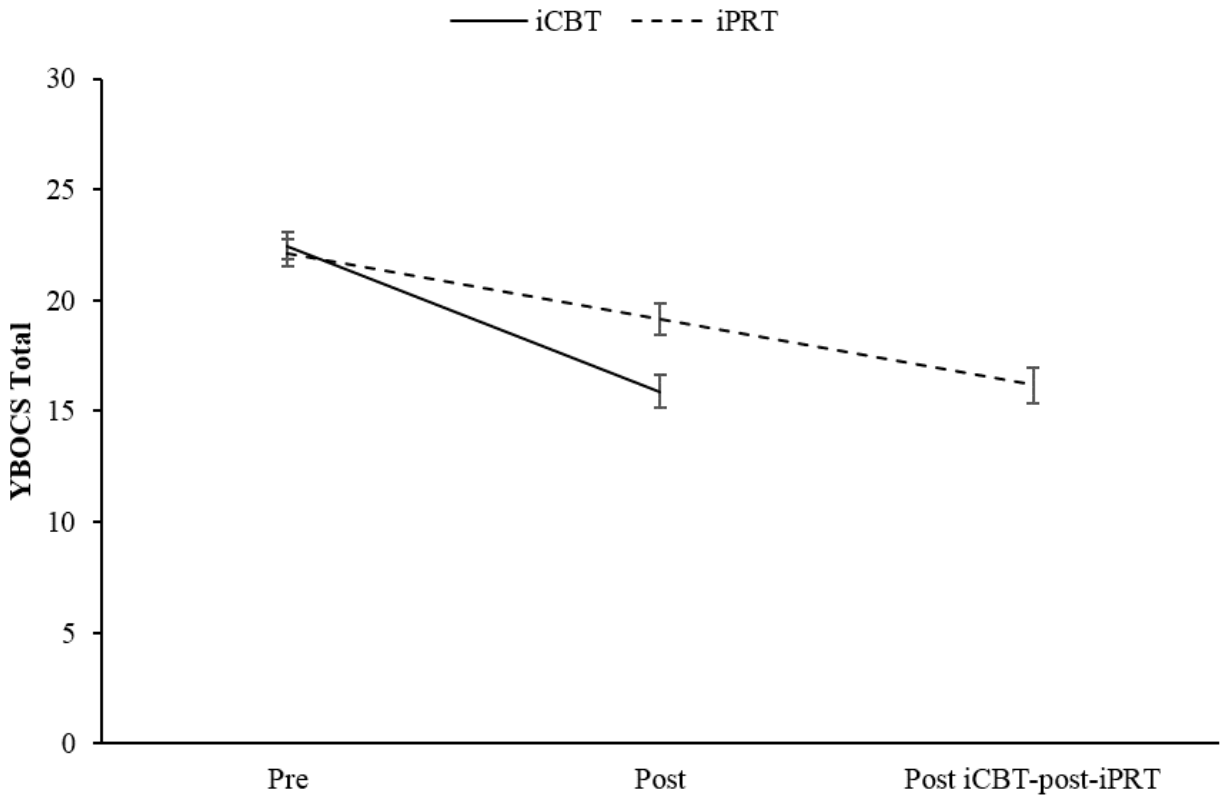
[c]ITT: intention-to-treat.

[d]iCBT: internet-based cognitive behavioral therapy.

[e]YBOCS: Yale-Brown Obsessive-Compulsive Scale.

[f]iPRT: internet-based progressive relaxation therapy.

Figure 2. Pre- and post-internet–based cognitive behavioral and progressive relaxation therapy estimated means including standard error bars. iCBT: internet-based cognitive behavioral therapy, iPRT: internet-based progressive relaxation therapy, YBOCS: Yale-Brown Obsessive-Compulsive Scale.



XSL•FO
RenderX

Table 5. Changes in the Yale-Brown Obsessive-Compulsive Scale scores for participants in the progressive relaxation versus cognitive behavioral therapy groups.

Analysis	Post-iPRT[a]		Post-iCBT (post-iPRT)[b]		Effect size (Within), Cohen *d* (95% CI)
	mean (SD)[c]	SE[d]	mean (SD)	SE	
ITT[e] analysis (n=58)	19.15 (6.45)	.69	16.16 (6.89)	.81	0.55 (0.21-0.88)
Completer analysis (n=46)	18.49 (6.35)	.76	15.26 (5.01)	.78	0.55 (0.22-0.88)

[a]Post-iPRT: postintervention internet-based progressive relaxation therapy.

[b]Post-iCBT (post-iPRT): postintervention internet-based cognitive behavioral therapy (post progressive relaxation therapy).

[c]Estimated marginal mean and standard deviation based on a mixed model analysis.

[d]SE: standard error.

[e]ITT: intention-to-treat.

Table 6. Reliability of improvement and recovery for intention-to-treat analysis at posttreatment.

Treatment	Reliable improvement			Reliable recovery		
	n (%)	χ^2	*P* value	n (%)	χ^2	*P* value
iCBT[a] (n=51)	25 (49)	4.43	.04	17 (33)	7.83	.01
iPRT[b] (n=55)	16 (29)	N/A[c]	N/A	6 (11)	N/A	N/A

[a]iCBT: internet-based cognitive behavioral therapy.

[b]iPRT: internet-based progressive relaxation therapy.

[b]N/A: not applicable.

Finally, separate regression analyses were performed using the total sample and the iCBT and iPRT groups. In the first regression model using the total sample, sociodemographic variables (gender, age, number of children, education, and marital status) were used to predict improvements in YBOCS scores. None of the sociodemographic variables explained a significant amount of variance in YBOCS improvement scores ($F_{5,96}$=0.42, *P*=.84). In a second regression model, indicators of disability (pretreatment global assessment of functioning, depression, anxiety scores, medication, number of hospitalizations) were used as predictors for YBOCS change scores in a subgroup of participants for whom full disability data sets were available, and none of these indicators explained a significant amount of variance in YBOCS improvement scores ($F_{5,95}$=2.06, *P*=.08). However, note that due to an administrative error, sociodemographic and disability data were not available for all participants. Nonetheless, there were no differences in pretreatment YBOCS scores between participants with and without missing sociodemographic (t_{175}=1.28, *P*=.20) and disability data (t_{175}=0.62, *P*=.53) used in the regression analyses.

In line with the results for the total sample, none of the sociodemographic variables or the indicators of illness severity were significant predictors for YBOCS change scores in the iCBT or iPRT group except pretreatment depression scores predicting YBOCS improvement scores in the iPRT group (t_{45}=2.41, *P*=.02), suggesting that participants with higher initial depressive symptoms showed greater YBOCS improvement scores; with every point on the HAM-D, participants in the iPRT group improved on average by 0.5 points more on the YBOCS.

Discussion

Principal Findings

Access to evidence-based treatment for psychological disorders can be facilitated by the digital revolution and the advent of online CBT, which is seen as an important component of contemporary mental health policies [60]. Such access is particularly important for a mental health condition such as OCD, which requires specialized treatment and is prone to effects of low help-seeking due to shame and stigma associated with the disorder [61]. This RCT aimed to evaluate the effect of therapist-assisted iCBT for OCD compared to an analogous active iPRT condition. This paper reports pre-post findings, with upcoming papers reporting follow-up findings and patterns of use associated with outcome.

Our findings demonstrate the large magnitude effect of a therapist-assisted iCBT for OCD. The study also established that a structured iPRT was efficacious, albeit less so than the iCBT intervention. While the addition of iCBT sequentially immediately following iPRT led to further significant symptom amelioration, the combined treatment was no more efficacious than iCBT alone. A similar pattern of results was found when we examined reliable improvement (ie, at least a 6-unit YBOCS change) and reliable recovery (reliable improvement plus YBOCS below 16). The iCBT condition was superior to iPRT, with around half of those in the iCBT treatment making a reliable improvement and a third making a reliable recovery compared to only 29% and 11%, respectively, in the iPRT condition. An exploration of predictors of treatment response found that pretreatment OCD symptom severity was the only significant predictor of change. No sociodemographic or

XSL•FO
RenderX

psychopathology severity variables predicted improvement. The small number of participants who deteriorated presented with an interesting profile; they tended to have lower severity scores on OCD, depression, and anxiety. This is contrary to what one might expect; however, previous literature has not generally reported deterioration statistics, although the relevant samples have been very small. The characterization of participants undertaking internet-based therapies who deteriorate is an important future research question and will require greater power, given the small numbers.

The magnitude of symptom amelioration is largely commensurate with previous studies of iCBT treatment indicating large effect sizes [32,62], although posttreatment YBOCS scores in this study were slightly higher than those reported in previous trials, while pretreatment YBOCS were either on par or slightly higher [33,35]. Our study used a more rigorous design by including an active control group that was comparable in terms of amount of content, prescribed therapist time, and mode of delivery, providing strong support for the efficacy of the specific CBT interventions over and above the effect of more general factors such as therapist support and time in treatment. While an inactive control might have been a useful addition from a design perspective, ethical considerations precluded this. Note that previous comparisons of iCBT against treatment as usual indicated no significant effects for the inactive control [36]. Overall, the findings support the notion that iCBT is an effective treatment for OCD, a disorder characterized particularly by shame, stigma, delayed help-seeking, and poor access to expert treatment [21-25].

Recovery figures in this study of around a third for iCBT were in the lower range compared to those reported in previous face-to-face and online treatment studies [33,35,58,63,64], although the expected superiority of iCBT over iPRT in recovery was supported. The slightly higher posttreatment YBOCS scores in our study certainly account for our findings. The nature of participants and recruitment strategies may partially explain these findings. Participants in this study were chronic in their presentation as indicated by over three-quarters reporting comorbidity, around 70% already on medications, and around half engaged in other forms of psychotherapy. While, on average, onset of OCD was reported as around 12 to 15 years, this compares to around 18 years in 1 similar study with better recovery rates [33]. Nonetheless, the literature generally asserts that around 25% to 70% of participants experience clinically significant change [33,35,58,63,64], although such definitions of recovery vary greatly between studies in the OCD area, and there has been a call to develop standard criteria [6,11]. One way around this is to use structured diagnostic interviews to assess recovery from diagnostic status, as originally intended, but participants were reticent to comply with the time required to undertake long interviews at posttreatment. Given revised DSM-5 criteria for OCD, future research will need to incorporate updated diagnostic interviews to assess recovery status.

Our results were based on ITT analyses of YBOCS severity data with available data biased toward older participants. It is possible that younger people are more transient or less likely to make themselves available for posttreatment assessments or they were impacted differentially more by the burden of multiple detailed assessments. Alternatively, they may have recovered more, may have deteriorated and dropped out of the study, or may have experienced decreased motivation to participate in further assessments due to the generally longer treatment (12 modules over 12 weeks for our study compared to 8 and 10 weeks for 5 and 10 modules in previous RCTs [35,33]). Nonetheless, treatment completion was high overall, supporting the degree to which internet-based treatments engage participants. Hence, while it is possible that effects were over- or underestimated, the large magnitude effect of iCBT is consistent with prior findings.

Previous literature has found mixed results with respect to predictors of treatment response [6,9-11,65]. On the one hand, that sociodemographic variables and disability did not predict outcome in this trial of an online treatment for OCD was encouraging in terms of suggesting the general utility of the intervention. However, it also leaves us none the wiser as to predicting which demographic group might most benefit from this intervention. Future research will need to examine participants from a broader range of social, educational, and cultural backgrounds and disability, chronicity, and comorbidity profiles in order to assess whether there are variables tied to differential effectiveness of the intervention. Future studies may also need to examine predictions away from the more controlled context of RCTs.

That greater pretreatment OCD symptom severity is a significant predictor of better outcome in internet-based and face-to-face treatments is consistent with much of the previous literature [10-12,65]. This finding is no surprise as those with greater symptom severity have greater scope for larger magnitude symptom amelioration. Nonetheless, it was encouraging to note that the results suggest that both iCBT and iPRT were effective for those with higher initial OCD severity. On the surface, this may contrast with a generally held expectation, embedded within national mental health policies and practice guidelines [60], that online treatments should be used to target only mild severity presentations. However, an examination of mean severity scores across studies suggests that those presenting for online treatments are generally in the mild-to-moderate severity range, with mean clinician-rated YBOCS scores of around 21 to 25 [32], in contrast to a slightly higher but broader range of scores (17 to 29) for face-to-face individual and group treatment studies [11].

Future research will benefit from further examining the differential effectiveness of treatment for different OCD symptom profiles. Little previous research has examined differences in outcome among different OCD subtypes, although the emergence of new measures such as the Dimensional Obsessive-Compulsive Scale [66] will allow such examination. Examination of subtype performance in online studies would be particularly useful. For instance, do individuals with obsessional presentations respond less well than do those with compulsions? Are individuals with contamination and washing presentations more responsive to the structured approach that is inherent in online treatments than are, say, individuals with obsessional checking or those with obsessional slowness? Such insights would allow the development of more targeted treatment guidelines.

XSL•FO
RenderX

One of the more interesting findings from this trial was that iPRT is moderately efficacious if embedded within a framework of managing anxiety in OCD situations, although results may have been influenced by a biased sample that maintained adherence in order to undertake iCBT. While older trials had previously concluded that the efficacy of PRT was limited for OCD [41], a more recent study by Twohig et al [17] reported that PRT had a large magnitude effect in an RCT, raising the possibility that PRT is more efficacious than had been believed in the treatment of OCD. For ethical and practical purposes, we embedded the PRT within the context of coping with anxiety in OCD-relevant situations rather than merely engaging participants in decontextualized progressive muscular relaxation training, meaning that participants may have found the information to be more personally relevant than in some PRT protocols. A recent online applied relaxation program was found to be effective in the management of anxiety in panic disorder [44]; hence, the anxiety management component of iPRT may have moderate but significant specific benefits in OCD.

Regardless of the reasons, while there was satisfactory adherence and relatively good effects for iPRT, these were still limited relative to iCBT, and the combined iPRT/iCBT did not perform any better than iCBT alone. Nonetheless, while guidelines to use CBT over interventions such as PRT are supported by the current findings, iPRT may prove useful in cases where the individual's capacity or willingness to undertake CBT is compromised. Future studies could examine the characteristics of affected individuals who respond specifically to the use of iPRT.

Future research will also need to pay greater attention to mechanisms of change. What is the impact on outcome of specific components of treatment protocols, treatment length or integrity, amount of time spent by assessors in interviewing participants, content of messages sent by therapists in responding to participant emails, and markers of degree of engagement with or by therapists? In this study, therapists were instructed to spend no more than 15 minutes per participant in writing responses to emails (ie, maximum 180 minutes per participant over 12 weeks). While some participants did not use the opportunity to write to their therapist, others wrote multiple emails with copious details. Investigating the impact of markers of engagement (eg, number of emails sent to therapist, number of words written in emails) could provide greater insights into mechanisms of change. Alternatively, examining therapeutic alliance and time spent on the modules and degree of homework adherence [67] could also inform about treatment processes. As our previous examination of an automated version of this iCBT indicated medium effect sizes in an uncontrolled naturalistic study [68], factors related to therapist assistance are likely to be important in facilitating greater efficacy in online programs. A following paper will examine some of these issues.

Given the evidence for the efficacy of this intervention alongside other RCTs using iCBT for OCD, we believe that such interventions hold promise as a routine treatment for individuals with OCD, particularly those with similar symptoms as in this study (eg, a majority with mild-moderate symptoms). For example, Andrews and Williams [69] note that 19 out of 20 individuals in their clinic, when given the option, chose iCBT over traditional CBT, meaning the iCBT became the standard treatment. This would be particularly the case for disorders such as OCD, where knowledge of treatment in the health care community has been low [22] and where availability of treatment is consequently limited [24,25]. The flexibility of the protocol means that various treatment constellations could be used. For example, our treatment center (which is subsidized by the Australian Government Department of Health and Ageing) currently offers a free automatic version of the therapy as well as a low-cost therapist-assisted version. Alternatively, community-based clinicians could be trained to provide motivational support while prescribing online training modules as the main component of treatment and then providing limited or brief face-to-face or telehealth support where required (eg, if patients experience difficulties in implementing ERP). This could minimize costs and maximize the advantages of internet-based treatment irrespective of funding models for mental health services. Alternatively, online programs can be used within a stepped care framework, which has been found to be effective in the treatment of anxiety [70].

Limitations and Strengths

A number of limitations have already been discussed throughout the paper. These include the lack of an inactive control arm, high rate of dropout at posttreatment assessment, and available data being biased toward older participants. Nonetheless, based on the evaluation framework from Öst [47], our study has several strengths. Broad recruitment strategies were used and participants were only excluded if they met primary criteria for other major disorders or reported current active suicidal ideation; thus, we are confident that the sample constituted a good representation of patients seeking online treatment for OCD. Furthermore, the use of well-trained blind evaluators, structured interviews to establish initial clinical diagnosis, and psychometrically sound outcome measures added to the quality of our study. Although the clinician contact was limited to 1 email per week for each participant, all therapists were trained using an online training module for e-therapists and supervised biweekly by an experienced clinical psychologist trained in e-therapy who provided supervision of e-therapists ensuring ongoing checks for treatment adherence and therapist competence. Equality of treatments, bone fide nature of the control condition, randomization to treatment condition, equality of maximum therapist engagement, and control of concomitant treatments were all strengths of this RCT. Although ITT analyses were used to control for this, participant attrition from data collection was a distinct limitation and may have impacted effect sizes.

Conclusions

Overall, this paper established the large magnitude effect of a 12-module therapist-assisted iCBT program for OCD and the moderate magnitude effect of iPRT when embedded within the framework of coping with anxiety in OCD-relevant situations. The iCBT program was significantly more efficacious than iPRT, and the sequential addition of iCBT immediately following iPRT was no more efficacious than iCBT alone. Recovery rates were not as high as those reported in previous literature, although participant characteristics and recruitment

XSL•FO
RenderX

strategies may account partially for these findings. While only pretreatment OCD symptom severity predicted outcome, younger people were more likely to drop out before assessment. Nonetheless, the study supported iCBT as a useful form of treatment for OCD, a disorder characterized by shame, delayed help-seeking, and poor access to expert treatment. The integration of digitally delivered treatment options into health systems and policies therefore seems an important development in managing the mental health challenges of communities.

Acknowledgments

This research was supported by a National Health and Medical Research Council (NHMRC) project grant (620506). We wish to acknowledge the contribution of Assistant Professor David Austin and Professor Britt Klein (investigators on the NHMRC project), Dr Jacinda Cadman, Dr Kate Fitzgerald, Dr Serafino Mancuso, Dr Debra Osbourne, Mr Peter Fogarty, Ms Mietta Kyrios, and Dr Imogen Rehm (research fellows and assistants).

Authors' Contributions

MK led the research; MK, MN, and RM were involved in the conceptualization of the study, research design, and methodology and intervention development; MK, CA, DBF, and DM conducted analyses; CA was responsible for study recruitment and intervention dissemination; MK, CA, and DBF drafted this paper; and CA, DBF, MN, RM, and DM edited this paper. All authors read and edited drafts of the manuscript and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT - EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 7MB - jmir_v20i8e242_app1.pdf]

References

1. Niederauer KG, Braga DT, Souza FPD, Meyer E, Cordioli AV. [Quality of life in individuals with obsessive-compulsive disorder: a review]. Rev Bras Psiquiatr 2007 Sep;29(3):271-278 [FREE Full text] [Medline: 17713688]
2. Diagnostic and Statistical Manual of Mental Disorders. 5th Edition. Washington: American Psychiatric Association; 2013.
3. Nedeljkovic M, Moulding R, Foroughi E, Kyrios M, Doron G. Cultural issues in understanding and treating obsessive compulsive and spectrum disorders. In: Steketee G, editor. Oxford Handbook of Obsessive Compulsive and Spectrum Disorders. Oxford: Oxford University Press; 2012:1-66.
4. Koran LM, Thienemann ML, Davenport R. Quality of life for patients with obsessive-compulsive disorder. Am J Psychiatry 1996 Jun;153(6):783-788. [doi: 10.1176/ajp.153.6.783] [Medline: 8633690]
5. British Psychological Society, Royal College of Psychiatrists. Guidance: core interventions in the treatment of obsessive-compulsive disorder and body dysmorphic disorder. National Collaborating Centre for Mental Health, National Institute for Health and Clinical Excellence 2006:1. [Medline: 21834191]
6. McKay D, Sookman D, Neziroglu F, Wilhelm S, Stein DJ, Kyrios M, et al. Efficacy of cognitive-behavioral therapy for obsessive-compulsive disorder. Psychiatry Res 2015 Feb 28;225(3):236-246. [doi: 10.1016/j.psychres.2014.11.058] [Medline: 25613661]
7. Diefenbach GJ, Abramowitz JS, Norberg MM, Tolin DF. Changes in quality of life following cognitive-behavioral therapy for obsessive-compulsive disorder. Behav Res Ther 2007 Dec;45(12):3060-3068. [doi: 10.1016/j.brat.2007.04.014] [Medline: 17559803]
8. Hertenstein E, Thiel N, Herbst N, Freyer T, Nissen C, Külz AK, et al. Quality of life changes following inpatient and outpatient treatment in obsessive-compulsive disorder: a study with 12 months follow-up. Ann Gen Psychiatry 2013 Feb 22;12(1):4 [FREE Full text] [doi: 10.1186/1744-859X-12-4] [Medline: 23433285]
9. Olatunji BO, Davis ML, Powers MB, Smits JAJ. Cognitive-behavioral therapy for obsessive-compulsive disorder: a meta-analysis of treatment outcome and moderators. J Psychiatr Res 2013 Jan;47(1):33-41. [doi: 10.1016/j.jpsychires.2012.08.020] [Medline: 22999486]
10. Kyrios M, Hordern C, Fassnacht DB. Predictors of response to cognitive behaviour therapy for obsessive-compulsive disorder. Int J Clin Health Psychol 2015 Sep;15(3):181-190. [doi: 10.1016/j.ijchp.2015.07.003]
11. Öst L, Havnen A, Hansen B, Kvale G. Cognitive behavioral treatments of obsessive-compulsive disorder: a systematic review and meta-analysis of studies published 1993-2014. Clin Psychol Rev 2015 Aug;40:156-169. [doi: 10.1016/j.cpr.2015.06.003] [Medline: 26117062]

XSL•FO
RenderX

12. Diefenbach G, Wootton B, Bragdon L, Moshier S, Tolin D. Treatment outcome and predictors of internet guided self-help for obsessive-compulsive disorder. Behav Ther 2015 Nov;46(6):764-774 [FREE Full text] [doi: 10.1016/j.beth.2015.06.001] [Medline: 26520219]
13. Fals-Stewart W, Marks AP, Schafer J. A comparison of behavioral group therapy and individual behavior therapy in treating obsessive-compulsive disorder. J Nerv Ment Dis 1993 Mar;181(3):189-193. [Medline: 8445378]
14. Fineberg NA, Brown A, Reghunandanan S, Pampaloni I. Evidence-based pharmacotherapy of obsessive-compulsive disorder. Int J Neuropsychopharmacol 2012 Sep;15(8):1173-1191. [doi: 10.1017/S1461145711001829] [Medline: 22226028]
15. Külz AK, Landmann S, Cludius B, Hottenrott B, Rose N, Heidenreich T, et al. Mindfulness-based cognitive therapy in obsessive-compulsive disorder: protocol of a randomized controlled trial. BMC Psychiatry 2014 Nov 18;14:314 [FREE Full text] [doi: 10.1186/s12888-014-0314-8] [Medline: 25403813]
16. Strauss C, Rosten C, Hayward M, Lea L, Forrester E, Jones A. Mindfulness-based exposure and response prevention for obsessive compulsive disorder: study protocol for a pilot randomised controlled trial. Trials 2015 Apr 16;16:167 [FREE Full text] [doi: 10.1186/s13063-015-0664-7] [Medline: 25886875]
17. Twohig MP, Hayes SC, Plumb JC, Pruitt LD, Collins AB, Hazlett-Stevens H, et al. A randomized clinical trial of acceptance and commitment therapy versus progressive relaxation training for obsessive-compulsive disorder. J Consult Clin Psychol 2010 Oct;78(5):705-716 [FREE Full text] [doi: 10.1037/a0020508] [Medline: 20873905]
18. Romanelli RJ, Wu FM, Gamba R, Mojtabai R, Segal JB. Behavioral therapy and serotonin reuptake inhibitor pharmacotherapy in the treatment of obsessive-compulsive disorder: a systematic review and meta-analysis of head-to-head randomized controlled trials. Depress Anxiety 2014 Aug;31(8):641-652. [doi: 10.1002/da.22232] [Medline: 24390912]
19. Feinstein SB, Fallon BA, Petkova E, Liebowitz MR. Item-by-item factor analysis of the Yale-Brown Obsessive Compulsive Scale symptom checklist. J Neuropsychiatry Clin Neurosci 2003;15(2):187-193. [doi: 10.1176/jnp.15.2.187] [Medline: 12724460]
20. Rasmussen SA, Tsuang MT. Clinical characteristics and family history in DSM-III obsessive-compulsive disorder. Am J Psychiatry 1986 Mar;143(3):317-322. [doi: 10.1176/ajp.143.3.317] [Medline: 3953865]
21. Mavrogiorgou P, Siebers F, Kienast T, Juckel G. [Help-seeking behavior and pathways to care for patients with obsessive-compulsive disorders]. Nervenarzt 2015 Sep;86(9):1130-1139. [doi: 10.1007/s00115-015-4298-5] [Medline: 25968650]
22. Kyrios M, Moulding R, Jones B. Obsessive compulsive disorder: integration of cognitive-behaviour therapy and clinical psychology care into the primary care context. Aust J Prim Health 2010;16(2):167-173. [Medline: 21128579]
23. Marques L, LeBlanc NJ, Weingarden HM, Timpano KR, Jenike M, Wilhelm S. Barriers to treatment and service utilization in an internet sample of individuals with obsessive-compulsive symptoms. Depress Anxiety 2010 May;27(5):470-475. [doi: 10.1002/da.20694] [Medline: 20455248]
24. Torres AR, Prince MJ, Bebbington PE, Bhugra DK, Brugha TS, Farrell M, et al. Treatment seeking by individuals with obsessive-compulsive disorder from the british psychiatric morbidity survey of 2000. Psychiatr Serv 2007 Jul;58(7):977-982. [doi: 10.1176/ps.2007.58.7.977] [Medline: 17602015]
25. Blanco C, Olfson M, Stein DJ, Simpson HB, Gameroff MJ, Narrow WH. Treatment of obsessive-compulsive disorder by U.S. psychiatrists. J Clin Psychiatry 2006 Jun;67(6):946-951. [Medline: 16848654]
26. Burns JM, Davenport TA, Durkin LA, Luscombe GM, Hickie IB. The internet as a setting for mental health service utilisation by young people. Med J Aust 2010 Jun 7;192(11 Suppl):S22-S26. [Medline: 20528703]
27. Powell J, Clarke A. Internet information-seeking in mental health: population survey. Br J Psychiatry 2006 Sep;189:273-277 [FREE Full text] [doi: 10.1192/bjp.bp.105.017319] [Medline: 16946364]
28. Wallach HS, Safir MP, Bar-Zvi M. Virtual reality cognitive behavior therapy for public speaking anxiety: a randomized clinical trial. Behav Modif 2009 May;33(3):314-338. [doi: 10.1177/0145445509331926] [Medline: 19321811]
29. Wilson JAB, Onorati K, Mishkind M, Reger MA, Gahm GA. Soldier attitudes about technology-based approaches to mental health care. Cyberpsychol Behav 2008 Dec;11(6):767-769. [doi: 10.1089/cpb.2008.0071] [Medline: 18991533]
30. Zinzow HM, Britt TW, McFadden AC, Burnette CM, Gillispie S. Connecting active duty and returning veterans to mental health treatment: interventions and treatment adaptations that may reduce barriers to care. Clin Psychol Rev 2012 Dec;32(8):741-753. [doi: 10.1016/j.cpr.2012.09.002] [Medline: 23063627]
31. Lal S, Adair CE. E-mental health: a rapid review of the literature. Psychiatr Serv 2014 Jan 1;65(1):24-32. [doi: 10.1176/appi.ps.201300009] [Medline: 24081188]
32. Wootton BM. Remote cognitive-behavior therapy for obsessive-compulsive symptoms: a meta-analysis. Clin Psychol Rev 2016 Feb;43:103-113. [doi: 10.1016/j.cpr.2015.10.001] [Medline: 26494179]
33. Andersson E, Enander J, Andrén P, Hedman E, Ljótsson B, Hursti T, et al. Internet-based cognitive behaviour therapy for obsessive-compulsive disorder: a randomized controlled trial. Psychol Med 2012 Oct;42(10):2193-2203 [FREE Full text] [doi: 10.1017/S0033291712000244] [Medline: 22348650]
34. Goodman WK, Price LH, Rasmussen SA, Mazure C, Fleischmann RL, Hill CL, et al. The Yale-Brown Obsessive Compulsive Scale. I. development, use, and reliability. Arch Gen Psychiatry 1989 Nov;46(11):1006-1011. [Medline: 2684084]
35. Wootton BM, Dear BF, Johnston L, Terides MD, Titov N. Remote treatment of obsessive-compulsive disorder: a randomized controlled trial. J Obsessive-Compulsive Related Dis 2013 Oct;2(4):375-384. [doi: 10.1016/j.jocrd.2013.07.002]

XSL•FO
RenderX

36. Mahoney AEJ, Mackenzie A, Williams AD, Smith J, Andrews G. Internet cognitive behavioural treatment for obsessive compulsive disorder: a randomised controlled trial. Behav Res Ther 2014 Dec;63:99-106. [doi: 10.1016/j.brat.2014.09.012] [Medline: 25461784]
37. Seol S, Kwon JS, Kim YY, Kim SN, Shin M. Internet-based cognitive behavioral therapy for obsessive-compulsive disorder in Korea. Psychiatry Investig 2016 Jul;13(4):373-382 [FREE Full text] [doi: 10.4306/pi.2016.13.4.373] [Medline: 27482237]
38. Kahn DA, Docherty JP, Carpenter D, Frances A. Consensus methods in practice guideline development: a review and description of a new method. Psychopharmacol Bull 1997;33(4):631-639. [Medline: 9493472]
39. Tolin DF, Hannan S, Maltby N, Diefenbach GJ, Worhunsky P, Brady RE. A randomized controlled trial of self-directed versus therapist-directed cognitive-behavioral therapy for obsessive-compulsive disorder patients with prior medication trials. Behav Ther 2007 Jun;38(2):179-191. [doi: 10.1016/j.beth.2006.07.001] [Medline: 17499084]
40. Bernstein D, Borkovec T, Hazlett-Stevens H. New Directions in Progressive Relaxation Training: A Guidebook for Helping Professionals. Wesport: Prager Publishers; 2000.
41. Lindsay M, Crino R, Andrews G. Controlled trial of exposure and response prevention in obsessive-compulsive disorder. Br J Psychiatry 1997 Aug;171:135-139. [Medline: 9337948]
42. Fals-Stewart W, Schafer J. The treatment of substance abusers diagnosed with obsessive-compulsive disorder: an outcome study. J Substance Abuse Treat 1992 Sep;9(4):365-370. [doi: 10.1016/0740-5472(92)90032-J]
43. Öst L. Tillämpad avslappning [Applied Relaxation]. Stockholm: Repro HSC; 1997.
44. Carlbring P, Ekselius L, Andersson G. Treatment of panic disorder via the Internet: a randomized trial of CBT vs. applied relaxation. J Behav Ther Experimental Psychiat 2003 Jun;34(2):129-140. [doi: 10.1016/S0005-7916(03)00026-0]
45. Kyrios M, Nedeljkovic M, Moulding R, Klein B, Austin D, Meyer D, et al. Study protocol for a randomised controlled trial of internet-based cognitive-behavioural therapy for obsessive-compulsive disorder. BMC Psychiatry 2014 Jul 25;14:209 [FREE Full text] [doi: 10.1186/1471-244X-14-209] [Medline: 25062747]
46. Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
47. Ost L. Efficacy of the third wave of behavioral therapies: a systematic review and meta-analysis. Behav Res Ther 2008 Mar;46(3):296-321. [doi: 10.1016/j.brat.2007.12.005] [Medline: 18258216]
48. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c332 [FREE Full text] [Medline: 20332509]
49. Mental Health Online. 2017. URL: http://www.mentalhealthonline.org.au [accessed 2018-07-09] [WebCite Cache ID 70p0f6eIc]
50. Diagnostic and Statistical Manual of Mental Disorders. 4th Edition, Text Revision. Washington: American Psychiatric Association; 2000.
51. First M, Williams J, Spitzer R, Gibbon M. Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Clinical Trials Version. New York: New York State Psychiatric Institute; 2007.
52. Woody SR, Steketee G, Chambless DL. Reliability and validity of the Yale-Brown Obsessive-Compulsive Scale. Behav Res Ther 1995 Jun;33(5):597-605. [Medline: 7598684]
53. Purcell R, Maruff P, Kyrios M, Pantelis C. Cognitive deficits in obsessive-compulsive disorder on tests of frontal-striatal function. Biol Psychiatry 1998 Mar 01;43(5):348-357. [Medline: 9513750]
54. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960 Feb;23:56-62 [FREE Full text] [Medline: 14399272]
55. Hamilton M. The assessment of anxiety states by rating. Br J Med Psychol 1959;32(1):50-55. [Medline: 13638508]
56. Moulding R, Nedeljkovic M, Kyrios M, Osborne D, Mogan C. Short-term cognitive-behavioural group treatment for hoarding disorder: a naturalistic treatment outcome study. Clin Psychol Psychother 2017 Jan;24(1):235-244. [doi: 10.1002/cpp.2001] [Medline: 26750388]
57. Snyder CF, Aaronson NK, Choucair AK, Elliott TE, Greenhalgh J, Halyard MY, et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. Qual Life Res 2012 Oct;21(8):1305-1314. [doi: 10.1007/s11136-011-0054-x] [Medline: 22048932]
58. Fisher PL, Wells A. How effective are cognitive and behavioral treatments for obsessive-compulsive disorder? A clinical significance analysis. Behav Res Ther 2005 Dec;43(12):1543-1558. [doi: 10.1016/j.brat.2004.11.007] [Medline: 16239151]
59. Farris SG, McLean CP, Van MPE, Simpson HB, Foa EB. Treatment response, symptom remission, and wellness in obsessive-compulsive disorder. J Clin Psychiatry 2013 Jul;74(7):685-690 [FREE Full text] [doi: 10.4088/JCP.12m07789] [Medline: 23945445]
60. Reynolds J, Griffiths KM, Cunningham JA, Bennett K, Bennett A. Clinical practice models for the use of e-mental health resources in primary health care by health professionals and peer workers: a conceptual framework. JMIR Ment Health 2015 Mar;2(1):e6 [FREE Full text] [doi: 10.2196/mental.4200] [Medline: 26543912]
61. Steketee G. Oxford Handbook of Obsessive Compulsive and Spectrum Disorders. Oxford: Oxford University Press; 2012.
62. Dèttore D, Pozza A, Andersson G. Efficacy of technology-delivered cognitive behavioural therapy for OCD versus control conditions, and in comparison with therapist-administered CBT: meta-analysis of randomized controlled trials. Cogn Behav Ther 2015 Feb;44(3):190-211. [doi: 10.1080/16506073.2015.1005660] [Medline: 25705787]

XSL•FO
RenderX

63. Whittal ML, Thordarson DS, McLean PD. Treatment of obsessive-compulsive disorder: cognitive behavior therapy vs. exposure and response prevention. Behav Res Ther 2005 Dec;43(12):1559-1576. [doi: 10.1016/j.brat.2004.11.012] [Medline: 15913543]
64. Simpson HB, Zuckoff AM, Maher MJ, Page JR, Franklin ME, Foa EB, et al. Challenges using motivational interviewing as an adjunct to exposure therapy for obsessive-compulsive disorder. Behav Res Ther 2010 Oct;48(10):941-948 [FREE Full text] [doi: 10.1016/j.brat.2010.05.026] [Medline: 20609435]
65. Andersson E, Ljótsson B, Hedman E, Enander J, Kaldo V, Andersson G, et al. Predictors and moderators of Internet-based cognitive behavior therapy for obsessive-compulsive disorder: results from a randomized trial. J Obsessive-Compulsive Rel Disord 2015 Jan;4:1-7. [doi: 10.1016/j.jocrd.2014.10.003]
66. Abramowitz JS, Deacon BJ, Olatunji BO, Wheaton MG, Berman NC, Losardo D, et al. Assessment of obsessive-compulsive symptom dimensions: development and evaluation of the Dimensional Obsessive-Compulsive Scale. Psychol Assess 2010 Mar;22(1):180-198. [doi: 10.1037/a0018260] [Medline: 20230164]
67. Kazantzis N, Whittington C, Zelencich L, Kyrios M, Norton PJ, Hofmann SG. Quantity and quality of homework compliance: a meta-analysis of relations with outcome in cognitive behavior therapy. Behavior Therapy 2016 May;47(5):755-772. [doi: 10.1016/j.beth.2016.05.002]
68. Klein B, Meyer D, Austin DW, Kyrios M. Anxiety online: a virtual clinic: preliminary outcomes following completion of five fully automated treatment programs for anxiety disorders and symptoms. J Med Internet Res 2011;13(4):e89 [FREE Full text] [doi: 10.2196/jmir.1918] [Medline: 22057287]
69. Andrews G, Williams AD. Up-scaling clinician assisted internet cognitive behavioural therapy (iCBT) for depression: a model for dissemination into primary care. Clin Psychol Rev 2015 Nov;41:40-48. [doi: 10.1016/j.cpr.2014.05.006] [Medline: 25043445]
70. Ho FY, Yeung W, Ng TH, Chan CS. The efficacy and cost-effectiveness of stepped care prevention and treatment for depressive and/or anxiety disorders: a systematic review and meta-analysis. Sci Rep 2016 Dec 05;6:29281 [FREE Full text] [doi: 10.1038/srep29281] [Medline: 27377429]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision
ERP: exposure and response prevention
HAM-A: Hamilton Anxiety Rating Scale
HAM-D: Hamilton Depression Rating Scale
iCBT: internet-based cognitive behavioral therapy
iPRT: internet-based progressive relaxation training
ICC: intraclass correlation coefficient
ITT: intention-to-treat analysis
MCAR: missing completely at random
NHMRC: National Health and Medical Research Council
OCD: obsessive-compulsive disorder
RCT: randomized controlled trial
SCID-CV: Structured Clinical Interview for DSM-IV Axis 1 Disorders, Clinician Version
YBOCS: Yale-Brown Obsessive-Compulsive Scale

Edited by G Eysenbach; submitted 04.12.17; peer-reviewed by B Wootton, J Reynolds; comments to author 28.12.17; revised version received 18.05.18; accepted 21.06.18; published 08.08.18.

Please cite as:
Kyrios M, Ahern C, Fassnacht DB, Nedeljkovic M, Moulding R, Meyer D
Therapist-Assisted Internet-Based Cognitive Behavioral Therapy Versus Progressive Relaxation in Obsessive-Compulsive Disorder: Randomized Controlled Trial
J Med Internet Res 2018;20(8):e242
URL: http://www.jmir.org/2018/8/e242/
doi:10.2196/jmir.9566
PMID:30089607

©Michael Kyrios, Claire Ahern, Daniel B Fassnacht, Maja Nedeljkovic, Richard Moulding, Denny Meyer. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 08.08.2018. This is an open-access article distributed under

XSL•FO
RenderX

the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Game Addiction Scale Assessment Through a Nationally Representative Sample of Young Adult Men: Item Response Theory Graded–Response Modeling

Yasser Khazaal[1], MD; Kyrre Breivik[2], PhD; Joel Billieux[3,4], PhD; Daniele Zullino[1], MD; Gabriel Thorens[1], MD; Sophia Achab[1], MD, PhD; Gerhard Gmel[5], PhD; Anne Chatton[4], MSc

[1]Geneva University, Geneve, Switzerland

[2]Regional Centre for Child and Youth Mental Health and Child Welfare, Uni Research Health, Bergen, Norway

[3]Addictive and Compulsive Behaviours Lab, Institute for Health and Behaviour, University of Luxembourg, Esch-sur-Alzette, Luxembourg

[4]Department of Mental Health and Psychiatry, Geneva University Hospitals, Geneva, Switzerland

[5]Alcohol Treatment Center, Lausanne University Hospital, Lausanne, Switzerland

Corresponding Author:
Yasser Khazaal, MD
Geneva University
Grand Pré 70c
Geneve, 1202
Switzerland
Phone: 41 223725550
Fax: 41 223202840
Email: yasser.khazaal@hcuge.ch

Abstract

Background: The 7-item Game Addiction Scale (GAS) has been validated under standard confirmatory factor analysis and exhibits good psychometric properties. Whether this scale satisfies the necessary conditions for consideration by item response theory (IRT) modeling remains unknown. However, the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5) recently proposed criteria, in its section 3, to define internet gaming disorder (IGD) to promote research on this possible condition.

Objective: The objective of our study was to (1) analyze GAS in the context of IRT (graded-response) modeling; (2) investigate differential item functioning (DIF), a feature of IRT modeling, in 2 subsamples; and (3) contribute to the ongoing (IGD) debate related to the validity of the DSM-5 criteria using GAS items as a proxy.

Methods: We assessed 2 large representative samples of Swiss men (3320 French-speaking and 2670 German-speaking) with GAS.

Results: All items comprised high discrimination parameters. GAS items such as relapse, conflict, withdrawal, and problems (loss of interests) were endorsed more frequently in more severe IGD stages, whereas items related to tolerance, salience (preoccupation), and mood modification (escape) were endorsed more widely among participants (including in less severe IGD stages). Several DIF effects were found but were classified as negligible.

Conclusions: The results of the analyses partly support the relevance of using IRT to further establish the psychometric properties of the GAS items. This study contributes to testing the validity of the IGD criteria, although cautious generalization of our findings is required with GAS being only a proxy of the IGD criteria.

(J Med Internet Res 2018;20(8):e10058) doi:10.2196/10058

KEYWORDS

internet addiction; internet gaming disorder; internet gaming; item response theory; game addiction scale

XSL•FO
RenderX

Introduction

In recent years, growing concerns have been expressed concerning public health issues related to excessive internet use [1] and online gaming [2,3], leading to numerous studies and debate about the possible addictive characteristics of some behaviors associated with the excessive use of internet games [4-7]. Thus, it is crucial to better understand and screen for potential disorders such as internet gaming addiction.

Many tools have been developed to this end, including the Game Addiction Scale (GAS) by Lemmens et al [8]. GAS was created to measure the following 7 criteria: *salience*, *tolerance*, *mood modification*, *relapse*, *withdrawal*, *conflict*, and *problems*. Validation of GAS in 2 samples of Dutch adolescent gamers showed good psychometric properties. GAS was, subsequently, cross culturally validated with 2 independent samples from two linguistic regions in Switzerland [9]. Standard confirmatory factor analysis (CFA) results revealed that the scale behaves similarly in both regions except for one item (withdrawal). This item showed a lack of invariance.

Standard CFA and item response theory (IRT) are two popular methods for establishing measurement invariance. Although both approaches share a number of similarities, they differ in many ways [10]. For instance, standard CFA models account for the covariance between test items, whereas IRT models account for examinee item responses [11]. The main difference between these methods, however, is that the relationship between the latent construct and the true score at the item level is linear in the standard CFA framework but nonlinear in the IRT framework [10]. Indeed, standard CFA often uses linear regression, but IRT typically uses a logistic model to estimate the probability of various types of item responses and thus, to describe item functioning along a continuum [12]. Under IRT, the primary purpose of administering a psychometric test is to locate the person taking it on the latent trait scale. If such a latent trait measure can be obtained for each person taking the test, two goals can be achieved. First, the respondent can be evaluated for the severity of the characteristic of interest and second, respondents can be compared to assign severity grades [13] under the appropriate IRT model. Within the IRT family, the logistic graded-response model (GRM) is a cumulative probability model developed by Samejima [14] and designed for Likert-type items.

However, the use of traditional IRT modeling rests on the following three fundamental assumptions: unidimensionality, local independence, and monotonicity [15]. Unidimensionality means that the test measures only one dimension. Strongly related to unidimensionality, local independence means that the item should be uncorrelated after conditioning on the latent trait [16]. Finally, monotonicity means that the probability of endorsement of item response categories increases with higher levels of the latent trait. To the best of our knowledge, no study has tested GAS against the monotonicity assumption, although previous studies have reported inconsistent results regarding dimensionality and local independence. Although most studies have found support for a unidimensional factorial structure [8,17-19], this was not the case in a large Norwegian study [20], which reported a better fit for a correlated 2-factor structure that distinguished between what they interpreted as core and peripheral criteria items. Earlier work on the French and German validation of GAS conducted on the this sample reported a good fit to a unidimensional factor structure but only after allowing for the correlation of 6 error terms, which suggests some local dependencies. Of note, however, scales are rarely strictly unidimensional. Thus, it is more a matter of whether the data are adequately unidimensional to produce relatively unbiased parameters using an IRT model despite some multidimensionality [21].

Accordingly, the first aim of this study was to explore whether it is appropriate to analyze GAS using IRT modeling. IRT provides an interesting feature to investigate the equivalence in the meaning of subgroup items; when such equivalence does not hold for item parameters, it is called differential item functioning (DIF) [22]. In addition, such items are of concern because they present a potential threat to the validity of the test. Regarding the validation of GAS referred to earlier, the withdrawal item did not seem to operate equivalently for both linguistic regions [9]. Many hypotheses were invoked, including a lack of precision for this concept when applied to game use [23] and a statistically significant difference because of the large sample size. A potential limitation of the study was that only weak (equal loadings) and not strong invariance (equal loadings and intercepts or thresholds) was tested. In IRT terminology, measurement noninvariance differentiates between the nonuniform DIF (different discrimination parameter or loading) and uniform DIF (equal factor loading but different threshold). Hence, a further aim of this study was to investigate a possible DIF effect associated with the group membership within the IRT framework.

Considering the concerns and debates related to potential internet gaming addiction [24], the American Psychiatric Association recently published, in section 3 (not yet accepted conditions requiring further research) of the *Diagnostic and Statistical Manual for Mental Disorders*, *5th Edition* (DSM-5) [25], the diagnostic criteria for internet gaming disorder (IGD). IGD is defined as a "persistent and recurrent use of the internet to engage in games...leading to clinically significant impairment or distress...during the past 12 months as indicated by 5 or more out of 9 criteria." These criteria are borrowed from substance use disorder and gambling disorder criteria [26], and the adequacy of such adaptation was criticized [4,7,27-29]. In particular, high engagement in video games might not always be considered an addiction but might simply reflect elevated healthy involvement [30].

In the context of the debates related to the IGD criteria, this study aims, in addition to its primary aims, to contribute to the discussion using the data driven by the analyses on a representative sample of young adult men.

Methods

Participants and Procedure

The data in this study are part of a longitudinal study, the Cohort Study on Substance Use Risk Factors, designed to assess

XSL•FO
RenderX

substance and game use among young Swiss men. This study protocol was approved by the Lausanne University Medical School's Ethics Committee for Clinical Research, and we obtained written informed consent from participants. The recruitment was conducted in 3 of 6 national army recruitment centers covering 21 of 26 cantons in the French- and German-speaking regions in Switzerland. Considering that military service is mandatory for adult men in Switzerland, the sample could be considered representative of their gender and age group.

During the recruitment period (August 2010-November 2011), 15,074 men received a mandatory appointment with the army recruitment center. Of 87.87% (13,245/15,074) men who were informed about the study, 57.10% (7563/13,245) provided their written consent to participate. Questionnaires were thus sent to their private addresses to ensure complete confidentiality of participants. Overall, 79.20% (5990/7563) participants completed the assessments (3320 French-speaking and 2670 German-speaking).

Instrument: Game Addiction Scale

We assessed participants with the 7-item version of GAS [8] translated into French and German. Because playing video games is often associated with other internet gaming-related behaviors (eg, gaming-related forums or chats and game broadcasts on apps such as YouTube) and considering that this was a large sample with diverse internet use habits who played a variety of games, the original 7-item GAS was modified to include the assessment of internet and gaming behaviors. For instance, the item "Do you play games to forget about real life?" was modified to "Do you play games or spend time on the internet to forget about real life?" Each of the 7 items was preceded by the statement "During the last 6 months, how often..." and was scored on a 5-point Likert scale (1=never, 2=rarely, 3=sometimes, 4=often, and 5=very often).

GAS was developed before the publication of DSM-5 based on a model that maintains that all addictions consist of some components (eg, salience, mood modification, tolerance, withdrawal, conflict, and relapse) [31]. The scale, nonetheless, partially covers the DSM-5 IGD criteria [32] (Table 1) [8]. However, one of the DSM-5 criteria, "jeopardized or lost a relationship, job or educational or career opportunity," is not explicitly proposed by GAS. In addition, the GAS item "problems" related to the DSM-5 criterion of "continue despite problems" is, instead, worded in relation to a loss of interest as "Have you neglected important activities...?" (Table 1). Furthermore, the time frame used in this study was the past 6 months rather than the 1-year time frame proposed by DSM-5.

Statistical Analysis

In this study, we used GRM because it is suitable for ordered polytomous variables [14]. GAS is a polytomous-ordered categorical scale containing 7 survey questions that measure gaming addiction on the internet. The items are labeled as salience, tolerance, mood modification, relapse, withdrawal, conflict, and problems and are ranked on a 5-point Likert scale from 1 (*never*) to 5 (*very often*). In GRM, the following two types of parameters were estimated: the discrimination parameter and the difficulty parameter. Because GRM is an ordered logistic model, difficulty parameters of each item were naturally estimated in the increasing order. Furthermore, the probability of observing outcome *k* or higher for item *i* and person *j* is as follows:

$$\Pr(Y_{ij} \geq k \mid \theta_j) = \exp[\alpha_i(\theta_j - \beta_{ik})]/\{1+\exp[\alpha_i(\theta_j - \beta_{ik})]\} \text{ with } \theta j \sim N(0,1)$$

where α_i represents the discrimination of item *i*, β_{ik} is the *k* th cutoff point for item *i*, and θ_j is the latent trait of person *j*.

Each item varies in difficulty and shares the same discrimination parameter. Of note, the discrimination parameter (also called slope) is a measure of the differential capability of an item. A high discrimination parameter suggests that an item has a high ability to differentiate subjects. In practice, a high discrimination parameter value means that the probability of endorsing an item response increases more rapidly as the latent trait or severity increases [33].

When discrimination is high (and the item response function is steep), the item provides more information on the latent trait and the information is concentrated around item difficulty. Items with low discrimination parameters, however, are less informative, and the information is scattered along a greater part of the latent trait range. With a logistic model for the item characteristic curve (ICC), Baker [13] proposed the following different ranges of values to better interpret the discrimination parameter: 0=nondiscriminative power; 0.01-0.34=very low; 0.35-0.64=low; 0.65-1.34=moderate; 1.35-1.69=high; >1.70=very high; and + infinity=perfect.

In GRM, 2 types of parameters are estimated, the discrimination and the threshold parameters. The number of thresholds is equal to the outcome categories minus 1. In this study, we had 5 alternative responses yielding 4 thresholds. The item threshold in the GRM model refers to the level of the latent variable an individual needs to endorse the item with 50% probability [34]. In addition, we presented ICCs, which are graphical functions that represent the respondents' latent trait as a function of the probability of endorsing the item [35]. Subsequently, ICCs were transformed into item information curves (IICs), which are a mathematical way to compute how much information each ICC can provide. Finally, IICs were summed, in turn, to obtain the test information function (TIF), which informs how well the instrument can estimate person locations. Globally, the information plots indicate the amount of psychometric information at each point along a latent severity dimension [36].

Model Fit Analysis

Prior to fitting a traditional item response model, a few prerequisites must be checked for the assessment of model fit, notably the assumptions of unidimensionality, local independence, and monotonicity. The flowchart in Figure 1 shows the steps leading to the use of IRT modeling.

XSL•FO
RenderX

Table 1. Game Addiction Scale (GAS).

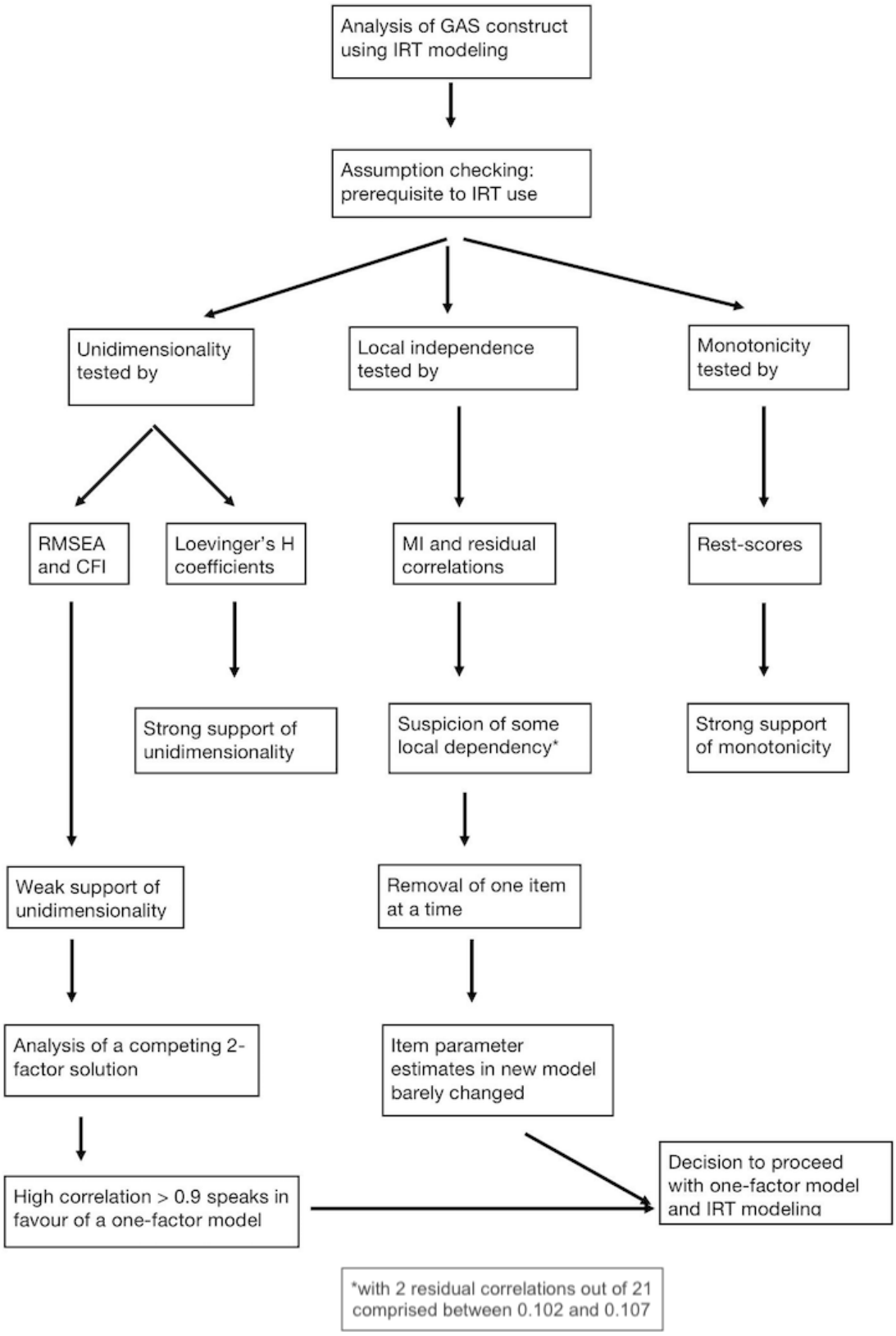
How often in the last 6 months...	Answer options[a], %					GAS items	DSM-5[b] criteria
	1	2	3	4	5		
Have you thought all day long about playing a game or spending time on the internet?						Salience	Preoccupation
All samples	48.5	24.8	15.0	7.2	4.4		
French	45.5	23.5	16.4	8.8	5.8		
German	52.3	26.4	13.2	5.3	2.7		
Have you played or stayed on the internet longer than intended?						Tolerance	Tolerance
All samples	36.0	21.4	24.9	12.9	4.8		
French	31.9	20.8	27.1	14.0	6.3		
German	41.0	22.3	22.3	11.5	2.9		
Have you played games or spent time on the internet to forget about real life?						Mood modification	Escape
All samples	61.1	19.3	12.4	4.8	2.5		
French	62.7	17.9	11.9	4.6	2.9		
German	59.1	20.9	13.0	5.0	2.0		
Have others unsuccessfully tried to reduce your time spent on games or the internet?						Relapse	Unsuccessful attempts to stop or reduce
All samples	69.8	15.7	9.9	3.4	1.3		
French	68.5	15.8	10.5	3.6	1.7		
German	71.3	15.6	9.1	3.1	0.8		
Have you felt upset when you were unable to play or to spend time on the internet?						Withdrawal	Withdrawal
All samples	78.5	13.6	5.8	1.6	0.6		
French	79.4	12.9	5.5	1.6	0.6		
German	77.4	14.4	6.1	1.5	0.6		
Have you had arguments with others (eg, family and friends) over your time spent on games on the internet?						Conflict	Deceiving Others
All samples	75.6	14.3	7.4	1.9	0.9		
French	76.1	13.5	7.1	2.2	1.0		
German	75.1	15.2	7.7	1.4	0.6		
Have you neglected important activities (eg, school, work, and sports) to play games or spent time on the internet?						Problems	Loss of interests
All samples	70.0	17.0	9.2	2.6	1.3		
French	68.1	17.2	10.2	3.0	1.4		
German	72.2	16.7	8.0	2.1	1.0		

[a]1=never, 2=rarely, 3=sometimes, 4=often, 5=very often.

[b]DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th Edition.

XSL•FO
RenderX

Figure 1. Flowchart of the steps leading to the decision to use item response theory (IRT) modeling. CFI: comparative fit index GAS: game addiction scale; MI: modification indices; RMSEA: root mean square error of approximation.



Unidimensionality

The unidimensionality assumption suggests that the correlation among these items could be explained by a single latent factor; this assumption was assessed using 2 different approaches, the fit of a unidimensional model in the categorical confirmatory factor analysis (CCFA; declaring the data as ordinal using the weighted least square mean and variance-adjusted estimator and the Mokken scaling method, a nonparametric IRT model

XSL•FO
RenderX

following an adaptation of Loevinger's *H* coefficients [37]. When testing the unidimensional assumption for an IRT model, it is more appropriate to use CCFA than standard CFA because the former (similar to an IRT model) treats the data as categorical. In addition, the acceptable and good fit is indicated by the root mean square error of approximation (RMSEA) of <0.08 and <0.06, respectively, and the comparative fit index (CFI) values of >0.90 and >0.95, respectively [38,39]. Furthermore, the *H* coefficients express the degree of homogeneity of a set of items. When $0.3 \leq H < 0.4$, the scale is considered weakly unidimensional; when $0.4 \leq H < 0.5$, it is considered moderately unidimensional; and when $H > 0.5$ [40,41], the scale is considered strong.

Local Independence

In local independence, it is assumed that a person's responses to questions are not statistically related to each other when the latent trait is held constant [42], that is, the response to one item should not influence the response to another item. Moreover, because local independence is closely related to the unidimensionality assumption, some authors argued that when the latter is true, local independence is obtained [22,43]. However, we tested for local independence by evaluating the matrix of residual correlations resulting from the CCFA model. Notably, residual correlations that are >0.1 are indicative of a possible local dependence [44,45].

Monotonicity

The monotonicity assumption is met when the probability of endorsing a response to a test item is nondecreasing with an increase in the value of the latent construct [46]; this assumption was examined through the results of the check monotonicity function of the Mokken package. The minimum violation default value was set to 0.3, and violations greater than this value were reported. In addition, the rest-score graphs, computed as the raw scale score minus the item score for each item, also served to detect monotonicity violation patterns. Graphically, rest-scores are on the x-axis, and the proportion of respondents in each rest-score group endorsing the item is on the y-axis [47]. We used the Mokken package to plot these graphs in this study.

After we found out that the IRT assumptions were tenable, we proceeded with the estimation of the item parameters for the whole sample and the detection of a possible DIF effect by regressing the group membership on all test items and the latent symptom severity dimension.

Differential Item Functioning

In DIF analyses, we compared a model, in which the alpha and beta parameters were constrained to be equal for the relevant subgroups, with a model, in which the parameters were left to be free. In addition, DIF was evaluated across linguistic groups with the help of the Lordif package [48], which uses a hybrid iterative technique in an ordinal regression. Of note, this approach tests the null hypothesis that α_i is equal for the 2 linguistic regions (absence of the nonuniform DIF) and the null hypothesis that β_{ij} is equal (absence of the uniform DIF). Because the chi-square test is highly sensitive to sample size [49], we decided that the change in pseudo R^2 also had to be a minimum of 0.035 to be flagged as a nonnegligible DIF effect [50].

Missing Values

The data from which this study was drawn were already analyzed for missingness in a previous study that performed hot decking [9]; this imputation technique implies that for each case with missing data, another case similar in characteristics to the case with the missing value is found but has responses for the item in question.

Sample Size Considerations

Sample size plays an important role in providing unbiased parameter estimates and accurate model fit information. Previous research has established guidelines concerning sample sizes needed to accurately estimate item parameters for the unidimensional GRM through simulation studies. For instance, it was reported [51] that a sample size of 375 respondents for a 15-item scale provided adequate discrimination and boundary parameter estimates. Reeve and Fayers [12] reported that GRM could be estimated with 250 respondents. However, around 500 respondents are recommended for accurate parameter estimates [12]. Stemming from a large-scale survey data, our sample widely fulfills this requirement.

In addition, we obtained all analyses and plots using the free R program (R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria) [52]. More specifically, the Mokken package served to test the monotonicity and unidimensionality of the scale. For the detection of local dependence problems, we fit the CCFA model using the Lavaan package. In addition, we were able to estimate the IRT-GRM parameters using the latent trait models (LTM) package, and the Lordif package served to evaluate the DIF effects if any.

Results

Sample Characteristics

The demographic and clinical characteristics of the sample (N=5983) have been described elsewhere [9]. Table 1 presents the item response distribution by region and for the entire sample in this study.

Unidimensionality

In a previous study [12], Velicer's minimum average partial test and parallel analyses [53] supported the 1-factor solution; this solution was also tested by the use of a standard CFA in an asymptotic distribution-free analysis to accommodate nonnormal variables. The 1-factor solution was only supported, however, after allowing for the correlation between 6 pairs of variables, indicating a certain degree of multidimensionality. This study found similar conflicting findings. For instance, the magnitude of Loevinger's coefficients ($H > 0.5$) indicated a strong common dimension, whereas the results of a unidimensional CCFA model showed an inadequate model fit with an RMSEA value of 0.107 and a CFI value of 0.97. In addition, a competing 2-factor CCFA model, which distinguishes between core and peripheral criteria items [20], obtained a more acceptable fit ($\chi^2_{13}=426.0$, $P<.001$;

XSL•FO
RenderX

RMSEA=0.073 and CFI=0.99) but was problematic because of the correlation (>0.9) between the 2 factors being very high and >0.85 cutoff set for the discriminative validity [54]. Therefore, we proceeded with the 1-factor solution, assuming that the effects of multidimensionality were negligible.

Local Independence

We examined local independence using modification indices and residual correlations in the CCFA model. On the one hand, the highest modification index was observed between salience and tolerance. On the other hand, having examined the concept of local independence through the residual correlation matrix, we observed that the residual correlation between salience and tolerance was 0.102, thereby marginally exceeding the cutoff value of 0.10 set by Kline. Another residual, the highest one, which also exceeded the cutoff value of 0.10, was observed between salience and conflict (0.107). These findings suggested that these item pairs (salience and tolerance as well as salience and conflict) might not be totally free of some local dependence bias. In addition, to explore the potential impact of these local dependencies, we examined whether the removal of 1 or 2 of the locally dependent items (eg, salience, tolerance, and conflict) had any noticeable effect on the size of the remaining IRT discriminative parameters in the original unidimensional model [55]. Consequently, the sizes of the remaining discriminative parameters were rather robust to such removals. The largest changes were for salience (−15%, [1.63−1.92]/1.92) and tolerance (−14%, [1.75−2.03]/2.03] with the removal of the tolerance and salience items, respectively. Furthermore, these modest changes supported GAS as being adequately unidimensional to obtain reasonably unbiased parameters when using traditional IRT models, despite some local dependencies.

Monotonicity

We found no violation of monotonicity in this study because rest-score graphs (from the Mokken scale) indicated that the probability of endorsing higher categories increased along the latent trait for all items. All in all, we decided that it is acceptable to use an IRT unidimensional model on GAS.

Item Response Theory Parameter Estimates

Table 2 presents results for item response modeling for GAS as well as the estimates of the parameters in the GRM. Figures 2-4 present the ICC, IIC, and TIF curves. Regarding the ranges proposed by Baker [13], we observed that all items had a very high discriminative power with a range of 1.92-2.93. In increasing order of strength, we found salience followed by tolerance, mood modification, problems, withdrawal, conflict, and relapse. Besides providing a reasonably good differentiation among individuals, large values of the parameter estimates also indicated that all items were highly related to the latent variable, gaming addiction.

In Table 2, it can also be observed that all thresholds were positive, except for those of salience and tolerance, the first threshold of which was negative. Moreover, these 2 items had the largest spread. Hence, their information functions exhibited a broader coverage on the continuum (below and above the mean), whereas the other items were better at discriminating people above the mean. In addition, we observed that all threshold parameters were not tightly clustered together, indicating that the item has adequate response options. Overall, the scale appears to cover a wide range of the item difficulty spectrum from −0.47 (with tolerance) to 3.15 (with withdrawal).

Figure 2 presents ICCs for the 7 items; these curves represent the probability that an individual selects a particular category at a given level of the latent construct. The x-axis represents the latent construct (or gaming addiction in this particular case), in which higher scores are indicative of higher game addiction. In contrast, the y-axis shows the probability of selecting each response option. In addition, each curve corresponds to one of the following 5 possible response alternatives: never, rarely, sometimes, often, and very often. Moving from left to right on the x-axis, the gaming addiction increases. Furthermore, Figure 2 shows that the response options for the respective items are monotonically related to game addiction and that each response option is most likely to be selected at some range of theta.

Consider, for example, ICCs with the largest and the smallest spread, that is, tolerance and relapse, respectively. For tolerance, subjects up to approximately 0.2 SD below the mean were more likely to endorse response category 1 (never); from 0.2 SD below the mean to 0.1 SD above, they were more likely to endorse category 2 (rarely); and from 0.1 to 1.2 SD above the mean, they were more likely to respond to category 3 (sometimes). In addition, from 1.2 to 2.0 SD above the mean, they exhibited the highest likelihood of endorsing category 4 (often). Finally, subjects most likely to choose category 5 (very often) were those with the intensity of gaming disorder symptoms of >2.0.

As GRM is defined in terms of cumulative probabilities, we also performed cumulative comparisons. The difficulties represented a point at which a person with $\theta=b_{ik}$ had a 50% chance of responding in category k or higher [56]. For example, looking at the estimated parameters for tolerance, we observed that a person with $\theta=-0.47$ has a 50% chance of answering 1 versus ≥2 and a person with $\theta=0.24$ has a 50% chance of answering 1 or 2 versus ≥3. Similarly, a person with $\theta=1.18$ has a 50% chance of answering 1, 2, or 3 versus ≥4, and a person with $\theta=2.17$ has a 50% chance of answering 1, 2, 3, or 4 versus 5. We noted that the ratings for tolerance span a broad range of the latent trait and that its discrimination parameter was high.

For relapse, subjects up to 0.8 SD above the mean were more likely to endorse response category 1 (never); subjects from 0.8 to 1.2 SD above the mean were more likely to endorse category 2 (rarely); and subjects from 1.2 to 1.8 SD above the mean were more likely to respond to category 3 (sometimes). From 1.8 to 2.5 SD above the mean, they exhibited the highest likelihood of endorsing category 4 (often); from 2.5 SD above the mean, they were more likely to choose category 5 (very often). With the highest discrimination parameter, we noted that the curves for relapse were more peaked than for tolerance and more concentrated toward the upper end of the trait.

XSL•FO
RenderX

Table 2. Estimates of discrimination and severity parameters for the Game Addiction Scale under the graded-response model with the LTM package.

Item	Discrimination, α_i[a]	Severity				
		β_{i1}	β_{i2}	β_{i3}	β_{i4}	Spread
Salience	1.92	−0.04	0.83	1.58	2.29	2.33
Tolerance	2.03	−0.47	0.24	1.18	2.17	2.64
Mood modification	2.13	0.35	1.06	1.83	2.56	2.21
Relapse	2.93	0.59	1.20	1.90	2.61	2.02
Withdrawal	2.56	0.92	1.64	2.42	3.15	2.23
Conflict	2.83	0.79	1.45	2.23	2.88	2.09
Problems	2.19	0.65	1.38	2.23	2.93	2.28

[a]α_i reflects the ability of item i to discriminate between different levels of game addiction severity (θ).

[b]β_{ik} is the *k* th cutoff point for item *i*. It is interpreted as the standardized level of game addiction severity where subsequent response options become more probable than the previous option.

Figure 2. Item characteristic curves: salience; tolerance; mood modification; relapse; withdrawal; conflict; and problems. GAS: Game Addiction Scale.

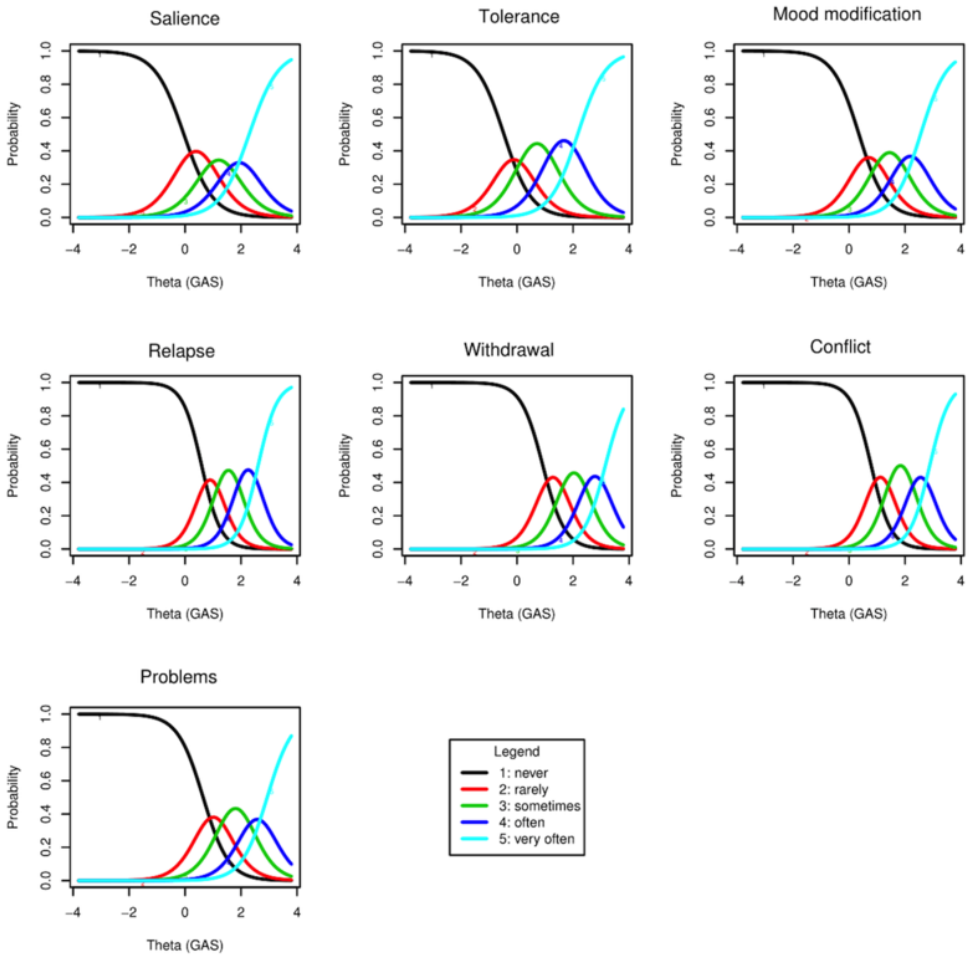


Figure 3 plots IICs of the 7 items. The shape of an IIC was determined both by its discrimination and threshold parameters; however, the steepness of the curves was determined by the magnitude of the discrimination index. Salience, tolerance, mood modification, and problems were less steep than relapse, conflict, and withdrawal, but they covered a wider range of the item severity spectrum. In turn, the latter best discriminated the population for the latent trait at a higher level.

Figure 4 presents TIF, which is the condensed information of each item in Figure 3. Applying the formula [12] reliability=1 − (1⁄information), we observed that the scale reliably assessed a wide range of individuals below and above the average. For instance, information scores of 5-12, which translate to a reliability range of 0.80-0.92, corresponded to participants from 0.3 SD below to 2.5 SD above the mean.

XSL•FO
RenderX

Figure 3. Item information curves. GAS: Game Addiction Scale.

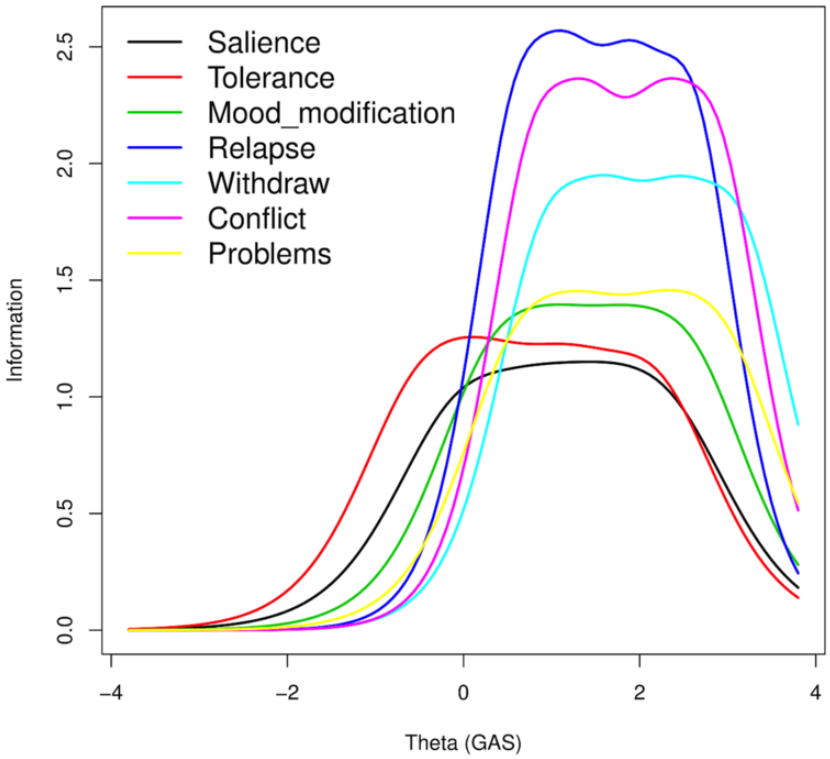
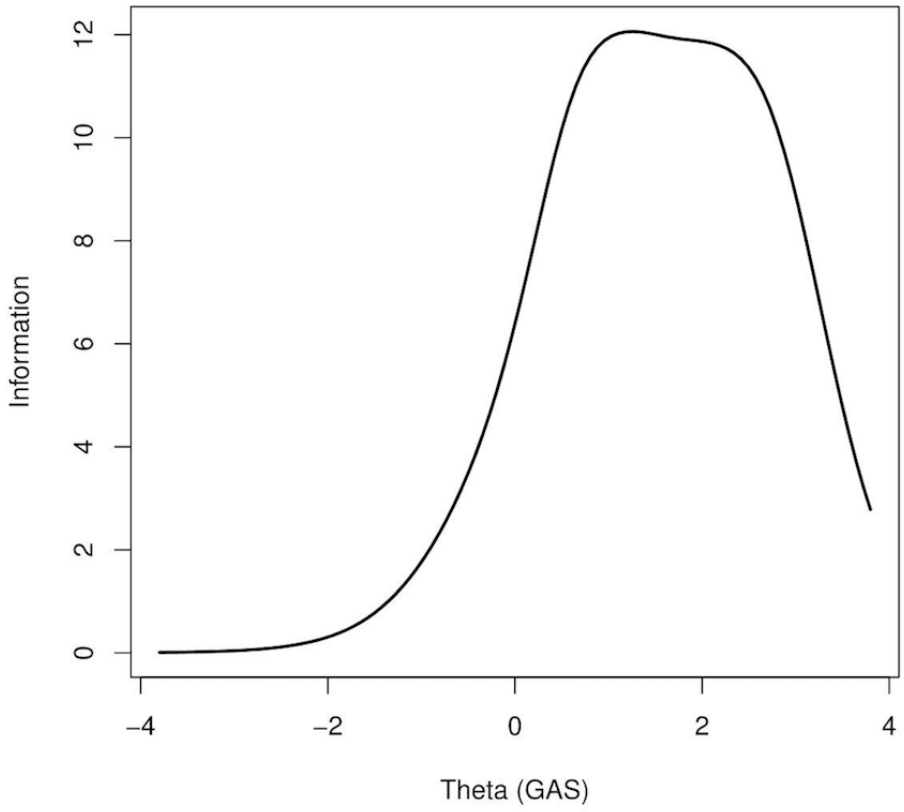
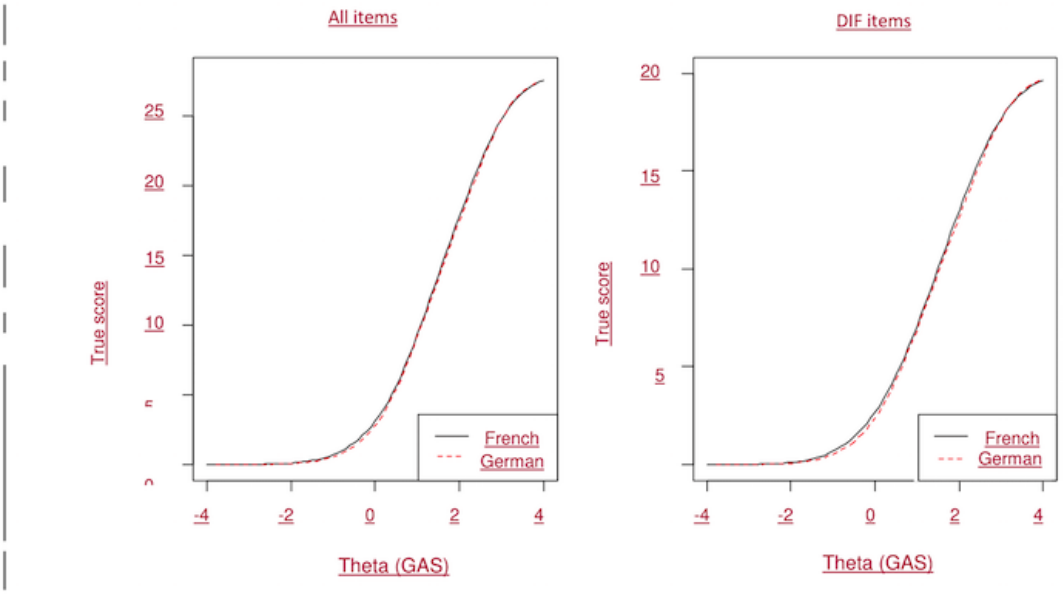


Figure 4. Test (scale) information function. GAS: Game Addiction Scale.



XSL•FO
RenderX

Figure 5. Test characteristic curve: all items (left) and differential item functioning (DIF) items (right). GAS: Game Addiction Scale.



Differential Item Functioning Parameter Estimates

The results obtained with Lordif software showed that 5 of 7 items (ie, salience, tolerance, mood modification, withdrawal, and conflict) were flagged for DIF using the change in the chi-square. Of note, relapse and problems were not flagged as DIF and thus, used as anchors. The nonuniform DIF, that is, a different slope or discrimination parameter between the 2 linguistic regions, was present in salience, withdrawal, and conflict, whereas the uniform DIF, resulting in different severity parameters, was observed in tolerance and mood modification. After the inspection of the pseudo R^2 (not shown here), all could be regarded as negligible when using Jodoin and Gierl's criteria (the largest being 0.0073) [50]. In addition, the test characteristic curves for all items (Figure 5, on left) and the DIF items only (Figure 5, on right) revealed that the impact of the DIF items was trivial at the scale level because the expected total score was virtually identical for the 2 linguistic groups along the whole latent trait continuum.

Discussion

Item Response Theory Modeling

In this study, using IRT modeling, we investigated the psychometric properties of GAS for the amount of information provided by the 7 items and the severity of the latent trait being measured. Although the monotonicity assumption was satisfied, the fit of the unidimensional model was somewhat unsatisfactory owing to the fact that 3 items appeared to be locally dependent. Although these dependencies had some impact on the IRT discriminative parameters, their impact on the performance of new coefficients was not large (maximum 15%). Indeed, referring to Baker's cutoff points, their estimates, except one, remained in the "very high range" category. In addition, when we modeled a 2-factor solution, the high correlation between the 2 dimensions was a matter of concern and ignoring this finding would have undermined their discriminative validity. Satisfied by the strong Loevinger's *H* coefficients, suggesting the occurrence of a strong primary factor, we decided to retain the 1-dimension model and concluded that it is reasonable to analyze GAS with a traditional IRT model. However, it is important to emphasize that the practical impact of ignoring multidimensionality probably depends on the intended use of the scale. Although the local dependencies shown in this study will probably exert a negligible impact on the scaling of individuals, the available research suggests that even minor violations of unidimensionality can exert an important impact on various aspects such as score reliability, differential functioning, and linking [57-59].

All items had high discrimination parameters and as a set, these items differentiated across a reasonable range of the trait. In accordance with Baker's interpretation, their discriminative power was very high [13] with the estimated parameters ranging from 1.92 to 2.93. Overall, the severity parameters (β_1-β_4), which reflect the range of the underlying construct, were between −0.47 and 3.15 for the whole sample, implying that the items show reasonable variability for the endorsement of response categories. Furthermore, no *null categories* existed because all item response categories were chosen by the respondents, null categories being referred to as "never chosen categories." Reportedly, none of the items in response categories seem to be superfluous owing to the fact that their response occupied a distinct portion of the ability continuum [60].

Internet Gaming Disorder Criteria Debate

Theoretical debate is ongoing about the IGD criteria in consideration of their ability to capture the features of addictive internet gaming and their potential tendency to conflate passion (ie, healthy repeated use) and disorder (ie, pathological addictive use). The following 4 criteria, described by some authors of the core addiction criteria [20,30], received more consensus than the other criteria: unsuccessful attempts to reduce or stop [4]; loss of interest in previous hobbies or activities [32,61,62]; continuation despite problems [4,30,62]; and jeopardized or

XSL•FO
RenderX

lost a relationship, job, or educational or career opportunity [62]. When observing such criteria, careful attention must be paid to possible coping motives (ie, related to a depressive disorder) before attributing any such symptoms to addictive behavior [61,63].

The following 5 criteria are more controversial:

1. Preoccupation (being absorbed by gaming and thinking about it): this criterion, thought to be related to cognitive salience, is considered a core criterion by some authors [30,64] but not others [20,61,65]. Preoccupation is commonly reported among high achievers [26,66] and is supposed to be common for gamers because of the social features of the games and flow-related engagement [65,67].
2. Withdrawal: Considered to be a core symptom in some studies [20,30], this criterion has, nonetheless, come under criticism (ie, difficulty distinguishing it from irritability related to the involuntary discontinuation of gaming). The withdrawal symptoms described for IGD were mostly irritability, restlessness, and sadness [68].
3. Tolerance: This criterion refers to the need to increasingly engage in games to feel as though one has played enough. Progression is, however, a part of the game process. This criterion is, therefore, difficult to conceptualize for IGD [69].
4. Escape: Despite the association between game involvement and escape motives [2,70,71], the specificity of this criterion and its link with possible primary disorders (ie, depression) has been discussed [4,26,64,66,72]. In some, but not all [65], IGD-related studies, low diagnostic accuracy was observed for this criterion [61,64].
5. Deceiving others (such as lying to relatives related to the number of games): This criterion is related to "excessive gaming despite problems" and conflicts. Considered as core by some authors [20,30], deceiving others is, however, sensitive to cultural aspects and interactions with relatives and age probably lead to low accuracy of the criterion in some adult studies [66].

Most debates related to the validity of the criteria were theoretically based and insufficiently data-driven [4,73] and thus, more empirical work is warranted [74]. Kiraly et al [62] examined how each IGD criterion performs at different severity levels using an IRT approach and demonstrated that some criteria, such as preoccupation, escape, continue despite problems, and jeopardized or lost a relationship, were endorsed more frequently in less severe IGD stages, whereas other criteria, such as tolerance, unsuccessful attempts to stop or reduce, loss of interest in previous hobbies or activities, and deceiving others, were reported only in more severe cases. However, the study was exposed to self-selected bias because of the Web-based recruitment of a convenience sample [75].

Reappraisal of Internet Gaming Disorder Criteria Using Game Addiction Scale as Proxy

Figure 4 shows that the information provided by GAS is reliable about respondents who are located between 0.3 SD below and 2.5 SD above the mean, suggesting that the scale does a good job of differentiating individuals below and above the average even though it is more precise at a higher level above the mean. Specifically, relapse (unsuccessful attempts to stop or reduce), conflict (deceiving others), withdrawal and problems (loss of interests) were the GAS items with a higher ability to discriminate IGD (endorsed more frequently in more severe IGD stages), whereas the items related to tolerance, salience (preoccupation), and mood modification (escape) were endorsed more widely among participants (included in less severe IGD stages). The results regarding preoccupation and escape were in concordance with those reported in previous studies [62,64], which showed large endorsement of the criteria. As reported in other studies, loss of interests [61,62], unsuccessful attempts to reduce or stop [62], deceiving others [62], and withdrawal [61,62] were more endorsed among participants with more severe IGD.

In contrast to the findings of this study, tolerance was endorsed by more severe cases in other studies [61]. This contradiction could be attributed to the differences in samples or in the wording of the criteria across scales (eg, "Have you ever felt the need to play more often or played for longer periods to feel that you have played enough?" vs "Have you played longer than intended?" in GAS). The wording used in GAS could realistically be interpreted as a form of loss of control or a form of enthusiasm related to the flow [76] induced by the mechanisms of game progression [69]. The wording used for this item in GAS is, perhaps, not entirely successful in capturing the intended meaning of tolerance [32], which might also be part of the reason that we found some local dependence between salience and tolerance in this study.

This study highlights that the IGD condition, as assessed by GAS and the proposed IGD criteria, involves different symptoms, some of which were widely disseminated across the sample and others that were characteristic of disorder severity. However, the GAS items differ from the IGD criteria in several ways. Hence, GAS has to be considered as a proxy measure of the IGD criteria and the findings must be interpreted accordingly.

As found in other studies, the preoccupation [61,62,65] and escape [64,66] criteria exhibited lower discriminatory power than that exhibited by other items. The deceiving other items had good discriminant capacity in this study and others [62], whereas some studies reported low diagnostic accuracy for this criterion [66]. This study was conducted on young adult men, and one may hypothesize that this item is more sensitive to differences in cultural contexts, family contexts, and age groups. In addition, discrepancies between the study results for this criterion could be attributed to differences in item wording across studies ("Have you had arguments with others?" in this study). Furthermore, we cannot exclude that the discriminative ability of the item is inflated in this study because of local dependencies in the model.

Differential Item Functioning

DIF occurs when items have a different relation with the construct in different subgroups; in our case, it is linguistic status. In this study, the discrimination and threshold parameters were very similar between the 2 linguistic groups as the uniform DIF and nonuniform DIF were found to be negligible, as shown

XSL•FO
RenderX

by the weak pseudo R^2. A change in beta showed no significant effect size, except for the withdrawal item, which was just above the 0.01 cutoff; this is the same item that was flagged for measurement invariance in a previous validation of this scale with the AMOS software. However, as can be seen in Figure 5, the curves are superposed. As we expected, the conclusions drawn from standard CFA analyses concerning the measurement invariance between the 2 linguistic regions are unambiguously supported by IRT analyses.

Limitations

This study has several limitations. First, although the sample is representative, it included only young men of about the same age group (almost 99% of them were between 18 and 24 years of age) from Switzerland, thereby limiting the generalizability of the results. Even though the military service is not mandatory for women in Switzerland, it enrolls a marginal number of them each year on a voluntary basis. Because of this marginal number and, more importantly, because no official figures of the female representation were available during the recruitment period, female army recruits were not considered in this study, further limiting the generalizability. However, the sample recruitment allowed us to overcome the self-selection biases reported in other studies [62]. Second, another limitation is related to the use of self-reported questionnaires with possible differences in understanding of questions, desirability bias, and recall bias and the difficulty in assessing the context of a given behavior. Other limitations of the study are directly related to the GAS instrument. In this study, several DSM-5 criteria, such as the loss of opportunities and relationships, were not included in GAS nor were other possibly important criteria for assessing IGD, such as craving or immersion. In addition, the time frame differed (6 months) than that proposed by DSM-5 (12 months). Furthermore, the study did not directly assess the internet-based or game activities used by participants. Thus, for example, we were not able to differentiate one game activity from another or a specific game activity from other types of internet use behavior, although the participants' answers might have related to a specific activity or a combination of activities. However, the advantage of such an approach is that other internet gaming-related activities, which can be time-consuming and performed in excess (eg, game broadcasts), are covered by the items.

Despite the variability across game mechanisms [77], it appears that video games are addictive among some users through refined rewards and processes contributing to the loss of control over game use [78]. In consideration of such similarities between the behavior associated with video games and that associated with other games, numerous studies have assessed games in general without focusing on a specific gaming behavior [62,79]. In addition, previous studies showed the suitability of assessing different internet behaviors (ie, internet gambling and internet gaming) using similar scales [80], whereas other studies concluded the differences between the problematic internet use and online gaming using different assessment tools and finding mostly between-group gender differences [81]. Hence, further studies with IRT analyses are warranted to increase our understanding of the similarities and differences across different types of excessive internet and game use behaviors.

Conclusions

This study partly supports the relevance of using IRT to further establish the psychometric properties of the GAS items. With respect to an overall picture of the symptoms assessed by GAS, relapse, conflict, withdrawal, and problems were endorsed more frequently in more severe IGD stages, whereas the items related to tolerance, preoccupation, and mood modification were endorsed more widely, including among participants in less severe IGD stages. However, these findings must be considered with caution because GAS measures something akin to the IGD criteria but does not measure these criteria *per se*.

Acknowledgments

The authors would like to thank the participants for their contribution to the study. They would like also to thank the Swiss National Science Foundation for the grants allowed for the study (FN 33CSC0-122679 and FN 33CS30-139467).

Conflicts of Interest

None declared.

References

1. Block JJ. Issues for DSM-V: internet addiction. Am J Psychiatry 2008 Mar;165(3):306-307. [doi: 10.1176/appi.ajp.2007.07101556] [Medline: 18316427]
2. Kuss DJ, Louws J, Wiers RW. Online gaming addiction? Motives predict addictive play behavior in massively multiplayer online role-playing games. Cyberpsychol Behav Soc Netw 2012 Sep;15(9):480-485. [doi: 10.1089/cyber.2012.0034] [Medline: 22974351]
3. Király O, Griffiths MD, King DL, Lee H, Lee S, Bányai F, et al. Policy responses to problematic video game use: A systematic review of current measures and future possibilities. J Behav Addict 2017 Sep 01:1-15. [doi: 10.1556/2006.6.2017.050] [Medline: 28859487]
4. Kardefelt-Winther D, Heeren A, Schimmenti A, van RA, Maurage P, Carras M, et al. How can we conceptualize behavioural addiction without pathologizing common behaviours? Addiction 2017 Oct;112(10):1709-1715. [doi: 10.1111/add.13763] [Medline: 28198052]

XSL•FO
RenderX

5. Aarseth E, Bean AM, Boonen H, Colder CM, Coulson M, Das D, et al. Scholars' open debate paper on the World Health Organization ICD-11 Gaming Disorder proposal. J Behav Addict 2017 Sep 01;6(3):267-270 [FREE Full text] [doi: 10.1556/2006.5.2016.088] [Medline: 28033714]
6. Billieux J, King DL, Higuchi S, Achab S, Bowden-Jones H, Hao W, et al. Functional impairment matters in the screening and diagnosis of gaming disorder. J Behav Addict 2017 Sep 01;6(3):285-289 [FREE Full text] [doi: 10.1556/2006.6.2017.036] [Medline: 28816514]
7. Billieux J, Schimmenti A, Khazaal Y, Maurage P, Heeren A. Are we overpathologizing everyday life? A tenable blueprint for behavioral addiction research. J Behav Addict 2015 Sep;4(3):119-123 [FREE Full text] [doi: 10.1556/2006.4.2015.009] [Medline: 26014667]
8. Lemmens JS, Valkenburg PM, Peter J. Development and Validation of a Game Addiction Scale for Adolescents. Media Psychology 2009 Mar 05;12(1):77-95. [doi: 10.1080/15213260802669458]
9. Khazaal Y, Chatton A, Rothen S, Achab S, Thorens G, Zullino D, et al. Psychometric properties of the 7-item game addiction scale among french and German speaking adults. BMC Psychiatry 2016 May 10;16:132 [FREE Full text] [doi: 10.1186/s12888-016-0836-3] [Medline: 27160387]
10. Raju NS, Laffitte LJ, Byrne BM. Measurement equivalence: a comparison of methods based on confirmatory factor analysis and item response theory. J Appl Psychol 2002 Jun;87(3):517-529. [Medline: 12090609]
11. Reise SP, Widaman KF, Pugh RH. Confirmatory factor analysis and item response theory: two approaches for exploring measurement invariance. Psychol Bull 1993 Nov;114(3):552-566. [Medline: 8272470]
12. Reeve BR, Fayers P. Applying item response theory modelling for evaluating questionnaire itemscale properties. In: Fayers P, Hays H, editors. Assessing Quality of Life in Clinical Trials: Methods and Practice. 2nd ed. New York: Oxford University Press; 2005:55-73.
13. Baker B. The Basics of Item Response Theory. 2001. The Basics Of Item Response Theory URL: http://www.edres.org/irt/baker/final.pdf [accessed 2018-06-26] [WebCite Cache ID 70SkGjBPh]
14. Samejima F. Estimation of latent ability using a response pattern of graded scores. Psychometrika 1969;34(1):1-97. [doi: 10.1007/BF03372160]
15. Rodriguez A, Reise SP, Haviland MG. Evaluating bifactor models: Calculating and interpreting statistical indices. Psychol Methods 2016 Jun;21(2):137-150. [doi: 10.1037/met0000045] [Medline: 26523435]
16. DeMars C. Item Response Theory. Oxford: Oxford University Press; 2010.
17. Gaetan S, Bonnet A, Brejard V, Cury F. French validation of the 7-item Game Addiction Scale for adolescents. Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology 2014 Jul;64(4):161-168 [FREE Full text] [doi: 10.1016/j.erap.2014.04.004]
18. Akin A, Ozcelik B, Usta F, Akin U, Secgin A. The validity and reliability of the Turkish version of the Game Addiction Scale. J Behav Addict (S1) 2015;4:61-62. [doi: 10.1556/JBA.4.2015.Suppl.1]
19. Festl R, Scharkow M, Quandt T. Problematic computer game use among adolescents, younger and older adults. Addiction 2013 Mar;108(3):592-599. [doi: 10.1111/add.12016] [Medline: 23078146]
20. Brunborg GS, Hanss D, Mentzoni RA, Pallesen S. Core and peripheral criteria of video game addiction in the game addiction scale for adolescents. Cyberpsychol Behav Soc Netw 2015 May;18(5):280-285. [doi: 10.1089/cyber.2014.0509] [Medline: 25826043]
21. Reise S, Cook K, Moore T. Evaluating the impact of multidimensionality on unidimensional item response theory model parameters. In: Reise SP, Revicki DA, editors. Handbook of Item Response Theory Modeling: Application to Typical Performance Assessment. New York: Routledge; 2015:13-40.
22. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. Med Care 2000 Sep;38(9 Suppl):II28-II42 [FREE Full text] [Medline: 10982088]
23. Starcevic V. Tolerance and withdrawal symptoms may not be helpful to enhance understanding of behavioural addictions. Addiction 2016 Dec;111(7):1307-1308. [doi: 10.1111/add.13381] [Medline: 27095413]
24. Petry NM, Rehbein F, Ko C, O'Brien CP. Internet Gaming Disorder in the DSM-5. Curr Psychiatry Rep 2015 Sep;17(9):72. [doi: 10.1007/s11920-015-0610-0] [Medline: 26216590]
25. American PA. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. Washington, DC: American Psychiatric Association; 2013:795-798.
26. Kuss DJ, Griffiths MD, Pontes HM. Chaos and confusion in DSM-5 diagnosis of Internet Gaming Disorder: Issues, concerns, and recommendations for clarity in the field. J Behav Addict 2017 Jun 01;6(2):103-109 [FREE Full text] [doi: 10.1556/2006.5.2016.062] [Medline: 27599673]
27. Kardefelt-Winther D. A critical account of DSM-5 criteria for internet gaming disorder. Addiction Research & Theory 2014 Jul 03;23(2):93-98. [doi: 10.3109/16066359.2014.935350]
28. Starcevic V, Aboujaoude E. Internet addiction: reappraisal of an increasingly inadequate concept. CNS Spectr 2017 Feb;22(1):7-13. [doi: 10.1017/S1092852915000863] [Medline: 26831456]
29. Billieux J, van RAJ, Heeren A, Schimmenti A, Maurage P, Edman J, et al. Behavioural Addiction Open Definition 2.0-using the Open Science Framework for collaborative and transparent theoretical development. Addiction 2017 Oct;112(10):1723-1724. [doi: 10.1111/add.13938] [Medline: 28891143]

XSL•FO
RenderX

30. Charlton JP, Danforth ID. Distinguishing addiction and high engagement in the context of online game playing. Computers in Human Behavior 2007 May;23(3):1531-1548. [doi: 10.1016/j.chb.2005.07.002]
31. Griffiths M. A 'components' model of addiction within a biopsychosocial framework. Journal of Substance Use 2009 Jul 12;10(4):191-197. [doi: 10.1080/14659890500114359]
32. Petry NM, Rehbein F, Gentile DA, Lemmens JS, Rumpf H, Mößle T, et al. An international consensus for assessing internet gaming disorder using the new DSM-5 approach. Addiction 2014 Sep;109(9):1399-1406. [doi: 10.1111/add.12457] [Medline: 24456155]
33. An X, Yung Y. SAS Institute. 2014. What It Is and How You Can Use the IRT Procedure to Apply It URL: https://support.sas.com/resources/papers/proceedings14/SAS364-2014.pdf [accessed 2018-07-19] [WebCite Cache ID 711z1vYVm]
34. Lipscomb J, Gotay C, Snyder C. Outcomes Assessment in Cancer: Measures, Methods and Applications. Cambridge: Cambridge University Press; 2005.
35. Mielenz TJ, Callahan LF, Edwards MC. Item response theory analysis of Centers for Disease Control and Prevention Health-Related Quality of Life (CDC HRQOL) items in adults with arthritis. Health Qual Life Outcomes 2016 Mar 12;14:43 [FREE Full text] [doi: 10.1186/s12955-016-0444-4] [Medline: 26969513]
36. Prisciandaro JJ, Tolliver BK. An item response theory evaluation of the young mania rating scale and the montgomery-asberg depression rating scale in the systematic treatment enhancement program for bipolar disorder (STEP-BD). J Affect Disord 2016 Nov 15;205:73-80 [FREE Full text] [doi: 10.1016/j.jad.2016.06.062] [Medline: 27420134]
37. Loevinger L. A systematic approach to the construction and evaluation of tests of ability. Psychological Monographs 1947;61(4):i-49. [doi: 10.1037/h0093565]
38. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
39. Hooper D, Coughlan J, Mullen M. Structural equation modeling: guidelines for determining model fit. Electronic Journal of Business Research Methods 2008;6:53-60.
40. Stochl J, Jones PB, Croudace TJ. Mokken scale analysis of mental health and well-being questionnaire item responses: a non-parametric IRT method in empirical research for applied health researchers. BMC Med Res Methodol 2012;12:74 [FREE Full text] [doi: 10.1186/1471-2288-12-74] [Medline: 22686586]
41. Sijtsma K, Molenaar I. Introduction to Nonparametric Item Response Theory. London: Sage Publications; 2002.
42. McDonald R. The dimensionality of tests and items. Br J Math Stat Psychol 1981;34(1):100-117. [doi: 10.1111/j.2044-8317.1981.tb00621.x]
43. Yang FM, Kao ST. Item response theory for measurement validity. Shanghai Arch Psychiatry 2014 Jun;26(3):171-177 [FREE Full text] [doi: 10.3969/j.issn.1002-0829.2014.03.010] [Medline: 25114494]
44. Kline R. Principles and Practice of Structural Equation Modeling. 3rd ed. New York: The Guilford Press; 2011.
45. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Med Care 2007 May;45(5 Suppl 1):S22-S31. [doi: 10.1097/01.mlr.0000250483.85507.04] [Medline: 17443115]
46. Van DLW, Hambleton R. Handbook of Modern Item Response Theory. New York: Spring-Verlag; 1997.
47. Hall TW, Reise SP, Haviland MG. An Item Response Theory Analysis of the Spiritual Assessment Inventory. International Journal for the Psychology of Religion 2007 Apr 20;17(2):157-178. [doi: 10.1080/10508610701244197]
48. Choi SW, Gibbons LE, Crane PK. lordif: An R Package for Detecting Differential Item Functioning Using Iterative Hybrid Ordinal Logistic Regression/Item Response Theory and Monte Carlo Simulations. J Stat Softw 2011 Mar 1;39(8):1-30 [FREE Full text] [Medline: 21572908]
49. Borsboom D. The attack of the psychometricians. Psychometrika 2006 Sep;71(3):425-440 [FREE Full text] [doi: 10.1007/s11336-006-1447-6] [Medline: 19946599]
50. Jodoin M, Gierl M. Evaluating Type I Error and Power Rates Using an Effect Size Measure With the Logistic Regression Procedure for DIF Detection. Applied Measurement in Education 2001 Oct 27;14(4):329-349 [FREE Full text] [doi: 10.1207/S15324818AME1404_2] [Medline: 25830810]
51. Jiang S, Wang C, Weiss DJ. Sample Size Requirements for Estimation of Item Parameters in the Multidimensional Graded Response Model. Front Psychol 2016;7:109 [FREE Full text] [doi: 10.3389/fpsyg.2016.00109] [Medline: 26903916]
52. R Core Team. R Foundation for Statistical Computing, Vienna, Austria. 2018. URL: http://www.r-project.org/ [accessed 2018-07-20] [WebCite Cache ID 7138ug5uN]
53. O'Connor BP. SPSS and SAS programs for determining the number of components using parallel analysis and velicer's MAP test. Behav Res Methods Instrum Comput 2000 Aug;32(3):396-402. [Medline: 11029811]
54. Brown T. Confirmatory Factor Analysis for Applied Research. 2nd ed. New York: The Guildford Press; 2015.
55. Revicki D, Chen W, Tucker C. Developing item banks for patient-reported health outcomes. In: Reise SP, Revicki DA. editors. Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment. New York: Routledge; 2015:334-363.
56. StataCorp. Release 141. College Station, TX. 2016. Stata Statistical Software URL: https://www.stata.com/ [accessed 2018-07-20] [WebCite Cache ID 7139KXL9r]

XSL•FO
RenderX

57. DeMars C. Application of the bifactor multidimensional item response theory model to testlet?based tests. J Educ Meas 2006;43(2):145-168.
58. Reise S, Cook K, Moore T. Evaluating the impact of multidimensionality on unidimensional item response theory model parameters. In: Reise SP, Revicki DA. editors. Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment. New York: Routledge; 2015:13-40.
59. Stucky B, Edelen M. Using hierarchical IRT models to create unidimensional measures from multidimensional data. In: Reise SP, Revicki DA. editors. Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment. New York: Routledge; 2015:183-206.
60. DeVellis R. In: Bickmann L, Rog DJ, editors. Scale Development: Theory and Applications. London, UK: Sage Publications; 2016:1-262.
61. Rehbein F, Kliem S, Baier D, Mößle T, Petry NM. Prevalence of Internet gaming disorder in German adolescents: diagnostic contribution of the nine DSM-5 criteria in a state-wide representative sample. Addiction 2015 May;110(5):842-851. [doi: 10.1111/add.12849] [Medline: 25598040]
62. Király O, Sleczka P, Pontes HM, Urbán R, Griffiths MD, Demetrovics Z. Validation of the Ten-Item Internet Gaming Disorder Test (IGDT-10) and evaluation of the nine DSM-5 Internet Gaming Disorder criteria. Addict Behav 2017 Jan;64:253-260. [doi: 10.1016/j.addbeh.2015.11.005] [Medline: 26632194]
63. Starcevic V, Khazaal Y. Relationships between Behavioural Addictions and Psychiatric Disorders: What Is Known and What Is Yet to Be Learned? Front Psychiatry 2017;8:53 [FREE Full text] [doi: 10.3389/fpsyt.2017.00053] [Medline: 28439243]
64. Lemmens JS, Valkenburg PM, Gentile DA. The Internet Gaming Disorder Scale. Psychol Assess 2015 Jun;27(2):567-582. [doi: 10.1037/pas0000062] [Medline: 25558970]
65. Deleuze J, Nuyens F, Rochat L, Rothen S, Maurage P, Billieux J. Established risk factors for addiction fail to discriminate between healthy gamers and gamers endorsing DSM-5 Internet gaming disorder. J Behav Addict 2017 Dec 01;6(4):516-524. [doi: 10.1556/2006.6.2017.074] [Medline: 29130328]
66. Ko C, Yen J, Chen S, Wang P, Chen C, Yen C. Evaluation of the diagnostic criteria of Internet gaming disorder in the DSM-5 among young adults in Taiwan. J Psychiatr Res 2014 Jun;53:103-110. [doi: 10.1016/j.jpsychires.2014.02.008] [Medline: 24581573]
67. Csikszentmihalyi M. TED. 2004. Flow, the Secret to Happiness: Ted URL: https://www.ted.com/talks/mihaly_csikszentmihalyi_on_flow?language=en [WebCite Cache ID 713C5Ho9v]
68. Kaptsis D, King DL, Delfabbro PH, Gradisar M. Withdrawal symptoms in internet gaming disorder: A systematic review. Clin Psychol Rev 2016 Feb;43:58-66. [doi: 10.1016/j.cpr.2015.11.006] [Medline: 26704173]
69. King DL, Delfabbro PH. Defining tolerance in Internet Gaming disorder: Isn't it time? Addiction 2016 Nov;111(11):2064-2065. [doi: 10.1111/add.13448] [Medline: 27297731]
70. Billieux J, Chanal J, Khazaal Y, Rochat L, Gay P, Zullino D, et al. Psychological predictors of problematic involvement in massively multiplayer online role-playing games: illustration in a sample of male cybercafé players. Psychopathology 2011;44(3):165-171. [doi: 10.1159/000322525] [Medline: 21372629]
71. Zanetta DF, Zermatten A, Billieux J, Thorens G, Bondolfi G, Zullino D, et al. Motivations to play specifically predict excessive involvement in massively multiplayer online role-playing games: evidence from an online survey. Eur Addict Res 2011;17(4):185-189. [doi: 10.1159/000326070] [Medline: 21494046]
72. Pontes HM, Király O, Demetrovics Z, Griffiths MD. The conceptualisation and measurement of DSM-5 Internet Gaming Disorder: the development of the IGD-20 Test. PLoS One 2014;9(10):e110137 [FREE Full text] [doi: 10.1371/journal.pone.0110137] [Medline: 25313515]
73. Saunders JB. Substance use and addictive disorders in DSM-5 and ICD 10 and the draft ICD 11. Curr Opin Psychiatry 2017 Jul;30(4):227-237. [doi: 10.1097/YCO.0000000000000332] [Medline: 28459730]
74. van RAJ, Schoenmakers TM, van DMD. Clinical validation of the C-VAT 2.0 assessment tool for gaming disorder: A sensitivity analysis of the proposed DSM-5 criteria and the clinical characteristics of young patients with 'video game addiction'. Addict Behav 2017 Jan;64:269-274. [doi: 10.1016/j.addbeh.2015.10.018] [Medline: 26526624]
75. Khazaal Y, van SM, Chatton A, Achab S, Zullino D, Rothen S, et al. Does self-selection affect samples' representativeness in online surveys? An investigation in online video game research. J Med Internet Res 2014;16(7):e164 [FREE Full text] [doi: 10.2196/jmir.2759] [Medline: 25001007]
76. Hamari J, Shernoff DJ, Rowe E, Coller B, Asbell-Clarke J, Edwards T. Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. Computers in Human Behavior 2016 Jan;54:170-179. [doi: 10.1016/j.chb.2015.07.045]
77. Thorens G, Billieux J, Megevand P, Zullino D, Rothen S, Achab S, et al. Capitalizing upon the Attractive and Addictive Properties of Massively Multiplayer Online Role-Playing Games to Promote Wellbeing. Front Psychiatry 2016;7:167 [FREE Full text] [doi: 10.3389/fpsyt.2016.00167] [Medline: 27799913]
78. Wei L, Zhang S, Turel O, Bechara A, He Q. A Tripartite Neurocognitive Model of Internet Gaming Disorder. Front Psychiatry 2017;8:285 [FREE Full text] [doi: 10.3389/fpsyt.2017.00285] [Medline: 29312016]

XSL•FO
RenderX

79. Przybylski AK, Weinstein N, Murayama K. Internet Gaming Disorder: Investigating the Clinical Relevance of a New Phenomenon. Am J Psychiatry 2017 Mar 01;174(3):230-236. [doi: 10.1176/appi.ajp.2016.16020224] [Medline: 27809571]
80. Khazaal Y, Achab S, Billieux J, Thorens G, Zullino D, Dufour M, et al. Factor Structure of the Internet Addiction Test in Online Gamers and Poker Players. JMIR Ment Health 2015;2(2):e12 [FREE Full text] [doi: 10.2196/mental.3805] [Medline: 26543917]
81. Király O, Griffiths MD, Urbán R, Farkas J, Kökönyei G, Elekes Z, et al. Problematic internet use and problematic online gaming are not the same: findings from a large nationally representative adolescent sample. Cyberpsychol Behav Soc Netw 2014 Dec;17(12):749-754 [FREE Full text] [doi: 10.1089/cyber.2014.0475] [Medline: 25415659]

Abbreviations

CCFA: categorical confirmatory factor analysis
CFA: confirmatory factor analysis
CFI: comparative fit index
DIF: differential item functioning
DSM-5: Diagnostic and Statistical Manual for Mental Disorders, 5th Edition
GAS: Game Addiction Scale
GRM: graded-response model
ICC: item characteristic curve
IIC: item information curve
IGD: internet gaming disorder
IRT: item response theory
RMSEA: root mean square error of approximation
TIF: test information function

Edited by G Eysenbach; submitted 06.02.18; peer-reviewed by J Grohol PsyD, N Khalili-Mahani; comments to author 08.03.18; revised version received 24.04.18; accepted 12.05.18; published 27.08.18.

Please cite as:
Khazaal Y, Breivik K, Billieux J, Zullino D, Thorens G, Achab S, Gmel G, Chatton A
Game Addiction Scale Assessment Through a Nationally Representative Sample of Young Adult Men: Item Response Theory Graded–Response Modeling
J Med Internet Res 2018;20(8):e10058
URL: http://www.jmir.org/2018/8/e10058/
doi:10.2196/10058
PMID:30150204

©Yasser Khazaal, Kyrre Breivik, Joel Billieux, Daniele Zullino, Gabriel Thorens, Sophia Achab, Gerhard Gmel, Anne Chatton. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 27.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Understanding Users' Vaping Experiences from Social Media: Initial Study Using Sentiment Opinion Summarization Techniques

Qiudan Li[1], PhD; Can Wang[1,2*], BE; Ruoran Liu[1,2*], BE; Lei Wang[1], PhD; Daniel Dajun Zeng[1,2,3], PhD; Scott James Leischow[4], PhD

[1]The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China
[2]University of Chinese Academy of Sciences, Beijing, China
[3]Department of Management Information Systems, Eller College of Management, The University of Arizona, Tucson, AZ, United States
[4]College of Health Solutions, Arizona State University, Phoenix, AZ, United States
[*]these authors contributed equally

Corresponding Author:
Daniel Dajun Zeng, PhD
Department of Management Information Systems
Eller College of Management
The University of Arizona
McClelland Hall 430
1130 E Helen Street
Tucson, AZ, 85721
United States
Phone: 1 520 621 4614
Fax: 1 520 621 2433
Email: zeng@eller.arizona.edu

Abstract

Background: E-liquid is one of the main components in electronic nicotine delivery systems (ENDS). ENDS review comments could serve as an early warning on use patterns and even function to serve as an indicator of problems or adverse events pertaining to the use of specific e-liquids—much like types of responses tracked by the Food and Drug Administration (FDA) regarding medications.

Objective: This study aimed to understand users' "vaping" experience using sentiment opinion summarization techniques, which can help characterize how consumers think about specific e-liquids and their characteristics (eg, flavor, throat hit, and vapor production).

Methods: We collected e-liquid reviews on JuiceDB from June 27, 2013 to December 31, 2017 using its public application programming interface. The dataset contains 27,070 reviews for 8058 e-liquid products. Each review is accompanied by an overall rating and a set of 4 aspect ratings of an e-liquid, each on a scale of 1-5: flavor accuracy, throat hit, value, and cloud production. An iterative dichotomiser 3 (ID3)-based influential aspect analysis model was adopted to learn the key elements that impact e-liquid use. Then, fine-grained sentiment analysis was employed to mine opinions on various aspects of vaping experience related to e-liquids.

Results: We found that flavor accuracy and value were the two most important aspects that affected users' sentiments toward e-liquids. Of reviews in JuiceDB, 67.83% (18,362/27,070) were positive, while 12.67% (3430/27,070) were negative. This indicates that users generally hold positive attitudes toward e-liquids. Among the 9 flavors, fruity and sweet were the two most popular. Great and sweet tastes, reasonable value, and strong throat hit made users satisfied with fruity and sweet flavors, whereas "strange" tastes made users dislike those flavors. Meanwhile, users complained about some e-liquids' steep or expensive prices, bad quality, and harsh throat hit. There were 2342 fruity e-liquids and 2049 sweet e-liquids. There were 55.81% (1307/2342) and 59.83% (1226/2049) positive sentiments and 13.62% (319/2342) and 12.88% (264/2049) negative sentiments toward fruity e-liquids and sweet e-liquids, respectively. Great flavors and good vapors contributed to positive reviews of fruity and sweet products. However, bad tastes such as "sour" or "bitter" resulted in negative reviews. These findings can help businesses and policy makers to further improve product quality and formulate effective policy.

XSL•FO
RenderX

Conclusions: This study provides an effective mechanism for analyzing users' ENDS vaping experience based on sentiment opinion summarization techniques. Sentiment opinions on aspect and products can be found using our method, which is of great importance to monitor e-liquid products and improve work efficiency.

(J Med Internet Res 2018;20(8):e252) doi:10.2196/jmir.9373

KEYWORDS

electronic nicotine delivery systems; e-cigarette; e-liquid; JuiceDB; sentiment opinion summarization; social media; vaping; infodemiology

Introduction

The market for electronic nicotine delivery systems (ENDS) or electronic cigarettes (e-cigarettes) is growing rapidly. According to data from Research and Markets, the global electronic cigarette market was expected to grow at a compound annual rate of 16.6% during 2017-2022 and hit US $27.7 billion by 2022 [1]. E-cigarettes are now the most commonly used tobacco product among youth [2]. More than 2 million middle and high school students used e-cigarettes in 2016 [3]. Among middle school students, 31% use e-cigarettes because they contain multiple flavors [4]. To protect Americans from dangers of tobacco and nicotine, the US Food and Drug Administration (FDA) extended its authority to e-cigarettes in 2016 [5]. ENDS or e-cigarette products heat a liquid (e-liquid) that may contain nicotine, as well as varying compositions of flavorings, propylene glycol (PG), vegetable glycerin (VG), and other ingredients into an aerosol that the user inhales [3]. All elements in the e-liquid form the unique "vaping" experience. For example, VG produces more vapor than PG and offers a slight sweetness; PG provides more "throat hit" and usually carries a stronger flavor [6-9]. Mining vaping experience with e-liquid products can help FDA policy makers understand use patterns and reasons users like or dislike products and thus make better decisions.

Social media such as Facebook, Twitter, and YouTube have recently become significant platforms for health surveillance and social intelligence [10,11], also providing new insights on e-cigarettes to help inform future research, regulations, surveillance, and enforcement efforts [12]. For example, Liang et al studied the prevalence and promotional strategies of protobacco content in Facebook, Wikipedia, and YouTube [13]. Chu et al examined marketing strategies of leading e-cigarette brands on multiple social networking sites including Facebook, Twitter, Google+, and Instagram [14]. Hua et al showed that e-cigarette use can have wide-ranging positive and negative effects by analyzing health effects in Web-based forums [15]. Kim et al used Twitter data to gain insights into e-cigarette marketing and locations of use [12]. Cole-Lewis et al conducted content analysis to identify key conversation trends and patterns over time using historical Twitter data [16]. Cole-Lewis et al adopted supervised machine learning-based predictive classification models to assess Twitter data for a range of e-cigarette-related factors, thus helping the development of public health communication, policies, and interventions regarding e-cigarettes [17]. Lazard et al uncovered key patterns and important e-cigarette topics on Twitter [18]. Harris et al analyzed messages and tweet patterns to mine the response to the Chicago Department of Public Health's e-cigarette campaign [19]. Huang et al quantified e-cigarette-related videos on YouTube, assessed their content, and characterized levels of engagement with the videos [20].

As two new social media platforms, Reddit and JuiceDB have been studied to analyze broadly discussed vaping methods and features including flavor, throat hit, and vapor production. In previous research, Wang et al have identified 8 categories of flavors on Reddit: fruits, cream, tobacco, menthol, beverages, sweet, seasonings, and nuts [21]. In JuiceDB, there were 9 flavor categories: sweet, fruity, rich, creamy, spiced, tobacco, cool, nutty, and coffee. The two category systems were fairly consistent, providing a good schema for future research. Li et al mined potential relationships between symptoms and e-liquid components, such as PG, VG, flavor extracts, and nicotine, using user-generated data collected from Reddit [22]. Jin et al performed e-liquid review rating prediction by jointly modeling review content and aspect ratings [23]. Zhan et al examined Reddit, JuiceDB, and Twitter as social media data sources for e-cigarette research and adopted latent Dirichlet allocation topic modeling techniques to identify 4 topics across platforms: promotions, flavor discussions, experience sharing, and regulation debates [24]. Chen et al analyzed polarities of e-liquid features by mining Web-based reviews [25]. These studies showed the importance of flavor in ENDS or e-cigarette products.

Despite the growing amount of literature regarding ENDS or e-cigarettes on social media, to date, no published studies have systematically mined users' e-liquid usage patterns from review data based on opinion summarization techniques. JuiceDB [26], one of the world's largest independent e-liquid and vape juice resources, has great influence in promoting e-liquid product use through user-generated content, that is, it allows users to share their vaping experience with different e-liquid products by leaving detailed comments, aspect ratings, and overall product ratings. This study aims to answer the following questions. Which factors have the most influence on users' sentiments toward e-liquids? What are the most popular flavors? Why do users like flavors and products? Data-driven findings could serve as an early warning on use patterns and even function to indicate problems or adverse events pertaining to use of specific e-liquids.

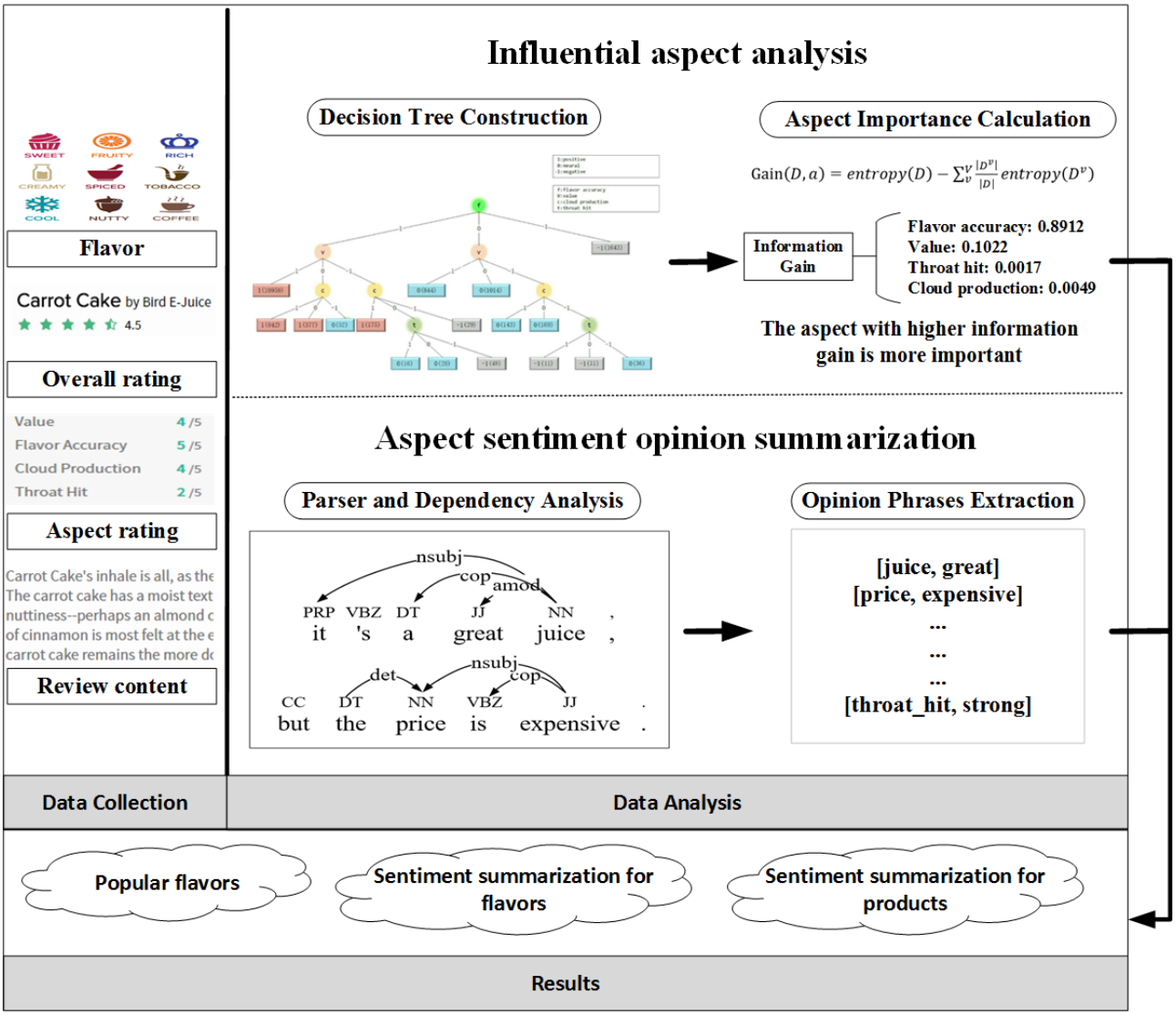
XSL•FO
RenderX

Methods

Framework

Figure 1 shows the framework for analyzing users' ENDS vaping experience. It consists of three components: data collection and preprocessing, data analysis, and results.

Figure 1. The framework to analyze users' electronic nicotine delivery systems vaping experience. Amod: adjectival modifier dependency relationship; CC: coordinating conjunction; cop: copula dependency relationship; det: determiner dependency relationship; DT: determiner; JJ: adjective; NN: noun, singular or mass; nsubj: nominal subject dependency relationship; PRP: personal pronoun; VBZ: verb, third person singular present.



Data Collection and Preprocessing

Since the first review by JuiceDB's API was published on June 27, 2013, we used API to collect e-liquid reviews on JuiceDB, one of the world's largest independent e-liquid and vape juice resources, from June 27, 2013 to December 31, 2017. The JuiceDB website provides flavor category information for each product. Registered users can provide reviews for e-liquids consisting of an overall rating and aspect ratings that respectively reflect their sentiments toward the product and its attributes. Each review is accompanied by an overall rating and a set of 4 e-liquid aspect ratings on a scale from 1 to 5: flavor accuracy, throat hit, value, and cloud production. The dataset contains 27,070 reviews for 8058 e-liquid products.

To better understand the sentiment toward products and aspects, discretization processing is necessary. A positive label is given to a product or aspect if the overall rating or aspect rating is ≥4; a neutral label is generated if the rating score is ≥3 and <4; and a negative label is assigned if the score is <3.

Data Analysis

This study aimed to understand users' e-liquid usage patterns by mining summarization, which helps explain reasons users like or dislike a product. The following processes were performed.

Influential Aspect Analysis

To evaluate the importance of aspects that influence a user's sentiment toward an e-liquid product, the iterative dichotomiser 3 (ID3) algorithm was adopted to construct a decision tree,

XSL•FO
RenderX

which has turned out to be an efficient method of identifying important features [27]. The key idea of the method was to compute feature importance based on information gain. An aspect with higher information gain has greater influence on users' sentiments toward an e-liquid product. First, both aspect ratings and overall ratings were discretized. Second, the ID3 algorithm computed the information gain of each aspect and split the dataset into subsets according to the value of the aspect with the largest information gain. This process was iterated on each subset until there was no available aspect. Finally, the importance of an aspect was computed as the normalized total information gain brought by the corresponding aspect. The aspect with a higher value was considered more important.

Aspect Sentiment Opinion Summarization

Opinions are aspect-sentiment pairs that summarize a user's sentiment toward a product at a fine granularity. Opinion summarization modeling aims to automatically mine aspect words and their corresponding sentiment words [28]. The model consists of the following two steps.

Step 1: Parser and Dependency Analysis

To identify words' part-of-speech tag and dependency in review sentences, Stanford Parser 3.4 [29], one module in the Stanford natural language processing toolbox, was adopted. For example, in "Flavor is great, definitely an adv," the adjective "great" modifies the noun "flavor."

Step 2: Opinion Phrases Extraction

Based on the above results, aspect-sentiment pairs were extracted. An aspect word is usually a noun. Term frequency was adopted to measure the importance of nouns, and we selected nouns whose term frequency was >20 as candidate words. Then, meaningful aspect words were manually selected. Sentiments are adjective words that modify the aspect words. The sentiment polarity of aspect-sentiment pairs was identified by the popular emotional word dictionary [30]. For example, an opinion phrase "great flavor" can be extracted from "A great flavor. Tastes like tobacco with waffles and maple syrup," and the corresponding sentiment polarity is positive.

Results

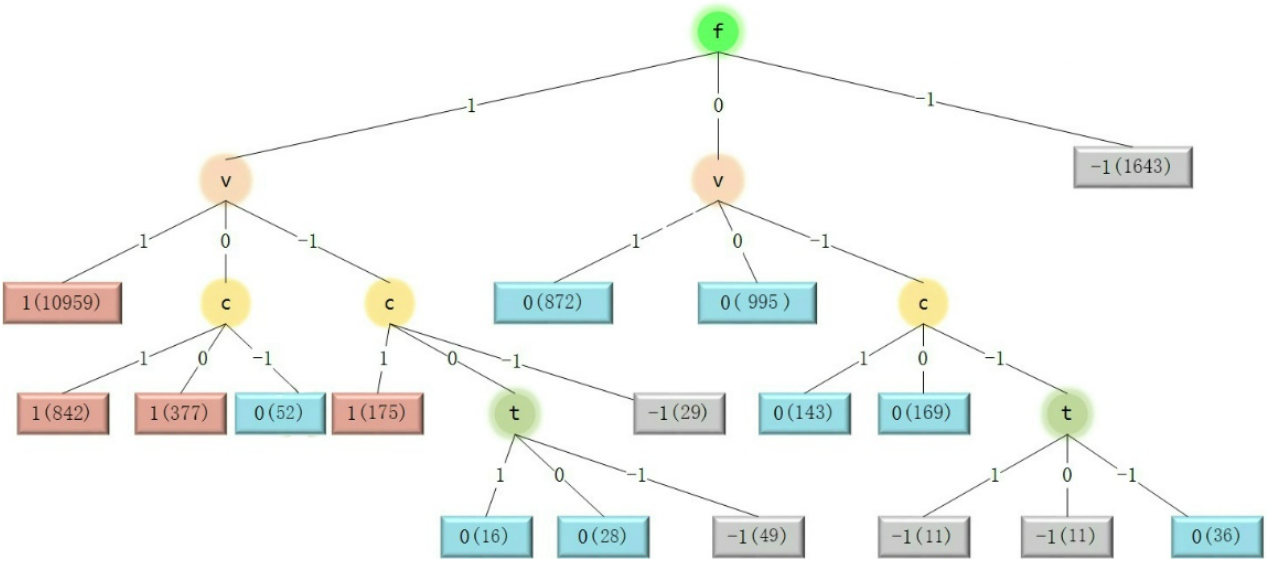
Influential Aspect Analysis

Aspect ratings such as flavor accuracy, throat hit, value, and cloud production reflect users' feelings about more specific aspects of an e-liquid product. The overall rating score is a mixture of product quality and the customer's overall interest in the product. Analyzing the relationship between aspect ratings and overall rating can help identify important aspects that influence users' interest in a product and impact marketing or product decisions.

Influential aspect analysis was performed on 16,407 reviews with no missing aspect ratings. The decision tree constructed in this analysis process is shown in Figure 2. Specifically, the label on the branch node means that this dataset is split into subsets according to the corresponding aspect. For example, the label "f" on the root node meant that the dataset was split into 3 subsets according to the value of the flavor accuracy aspect. The label on the edge from a parent node to a child node represented a condition. As another example, the label "1" on the edge from the root node to the leftmost child node indicated that reviews were split into the leftmost child node if their flavor accuracy aspects were positive. The label on the leaf node was the predicted sentiment of reviews that belong to this node, and the number in parentheses referred to the number of reviews on the node.

Table 1 shows the normalized information gain computed with the ID3 algorithm. The aspect with higher information gain is more important. According to results, flavor accuracy and value were the two most important aspects that influence users' sentiments toward e-liquids.

Figure 2. The decision tree constructed on reviews without missing aspect ratings. C: cloud production; f: flavor accuracy; t: throat hit; v: value; 1: positive; 0: neutral; -1: negative.



XSL•FO
RenderX

Table 1. Normalized information gain of each aspect.

Aspect	Normalized information gain
Flavor accuracy	0.8912
Value	0.1022
Cloud production	0.0049
Throat hit	0.0017

Table 2. The number of reviews for each flavor category.

Flavor	Number of reviews
Coffee	282
Cool	1609
Creamy	4056
Fruity	9653
Nutty	625
Rich	3268
Spiced	1089
Sweet	5128
Tobacco	1360

Table 3. Sentiment analysis of reviews for fruity and sweet flavors.

Reviews	Fruity (n=9653), n (%)	Sweet (n=5128), n (%)
Positive	6381 (66.10)	3315 (64.65)
Negative	1233 (12.77)	739 (14.41)
Neutral	2039 (21.12)	1074 (20.94)

Statistics of Reviews for Each Flavor Category

The numbers of reviews for each flavor category are listed in Table 2. Flavors with more reviews were more popular. Table 2 shows that fruity and sweet were the two most popular categories.

Furthermore, we counted the numbers of positive, negative, and neutral reviews for these two popular flavors. As shown in Table 3, both flavors had more positive reviews than negative reviews. Sweet flavors had a higher percentage of negative reviews than fruity flavors.

Opinion Sentiment Summarization

By mining the opinion sentiment summarizations of flavor accuracy, throat hit, value, and cloud production aspects for different flavors, decision makers and businesses have the opportunity to know why users like or dislike the aspect, thus gaining better understanding of users' vaping experience. Multimedia Appendix 1 shows identified aspect words. Flavor-related words included "flavor," "juice," "vape," "taste," "aftertaste," etc. Value-related words included "price," "value," "quality," etc. Cloud production-related words included "vapor production," "vapor," "cloud production," etc. Throat hit-related words included "throat," "hit," "throat hit," etc.

Multimedia Appendix 2 shows opinion summarization of the flavor accuracy aspect for fruity and sweet flavors. Fruity flavors cover a wide range, and since different flavors have different tastes, they have the most positive and negative reviews. Users were satisfied with fruity and sweet flavors with tastes such as "great," "sweet," "good," "strong," and "nice;" "weak," "sour," "bad," and "terrible" tastes made users dislike fruity flavor.

Reviews with value aspect ratings ≥4 and <3 were used to generate positive and negative opinions for value aspects, respectively. Multimedia Appendix 3 shows opinion summarization of value aspect for fruity and sweet flavors. There were more opinions about price and quality, indicating that they were two key concerns about value. Products with "great," "good," and "reasonable" prices can attract more user attention; "steep," "expensive," and "crazy" prices can make users dislike the product.

Multimedia Appendix 4 shows opinion summarization of the throat hit aspect for fruity and sweet flavors. Users liked fruity and sweet flavors with a throat hit that was "strong," "good," "nice," and "perfect"; users disliked these flavors when the throat hit was "nonexistent," "weak," "unpleasant," and "harsh." Specifically, users preferred strong throat hit the most and disliked harsh throat hit the most.

XSL•FO
RenderX

Table 4. The number of products for each flavor category.

Flavor	Number of products
Coffee	104
Cool	459
Creamy	920
Fruity	2342
Nutty	222
Rich	1033
Spiced	439
Sweet	2049
Tobacco	490

Table 5. Sentiment analysis of products for fruity and sweet flavors.

Sentiment	Fruity (n=2342), n (%)	Sweet (n=2049), n (%)
Positive products (%)	1307 (55.81)	1226 (59.83)
Neutral products (%)	716 (30.57)	559 (27.28)
Negative products (%)	319 (13.62)	264 (12.88)

Multimedia Appendix 5 shows opinion summarization for the cloud production aspect for fruity and sweet flavors. Generally, users were satisfied with “great” and “good” cloud production and were not satisfied with “poor” cloud production.

Product Statistics for Each Flavor Category

We regarded products whose average overall ratings were ≥4 as positive products, <4 and ≥3 as neutral products, and <3 as negative products. Then, we counted the number of products for each flavor category. The result is shown in Table 4.

Fruity and sweet products were the two most popular e-liquids. The sentiment distribution of products for fruity and sweet flavor is presented in Table 5. Furthermore, we extracted opinions for fruity and sweet products.

Positive and Negative Product Opinions

The positive and negative opinions for fruity and sweet products are shown in Multimedia Appendix 6. In addition to “great flavor” and “good juice,” users also expressed their love for fruity products with “great vape.” This suggested that good vapor contributes to positive reviews of fruity products. However, negative reviews were attributed to bad tastes, which were expressed by “soapy flavor,” “odd taste,” and so on.

Discussion

Principal Findings

This study provides a sentiment analysis of users’ ENDS vaping experience from review sites. By analyzing influential factors and opinions, we revealed users’ e-liquid preferences. Our findings may help businesses and policy makers better understand the advantages, disadvantages, and potential health risks of e-cigarette products, thus helping them to further improve product design and provide decision-making references.

Based on results obtained by the ID3 algorithm, flavor accuracy (normalized information gain=0.8912) and value (normalized information gain=0.1022) were the two most important aspects that influence users’ sentiments toward e-liquids. For the value aspect, users were concerned with price and quality; thus, a business can attract users by providing inexpensive and high-quality products, and policy makers can develop policies to manage and monitor their price and quality.

Previous research has shown that flavor has been found to be an attractive factor to ENDS users [31,32,33]. It is broadly used in Web-based social media advertisements and offline store promotions to increase the appeal of e-cigarette products [34]. Fruity and sweet were the two most popular flavors. Users’ flavor preference closely related to positive or negative content. By using sentiment opinion summarization techniques, we could reveal more information and flavor patterns among users. Opinion summarization gave reasons users like or dislike flavors. For example, opinions such as “good/great juice” were usually adopted to express users’ positive sentiments toward e-liquids. Opinions such as “sweet/strong flavor” indicated that users liked fruity and sweet flavors because of sweet and strong tastes. The result was consistent with a previous study [35], indicating that candy-like flavors could increase the appeal to starters because they mask the heavy cigarette taste; furthermore, adding candy-like flavors could potentially be perceived as enjoyable. We found that good or great or nice juice and fruity or sweet flavor might make users dependent on or be addicted to the product. Words such as “adv (all day vape)” and “addicting/be addicted to” were used to describe these feelings. Among 8186 posts containing adjectives in the positive opinion summarization for fruity and sweet flavor, the number of posts containing “adv” and “addicting/be addicted to” were 1110 and 46, respectively. For example, some users expressed their feelings as follows: “This juice was absolutely delicious and a great adv;” “this is my adv (all day vape), i love the taste of the

XSL•FO
RenderX

smooth caramel paired with the crisp green apple flavor. Very addicting!!!!;" "Sweet flavor that is nice for an ADV;" "I am addicted to this juice."

Opinions such as "bad juice/terrible flavor/harsh throat hit" described why users disliked these flavors. E-cigarette flavorings could potentially be harmful to users. Prior research has found that the majority of users reported negative sentiments about symptoms. Negative symptom words included "dry," "nausea," "burn," "hurt," "sore," "tingle," "fatigue," "sick," "toothache," "cough," and "headache" [22,24]. Among 929 posts containing adjectives in the negative opinion summarization for fruity and sweet flavor, the number of posts containing negative symptoms was 38. For example, users described symptoms caused by flavors as follows: "I do get a headache from all the sweeteners if I vape too much too quickly;" "Lemon vapes give me a headache;" and "The harsh throat hit makes me cough."

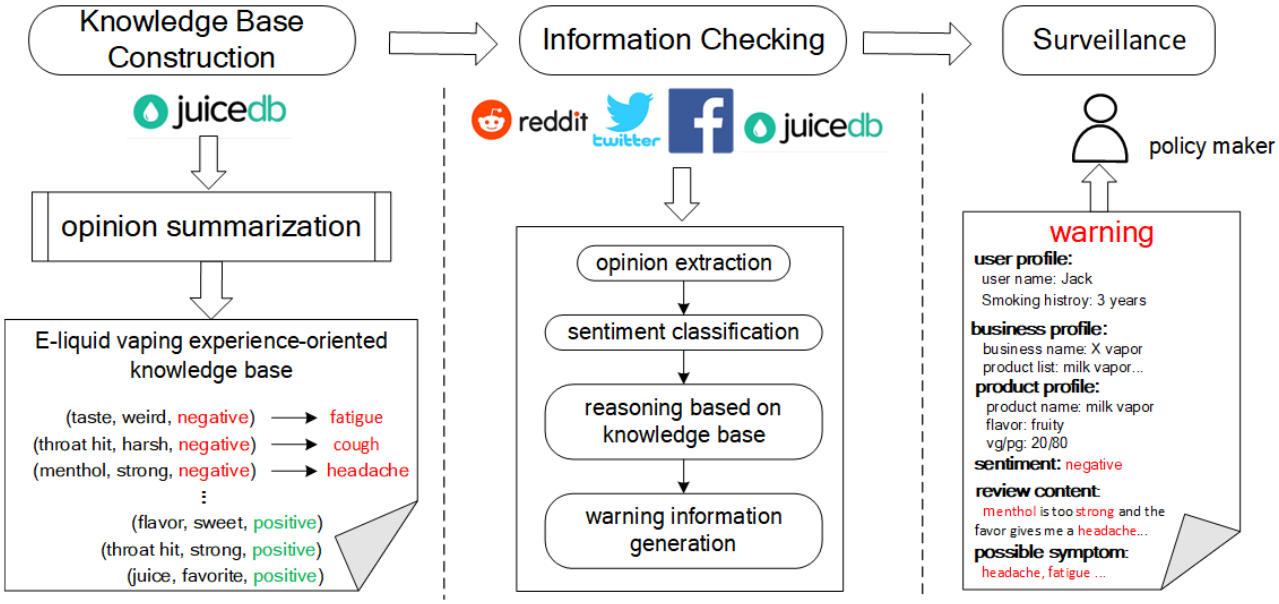
Our research shows that both attractiveness and negative symptoms of fruity and sweet products had effects on users' health. Policy makers need to pay more attention to these products and take appropriate regulatory action to reduce health risk. For example, they may formulate a comprehensive policy to manage ingredients, dosage, and sales of such products.

The proposed method for analyzing vaping behavior also has the potential to be used for surveillance and detection of health-related activities on other platforms. Figure 3 shows an application scenario of the proposed framework, which can be used to monitor e-liquid product information automatically. Consider a simple example. First, we can construct an e-liquid vaping experience-oriented knowledge base, including "throat hit, harsh, negative, cough," "menthol, strong, negative, and headache." Furthermore, we may automatically monitor incoming information from multiple platforms including Reddit, Twitter, Facebook, and JuiceDB. When the discovery of e-liquid may be harmful to human health, the system will generate prompt warnings. For instance, incoming posts like "The menthol is strong, I feel headache" and "After vaping it all day, all week, all month, I begin to cough" will be labeled as negative, highlighted, and sent to regulatory authorities. At the same time, prevention messages could be delivered to users at risk for harm associated with e-liquid use, thus realizing automatic supervision of product information across platforms.

Contributions

The rapid growth of ENDS, or e-cigarettes, indicates the importance of research in this field. Social media plays an indispensable role in providing new insights on e-cigarettes to help inform future research, regulations, and surveillance. Previous research has mainly utilized social media including Twitter, Facebook, YouTube, and Reddit as data sources to study e-cigarettes. Review sites such as JuiceDB provide a novel channel for users to discuss vaping methods and features; however, systematic studies on mining users' e-liquid usage patterns from review websites are still missing. This study contributes to the field by analyzing users' ENDS vaping experience from reviews using sentiment summarization. Specifically, we found that flavor accuracy and value were the two most important aspects that influence users' sentiments toward e-liquids. Of reviews in JuiceDB, 67.83% (18,362/27,070) were positive, while 12.67% (3430/27,070) were negative. This indicates that users generally hold positive attitudes toward e-liquids. Among the 9 flavors, fruity and sweet were the two most popular. Great and sweet tastes, reasonable values, and strong throat hit satisfied users with "fruity" and "sweet" flavors, whereas "strange" tastes made users dislike these flavors. Meanwhile, users complained about steep or expensive prices, bad quality, and harsh throat hit of some e-liquids. There were 2342 fruity e-liquids and 2049 sweet e-liquids. There were 55.81% (1307/2342) and 59.83% (1226/2049) positive sentiments and 13.62% (319/2342) and 12.88% (264/2049) negative sentiments toward fruity e-liquids and sweet e-liquids, respectively. Great flavor and good vapor contributed to positive reviews of fruity and sweet products.

Figure 3. Framework showing automatic supervision of e-liquid product information.



XSL•FO
RenderX

However, bad tastes such as “sour” and “bitter” resulted in negative reviews. Mined data-driven findings can help businesses and policy makers to further improve product quality and formulate effective policy.

Limitations

We collected review data only from JuiceDB—feasible for our current research. However, several other social media platforms, such as Twitter, Facebook, and E-cigarette Forum, could be jointly used to implement cross-platform sentiment analysis.

Another limitation of this paper was incomplete demographic information. Because JuiceDB does not provide complete personal characteristics, specifically, age and gender, we could not divide our dataset into several subgroups to analyze different usage patterns among different age or gender groups.

Finally, this study used only sentiment summarization methods to mine users’ ENDS vaping experiences. Many other data mining tools could be applied to explore the dataset further. For instance, more advanced topic association methods could be adopted to discover associations between flavors and symptoms.

Future Research

We envision three possible approaches for future study. First, the influential aspect analysis model could be extended by integrating aspect ratings and review content. In this study, we applied the ID3 algorithm to identify the relationship between aspect ratings and overall ratings; however, the review content provides more detailed semantic description information about aspect ratings. We believe that integrating these two kinds of information could produce more insights about what aspects influence users’ attitude toward e-liquid products.

Second, the aspect sentiment opinion summarization model provides basic components for analyzing aspect and product opinions. More advanced algorithms can be used to extend the model, to cluster similar opinions, and to generate more explainable opinions.

Finally, other social media platforms such as other review sites, Twitter, Reddit, etc can be considered to implement cross-platform sentiment analysis. It will be challenging and meaningful to develop a tool to monitor e-liquid product information automatically and profvide timely, valuable signals for management departments to make better decisions.

Conclusion

This study provides an effective mechanism for analyzing users’ ENDS vaping experience based on sentiment opinion summarization techniques. Sentiment opinions for aspect and product can be found using our method, which is of great importance for monitoring e-liquid products and improving work efficiency of management departments. We hope that the characteristics we reported in this paper can be useful for other researchers and policy makers.

Acknowledgments

This work was supported by the National Key R&D Program of China under Grant No. 2016QY02D0305, the US National Institutes of Health under Grant No. 5R01DA037378-05, National Natural Science Foundation of China under Grant Nos. 61671450, 71621002, 71272236, and The Key Research Program of the Chinese Academy of Sciences under Grant No. ZDRW-XH-2017-3.

Authors' Contributions

QL, DDZ, and SJL conceived the idea for this study. QL designed the study, conducted data analysis and wrote the manuscript. CW, RL, and DDZ contributed to the manuscript and interpretation of study findings. LW and SJL contributed to the manuscript and provided critical feedback on it. All authors read and approved the final manuscript.

Conflicts of Interest

SJL has served as a paid consultant to or conducted research for Pfizer, GSK, Cypress BioScience, and McNeil Consumer. McNeil Consumer is collaborating with GSK on a current study on nicotine replacement, which is being conducted by SJL, and GSK markets bupropion.

Multimedia Appendix 1

Aspect words.

[PNG File, 71KB - jmir_v20i8e252_app1.png]

Multimedia Appendix 2

Opinion sentiment summarization for flavor accuracy aspect.

[PNG File, 155KB - jmir_v20i8e252_app2.png]

Multimedia Appendix 3

Opinion sentiment summarization for value aspect.

XSL•FO
RenderX

[PNG File, 181KB - jmir_v20i8e252_app3.png]

Multimedia Appendix 4

Opinion sentiment summarization for throat hit aspect.

[PNG File, 226KB - jmir_v20i8e252_app4.png]

Multimedia Appendix 5

Opinion sentiment summarization for cloud production aspect.

[PNG File, 158KB - jmir_v20i8e252_app5.png]

Multimedia Appendix 6

Opinions on products belonging to fruity and sweet categories.

[PNG File, 168KB - jmir_v20i8e252_app6.png]

References

1. PR Newswire. Global E-Cigarette Market 2017-2022: Growing Health Awareness, Advancement in Electronic Device Technology, Smoke & Ash Less Vaping - Research and Markets URL: https://tinyurl.com/y7el52o2 [WebCite Cache ID 6ytcvhnYU]
2. Centers for Disease Control and Prevention. Electronic Cigarettes URL: https://www.cdc.gov/tobacco/basic_information/e-cigarettes/index.htm [accessed 2017-09-27] [WebCite Cache ID 6tmwujQZ0]
3. US Food & Drug Administration. Vaporizers, E-Cigarettes, and other Electronic Nicotine Delivery Systems (ENDS) URL: https://www.fda.gov/TobaccoProducts/Labeling/ProductsIngredientsComponents/ucm456610.htm [accessed 2017-09-27] [WebCite Cache ID 6tmx9YC2e]
4. Centers for Disease Control & Prevention. Reasons for Electronic Cigarette Use Among Middle and High School Students — National Youth Tobacco Survey, United States, 2016 URL: https://www.cdc.gov/mmwr/volumes/67/wr/mm6706a5.htm [accessed 2018-04-23] [WebCite Cache ID 6yuD2ufXl]
5. U.S. Food & Drug Administration. FDA takes significant steps to protect Americans from dangers of tobacco through new regulation URL: https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm499234.htm [accessed 2017-09-27] [WebCite Cache ID 6tmxWx8zB]
6. Fan D. LinkedIn. 2014. PG vs VG – All Things You Should Know About E-liquid URL: https://www.linkedin.com/pulse/20140610083157-323109215-pg-vs-vg-all-things-you-should-know-about-e-liquid [accessed 2018-05-15] [WebCite Cache ID 6zQFkatf1]
7. Vaperanks. Propylene Glycol vs Vegetable Glycerin E-Liquid ? What's the Difference? URL: http://vaperanks.com/propylene-glycol-vs-vegetable-glycerin-e-liquid-whats-the-difference/ [accessed 2017-09-27] [WebCite Cache ID 6tmydrnUy]
8. eCig One. PG vs. VG URL: http://ecigone.com/e-cigarette-basics/pg-vs-vg/ [accessed 2017-09-27] [WebCite Cache ID 6tmynuwUg]
9. Vapei. E-Liquid Beginners Guide URL: https://www.vapei.com/blog/e-liquid-beginners-guide/ [accessed 2017-11-10] [WebCite Cache ID 6urg1DHAH]
10. Wang F, Carley KM, Zeng D, Mao W. Social Computing: From Social Informatics to Social Intelligence. IEEE Intell. Syst 2007 Mar;22(2):79-83. [doi: 10.1109/MIS.2007.41]
11. Yan P, Chen H, Zeng D. Syndromic surveillance systems. In: Ann. Rev. Info. Sci. Tech. Hoboken, New Jersey: Wiley Online Library; Nov 05, 2009:425-495.
12. Kim AE, Hopper T, Simpson S, Nonnemaker J, Lieberman AJ, Hansen H, et al. Using Twitter Data to Gain Insights into E-cigarette Marketing and Locations of Use: An Infoveillance Study. J Med Internet Res 2015;17(10):e251 [FREE Full text] [doi: 10.2196/jmir.4466] [Medline: 26545927]
13. Liang Y, Zheng X, Zeng DD, Zhou X, Leischow SJ, Chung W. Exploring how the tobacco industry presents and promotes itself in social media. J Med Internet Res 2015;17(1):e24 [FREE Full text] [doi: 10.2196/jmir.3665] [Medline: 25608524]
14. Chu KH, Sidhu AK, Valente TW. Electronic Cigarette Marketing Online: a Multi-Site, Multi-Product Comparison. JMIR Public Health Surveill 2015;1(2):e11 [FREE Full text] [doi: 10.2196/publichealth.4777] [Medline: 27227129]
15. 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]
16. Cole-Lewis H, Pugatch J, Sanders A, Varghese A, Posada S, Yun C, et al. Social Listening: A Content Analysis of E-Cigarette Discussions on Twitter. J Med Internet Res 2015;17(10):e243 [FREE Full text] [doi: 10.2196/jmir.4969] [Medline: 26508089]

XSL•FO
RenderX

17. Cole-Lewis H, Varghese A, Sanders A, Schwarz M, Pugatch J, Augustson E. Assessing Electronic Cigarette-Related Tweets for Sentiment and Content Using Supervised Machine Learning. J Med Internet Res 2015;17(8):e208 [FREE Full text] [doi: 10.2196/jmir.4392] [Medline: 26307512]
18. Lazard AJ, Saffer AJ, Wilcox GB, Chung AD, Mackert MS, Bernhardt JM. E-Cigarette Social Media Messages: A Text Mining Analysis of Marketing and Consumer Conversations on Twitter. JMIR Public Health Surveill 2016 Dec 12;2(2):e171 [FREE Full text] [doi: 10.2196/publichealth.6551] [Medline: 27956376]
19. Harris JK, Moreland-Russell S, Choucair B, Mansour R, Staub M, Simmons K. Tweeting for and against public health policy: response to the Chicago Department of Public Health's electronic cigarette Twitter campaign. J Med Internet Res 2014;16(10):e238 [FREE Full text] [doi: 10.2196/jmir.3622] [Medline: 25320863]
20. Huang J, Kornfield R, Emery SL. 100 Million Views of Electronic Cigarette YouTube Videos and Counting: Quantification, Content Evaluation, and Engagement Levels of Videos. J Med Internet Res 2016 Mar 18;18(3):e67 [FREE Full text] [doi: 10.2196/jmir.4265] [Medline: 26993213]
21. Wang L, Zhan Y, Li Q, Zeng DD, Leischow SJ, Okamoto J. An Examination of Electronic Cigarette Content on Social Media: Analysis of E-Cigarette Flavor Content on Reddit. Int J Environ Res Public Health 2015 Nov;12(11):14916-14935 [FREE Full text] [doi: 10.3390/ijerph121114916] [Medline: 26610541]
22. Li Q, Zhan Y, Wang L, Leischow SJ, Zeng DD. Analysis of symptoms and their potential associations with e-liquids' components: a social media study. BMC Public Health 2016 Jul 30;16:674 [FREE Full text] [doi: 10.1186/s12889-016-3326-0] [Medline: 27475060]
23. Jin Z, Li Q, Zeng DD, Zhan Y, Liu R, Wang L, et al. Jointly modeling review contentaspect ratings for review rating prediction. 2016 Jul 17 Presented at: Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval; 2016; Pisa, Italy p. 893-896. [doi: 10.1145/2911451.2914692]
24. Zhan Y, Liu R, Li Q, Leischow SJ, Zeng DD. Identifying Topics for E-Cigarette User-Generated Contents: A Case Study From Multiple Social Media Platforms. J Med Internet Res 2017 Jan 20;19(1):e24 [FREE Full text] [doi: 10.2196/jmir.5780] [Medline: 28108428]
25. Chen Z, Zeng DD. Mining online e-liquid reviews for opinion polarities about e-liquid features. BMC Public Health 2017 Dec 07;17(1):633 [FREE Full text] [doi: 10.1186/s12889-017-4533-z] [Medline: 28683797]
26. JuiceDB. URL: https://www.juicedb.com/ [accessed 2018-03-30] [WebCite Cache ID 6yIjEAJNe]
27. Quinlan JR. Induction of decision trees. In: Mach Learn. Dordrecht: Kluwer Academic Publishers; Mar 1986:81-106.
28. Li Q, Jin Z, Wang C, Zeng DD. Mining opinion summarizations using convolutional neural networks in Chinese microblogging systems. Knowledge-Based Systems 2016 Sep;107:289-300 [FREE Full text] [doi: 10.1016/j.knosys.2016.06.017]
29. The Stanford Natural Language Processing Group. The Stanford Parser: A statistical parser URL: https://nlp.stanford.edu/software/lex-parser.html [accessed 2018-04-15] [WebCite Cache ID 6ygqX99L4]
30. UIC Computer Science. Opinion Mining, Sentiment Analysis, and Opinion Spam Detection URL: https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html [accessed 2017-11-10] [WebCite Cache ID 6urdXaED2]
31. Choi K, Fabian L, Mottey N, Corbett A, Forster J. Young adults' favorable perceptions of snus, dissolvable tobacco products, and electronic cigarettes: findings from a focus group study. Am J Public Health 2012 Nov;102(11):2088-2093 [FREE Full text] [doi: 10.2105/AJPH.2011.300525] [Medline: 22813086]
32. Etter J, Zäther E, Svensson S. Analysis of refill liquids for electronic cigarettes. Addiction 2013 Sep;108(9):1671-1679. [doi: 10.1111/add.12235] [Medline: 23701634]
33. McDonald EA, Ling PM. One of several 'toys' for smoking: young adult experiences with electronic cigarettes in New York City. Tob Control 2015 Nov;24(6):588-593 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051743] [Medline: 25564287]
34. Cheney M, Gowin M, Wann TF. Marketing practices of vapor store owners. Am J Public Health 2015 Jun;105(6):e16-e21. [doi: 10.2105/AJPH.2015.302610] [Medline: 25880960]
35. Kostygina G, Glantz SA, Ling PM. Tobacco industry use of flavours to recruit new users of little cigars and cigarillos. Tob Control 2016 Jan;25(1):66-74 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051830] [Medline: 25354674]

Abbreviations

Adv: all day vape
API: application programming interface
ENDS: electronic nicotine delivery systems
FDA: Food and Drug Administration
ID3: iterative dichotomiser 3
PG: propylene glycol
TF: term frequency
VG: vegetable glycerin

XSL•FO
RenderX

Edited by G Eysenbach; submitted 10.11.17; peer-reviewed by A Mavragani, R Hilscher; comments to author 29.03.18; revised version received 21.05.18; accepted 10.07.18; published 15.08.18.

Please cite as:
Li Q, Wang C, Liu R, Wang L, Zeng DD, Leischow SJ
Understanding Users' Vaping Experiences from Social Media: Initial Study Using Sentiment Opinion Summarization Techniques
J Med Internet Res 2018;20(8):e252
URL: http://www.jmir.org/2018/8/e252/
doi:10.2196/jmir.9373
PMID:30111530

©Qiudan Li, Can Wang, Ruoran Liu, Lei Wang, Daniel Dajun Zeng, Scott James Leischow. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 15.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Information and Communication Technologies Interest, Access, and Use: Cross-Sectional Survey of a Community Sample of Urban, Predominantly Black Women

Sarah M Jabour[1], BA; Alexis Page[1], BA; Seventy F Hall[2], BS; Lycinda Rodriguez[3], MS; Wendy C Shields[4], MPH; Anika AH Alvanzo[1], MD, MS

[1]Division of General Internal Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD, United States

[2]School of Social Work, University of Buffalo, Buffalo, NY, United States

[3]Department of Psychological, Health and Learning Sciences, University of Houston, Houston, TX, United States

[4]Department of Health Policy and Management, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

Corresponding Author:
Anika AH Alvanzo, MD, MS
Division of General Internal Medicine
School of Medicine
Johns Hopkins University
1830 East Monument, Room 8069
Baltimore, MD, 21287
United States
Phone: 1 4105500094
Fax: 1 4105500060
Email: aalvanz1@jhmi.edu

Abstract

Background: Information and communication technologies (ICT) offer the potential for delivering health care interventions to low socioeconomic populations who often face barriers in accessing health care. However, most studies on ICT for health education and interventions have been conducted in clinical settings.

Objective: The aim of this study was to examine access to and use of mobile phones and computers, as well as interest in, using ICT for receipt of behavioral health information among a community sample of urban, predominately black, women with low socioeconomic status.

Methods: Participants (N=220) were recruited from hair salons and social service centers and completed audio-computer assisted self-interviews.

Results: The majority of the participants (212/220, 96.3%) reported use of a cell phone at least weekly, of which 89.1% (189/212) used smartphones and 62.3% (137/220) reported computer use at least weekly. Of the women included in the study, 51.9% (107/206) reported using a cell phone and 39.4% (74/188) reported using a computer to access health and/or safety information at least weekly. Approximately half of the women expressed an interest in receiving information about stress management (51%-56%) or alcohol and health (45%-46%) via ICT. Smartphone ownership was associated with younger age (odds ratio [OR] 0.92, 95% CI 0.87-0.97) and employment (OR 5.12, 95% CI 1.05-24.95). Accessing health and safety information weekly by phone was associated with younger age (OR 0.96, 95% CI 0.94-0.99) and inversely associated with higher income (OR 0.42, 95% CI 0.20-0.92).

Conclusions: Our findings suggest that ICT use, particularly smartphone use, is pervasive among predominantly black women with low socioeconomic status in urban, nonclinical settings. These results show that ICT is a promising modality for delivering health information to this population. Further exploration of the acceptability, feasibility, and effectiveness of using ICT to disseminate behavioral health education and intervention is warranted.

(J Med Internet Res 2018;20(8):e248) doi:10.2196/jmir.9962

KEYWORDS

women; internet communication technology; urban; alcohol; stress; smartphones

XSL•FO
RenderX

Introduction

Information and communication technologies (ICT) use has become widespread across the United States. An estimated 73% of American adults own computers and as many as 92% own a cell phone, 68% of which are smartphones, cellular devices with internet capabilities [1]. Although earlier studies found racial and ethnic disparities in ICT accessibility [2-4], recent studies have suggested that this divide is narrowing, particularly related to cell phone use. Black Americans are equally as likely as white Americans to own a cell phone, more likely to report cell phones as a primary internet source and use a wider array of cell phone data functions compared to their white American counterparts [5]. The availability of ICT offers a promising vehicle for dissemination of health education and interventions in a convenient delivery format [6,7].

ICT may be particularly useful for intervening on potentially stigmatizing conditions, such as behavioral health disorders involving substance misuse and/or mental health. Multiple studies have shown that people are more likely to disclose sensitive behaviors when the data are collected using computers as compared to more traditional methods, such as a face-to-face interview [8-10]. Evidence suggests that internet-based interventions can effectively reduce drinking [7,11-13] and are a promising way to improve mental health symptoms [14,15]. This may be especially relevant for minority populations living in low-income, urban neighborhoods, as members of minority populations are less likely to access health care in traditional medical settings; and services for co-occurring mental health and substance misuse are less likely to be found in urban neighborhoods [16-19]. Thus, ICT-based interventions warrant exploration of their potential for reducing health disparities by engaging vulnerable populations in urban medically underserved environments who are often underrepresented in traditional clinical settings [20,21].

Although age, income, and educational attainment remain barriers to accessing health information through ICT [5,21-23], findings suggest that the majority of low-income patients recruited from urban health care settings have access to, and are generally interested in, the use of ICT for health advice and communication with their families' health providers [22,24]. However, less is known about ICT use and access of behavioral health information among nonclinical urban populations. This exploratory analysis aimed to address gaps in the literature by examining the availability and use of ICT as well as the interest in ICT-based information on alcohol use and stress symptoms management among women recruited from urban community-based sites.

Methods

Participants

Participants (N=220) were recruited for a larger study examining the prevalence and relationship between trauma, posttraumatic stress, and high-risk drinking in predominantly black women. Between June 2014 and September 2016, women were recruited from 3 urban hair salons (n=57) and 4 Community Action Partnership Centers (CAPCs). CAPCs are city social service centers providing a variety of resources, including housing and energy assistance (n=163). English speaking women between the ages of 21 to 65 years old were eligible. Women who were pregnant, currently enrolled in treatment for substance use, or unable to provide informed consent were excluded.

Procedures and Measures

Research assistants approached women in the waiting rooms of the hairs salons and CAPCs. After completing a screening questionnaire, eligible women provided informed consent and used a tablet computer equipped with headphones to complete a battery of instruments via audio-computer assisted self-interview (ACASI). The survey included demographic questions and the Technology Use Survey, a 20-item questionnaire on cell phone, computer, and internet use adapted from an emergency department study examining parents' access of child health and safety information [24]. The survey was modified to include questions on access to and interest in receiving alcohol and stress management health information via ICT. The study was approved by the Johns Hopkins Institutional Review Board.

Data Analyses

Descriptive statistics (frequencies, means, and standard errors) were used to describe the sample. Exploratory logistic regression analyses were used to examine the women's use of ICT based on demographic variables of interest (age, education, employment, and income). The frequency of cell phone and computer use was dichotomized into daily and less than daily use, and frequency of accessing health and safety information was dichotomized into weekly and less than weekly access. Most of the independent variables were binary except age, which was continuous, and education (< high school, high school or general equivalency degree, > high school). Annual household income was also dichotomized (<US $25,000 and ≥US $25,000). Given that this study targeted predominantly black women, race was dichotomized into black and nonblack. However, race was excluded from the logistic regression analyses because it failed criteria for inclusion (P<.10) during univariable analyses. Significance was determined at alpha level of .05. Analyses were performed using SPSS version 24.0 for Windows.

Results

Demographics

Most women were in their mid-forties (mean [SD] of 44.7 [12.11] years), black (185/220, 84.1%), had at least a high school education (173/220, 78.6%), and reported annual household income of less than US $25,000 (140/220, 64.5%). However, the average household income was largely influenced by women recruited from the CAPCs, as more than half of these women (84/163, 51.5%) reported household incomes of less than US $10,000 annually. When compared to women recruited from the CAPCs, women recruited from salons were roughly four years younger (41.8 vs 45.7; t= 2.07, P=.04), more likely to be black (P=.03), married (χ^2_2=15.06; P=.001), and employed (χ^2_1=33.00; P<.001), were more educated (χ^2_2=37.47; P<.001),

XSL•FO
RenderX

and had higher annual household incomes (χ^2_1=109.17; *P*<.001). Sample demographic characteristics are presented in Table 1.

Cell Phone Use

The results from the cell phone section of the questionnaire are presented in Table 2. The overwhelming majority of women (212/220, 96.4%) reported using a cell phone at least weekly. Of those reporting at least weekly cell phone use, the majority reported personal ownership of the phone (206/212, 97.2%), of which a majority were smartphones (189/212, 89.1%). Most women (181/206, 87.9%) had a monthly cell phone plan and reported <5 days in which their phones were not working in the previous three months (180/205, 87.9%). A substantial majority used their cell phones to send or receive emails (145/206, 70.4%) or to access the internet daily (148/206, 71.8%). As shown in Table 2, there were some statistically significant differences for certain characteristics of cell phone use between women recruited from the salons and those recruited from the CAPCs, but not for smartphone ownership or daily internet use via cell phones.

Computer Use

Two-thirds of the women (147/220, 66.7%) reported at least weekly use of a computer and it was found that the women were accessing computers in multiple locations (Table 3). The home was the most frequent place where participants reported accessing the internet via computer (115/187, 61.5%), followed by work (58/187, 31.0%), and the library (54/187, 28.9%). Just over half reported using the computer to email (97/188, 51.6%) or access the internet (107/188, 56.9%) on a daily basis. Women recruited from salons were more likely to report daily email messaging (41/54, 75.9% vs 56/134, 41.8%; χ^2_4=24.48; *P*<.001) and internet access via a computer (42/54, 77.8% vs 65/134, 48.5%; χ^2_4=18.59; *P*<.001), respectively, when compared to women recruited from the CAPCs.

Table 1. Demographic characteristics of participants.

Characteristic	Total recruited (n=220)	Recruited from salons (n=57)	Recruited from CAPCs[a] (n=163)	*P* value
Age, mean (SD)	44.7 (12.11)	41.8 (13.0)	45.7 (11.7)	.04
Race, n (%)				.03
Black	185 (84.1)	52 (91.2)	133 (81.6)	
White	22 (10.0)	1 (1.8)	21 (12.9)	
Other	13 (5.9)	4 (7.0)	9 (5.5)	
Education level, n (%)				<.001
≤ High school or GED[b]	47 (21.4)	2 (3.5)	45 (27.6)	
High school or GED	81 (36.8)	12 (21.1)	69 (42.3)	
>High school	92 (41.8)	43 (75.4)	49 (30.1)	
Marital status, n (%)				.001
Single (never married)	114 (51.8)	25 (43.9)	89 (54.6)	
Married or living as married	43 (19.6)	21 (36.8)	22 (13.5)	
Previously married (divorced, separated, or widowed)	63 (28.6)	11 (19.3)	52 (31.9)	
Employment status, n (%)[c]				
Employed for wages	87 (39.9)	41 (71.9)	46 (28.6)	<.001
Self-employed	8 (3.7)	2 (3.5)	6 (3.7)	1.00
Unemployed	38 (17.4)	5 (8.8)	33 (20.5)	.07
Disabled	50 (22.9)	0 (0.0)	50 (31.1)	<.001
Homemaker	13 (6.0)	0 (0.0)	13 (8.1)	.02
Student	10 (4.6)	4 (7.0)	6 (3.7)	.29
Retired	10 (4.6)	5 (8.8)	5 (3.1)	.13
Household income, n (%)				<.001
<US $25,000	140 (64.5)	3 (5.6)	137 (84.0)	
≥US $25,000	77 (35.5)	51 (94.4)	26 (16.0)	

[a]CAPC: Community Action Partnership Center.

[b]GED: General Equivalency Diploma.

[c]Respondents able to select multiple options.

XSL•FO
RenderX

Table 2. Cell phone use characteristics of participants.

Cell phone use characteristics	Total (n=220)	Salons (n=57)	CAPCs[a] (n=163)	*P* value
Do you use a cell phone at least once per week?				0.12
No	8 (3.6)	0 (0.0)	8 (4.9)	
Yes	212 (96.4)	57 (100)	155 (95.1)	
What kind of cell phone do you use most of the time?[b]				.33
Smartphone	189 (89.2)	53 (93.0)	136 (87.7)	
Other, nonsmartphone	23 (10.8)	4 (7.0)	19 (12.3)	
Does this phone belong to you or to someone else?[b]				.20
Belongs to me	206 (97.2)	57 (100)	149 (96.1)	
Belongs to someone else	6 (2.8)	0 (0.0)	6 (3.9)	
How long have you had the same phone number?[c]				<.001
≤1 year	57 (27.7)	5 (8.8)	52 (34.9)	
>1 year	146 (70.9)	49 (86.0)	97 (65.1)	
Don’t know	3 (1.5)	3 (5.3)	0 (0.0)	
In the last 3 months, how many days was your phone not working for any reason (disconnected, dead battery, etc)?[c]				<.001
0 days	127 (62.0)	47 (82.5)	80 (54.1)	
1-4 days	53 (25.9)	9 (15.8)	44 (29.7)	
>5 days	25 (12.2)	1 (1.8)	24 (16.2)	
What type of cell phone plan do you have?[c]				.004
Pay per month	181 (87.9)	56 (98.2)	125 (83.9)	
Pay as you go (you have to add minutes)	6 (2.9)	1 (1.8)	5 (3.4)	
Other	19 (9.2)	0 (0.0)	19 (12.8)	
Thinking just about cell phones, how often do you use a cell phone to:[c]				
Send or get text messages				.30
Daily	175 (85.0)	52 (91.2)	123 (82.6)	
Weekly to monthly	14 (6.8)	3 (5.3)	11 (7.4)	
Rarely to never	17 (8.3)	2 (3.5)	15 (10.1)	
Send or get email messages				.005
Daily	145 (70.4)	49 (86.0)	96 (64.4)	
Weekly to monthly	10 (4.9)	2 (3.5)	8 (5.4)	
Rarely or never	51 (24.8)	6 (10.5)	45 (30.2)	
Access the internet				.82
Daily	148 (71.8)	43 (75.4)	105 (70.5)	
Weekly to monthly	12 (5.8)	3 (5.3)	9 (6.0)	
Rarely or never	46 (22.3)	11 (19.3)	35 (23.5)	

[a]CAPC: Community Action Partnership Center.

[b]Includes responses only from those reporting at least weekly cell phone use.

[c]Includes responses only from those reporting at least weekly cell phone use and cell phone belongs to them.

XSL•FO
RenderX

Table 3. Computer use characteristics of participants.

Computer use characteristics	Total (n=220)	Salons (n=57)	CAPCs[a] (n=163)	*P* value
How often do you use a computer for any reason?				<.001
Daily	98 (44.5)	44 (77.2)	54 (33.1)	
Weekly to monthly	49 (22.3)	9 (15.8)	40 (24.5)	
Rarely to never	73 (33.2)	4 (7.0)	69 (42.3)	
Where do you use a computer to access the internet?[b,c]				
I don't use the internet	10 (5.3)	1 (1.9)	9 (6.7)	.29
Home	115 (61.5)	40 (75.5)	75 (56.0)	.02
Work	58 (31.0)	33 (62.3)	25 (18.7)	<.001
Library	54 (28.9)	6 (11.3)	48 (35.8)	.001
Other	38 (20.3)	11 (20.8)	27 (20.1)	.40
Thinking just about computers, how often do you use a computer to:[c]				
Send or get email messages				<.001
Daily	97 (51.6)	41 (75.9)	56 (41.8)	
Weekly to monthly	30 (16.0)	9 (16.7)	21 (15.7)	
Rarely to never	61 (32.4)	4 (7.4)	57 (42.5)	
Access the internet				<.001
Daily	107 (56.9)	42 (77.8)	65 (48.5)	
Weekly to monthly	34 (18.1)	9 (16.7)	25 (18.7)	
Rarely to never	47 (25.0)	3 (5.6)	44 (32.8)	

[a]CAPC: Community Action Partnership Center.

[b]Respondents able to select multiple options.

[c]Excludes those reporting never using a computer.

ICT and Access of Health and Safety Information

Approximately half of women reported using a cell phone (107/206, 52.0%) and one-third used a computer (74/188, 39.4%) to access health information at least weekly (Table 4). Just over half were interested in accessing stress management information on their cell phones (115/212, 54.2%) and computers (96/188, 51.5%), and just under half reported interest in accessing alcohol-related health information via cell phones (94/211, 44.5%) and computers (84/188, 44.7%).

Demographics and Information and Communication Technology Use

The exploratory logistic regression results are shown in Tables 5 and 6. Younger age was associated with all cell phone use and access variables (ie, smartphone ownership, daily text messaging, daily email, daily internet access, and weekly access of health and safety information via cell phone) but not with any of the computer use variables. Compared to women with less than a high school education, women with greater than a high school education were more likely to report daily use of a cell phone to access the internet. Employment was associated with owning a smartphone. Women with annual household incomes of ≥US $25,000 were four times more likely to use a computer daily for email messaging and almost three times more likely to access the internet daily on a computer. There was no association between income and cell phone use with the exception that women with higher incomes were less likely to report weekly access health and safety information via their cell phones.

XSL•FO
RenderX

Table 4. ICT (information and communication technology) and access of health and safety information.

Characteristic	Total (n=220)	Salons (n=57)	CAPCs[a] (n=163)	*P* value
Thinking just about mobile phones, how often do you use a mobile phone to:				
Get health or safety information[b]				.10
Daily	78 (37.9)	15 (26.3)	63 (42.3)	
Weekly	29 (14.1)	9 (15.8)	20 (13.4)	
Monthly	20 (9.7)	9 (15.8)	11 (7.4)	
Rarely to never	79 (38.3)	24 (42.1)	55 (36.9)	
Would you like to be able to use a mobile phone for information about dealing with stress?[c]				.09
No	97 (45.8)	32 (56.1)	65 (41.9)	
Yes	115 (54.2)	25 (43.9)	90 (58.1)	
Would you like to be able to use a mobile phone to get information about alcohol and health?[c]				53.
No	117 (55.5)	35 (61.4)	82 (53.2)	
Yes	94 (44.5)	22 (38.6)	72 (46.8)	
Thinking just about computers, how often do you use a computer to:[d]				
Get health or safety information				.01
Daily	47 (25.0)	14 (25.9)	33 (24.6)	
Weekly	27 (14.4)	9 (16.7)	18 (13.4)	
Monthly	23 (12.2)	13 (24.1)	10 (7.5)	
Rarely to never	91 (48.4)	18 (33.3)	73 (54.5)	
Would you like to be able to use a computer to get information about dealing with stress?				.75
No	92 (48.9)	25 (46.3)	67 (50.0)	
Yes	96 (51.1)	29 (53.7)	67 (50.0)	
Would you like to be able to use a computer to get information about alcohol and health?				.63
No	104 (55.3)	28 (51.9)	76 (56.7)	
Yes	84 (44.7)	26 (48.1)	58 (43.3)	

[a]CAPC: Community Action Partnership Center.

[b]Includes responses only from those reporting at least weekly cell phone use and cell phone belongs to them.

[c]Includes responses only from those reporting at least weekly cell phone use.

[d]Excludes those reporting never using a computer.

XSL•FO
RenderX

Table 5. Demographic associations with cell phone use.

Characteristic	Smartphone ownership, OR[a] (95% CI)	Daily text messages, OR (95% CI)	Daily email messages, OR (95% CI)	Daily internet access, OR (95% CI)	Weekly health/safety info, OR (95% CI)
Age	0.92 (0.87-0.97)[b]	0.88 (0.83-0.93)[b]	0.93 (0.90-0.96)[b]	0.89 (0.85-0.93)[b]	0.96 (0.94-0.99)[b]
Education					
<High school or GED[c] (reference)	1.00	1.00	1.00	1.00	1.00
High school or GED	0.45 (0.13-1.55)	1.06 (0.35-3.25)	1.26 (0.54-2.95)	1.18 (0.47-2.96)	1.30 (0.58-2.94)
>High school	0.99 (0.21-4.65)	0.52 (0.15-1.85)	2.09 (0.78-5.61)	5.10 (1.62-16.00)[b]	2.42 (0.99-5.93)
Employment[d]					
Not working (reference)	1.00	1.00	1.00	1.00	1.00
Working	5.12 (1.05-24.95)[b]	2.28 (0.71-7.30)	1.24 (0.55-2.80)	1.23 (0.50-3.05)	1.85 (0.91-3.74)
Income					
<US $25,000 (reference)	1.00	1.00	1.00	1.00	1.00
≥US $25,000	1.07 (0.27-4.22)	3.34 (0.96-11.59)	2.50 (1.00-6.26)[e]	0.57 (0.20-1.57)	0.42 (0.20-0.92)[b]

[a]OR: odds ratio.

[b]Statistically significant.

[c]GED: General Equivalency Diploma.

[d]Working defined as employed for wages and self-employed; not working defined as out of work, homemaker, student, retired, or disabled.

[e]Appears statistically significant only because of rounding.

Table 6. Demographic associations with computer use.

Characteristic	Daily email messages, OR[a] (95% CI)	Daily internet access, OR (95% CI)	Weekly health or safety info, OR (95% CI)
Age	1.00 (0.97-1.03)	1.00 (0.97-1.02)	1.00 (0.97-1.02)
Education			
<High school or GED[b] (reference)	1.00	1.00	1.00
High school or GED	0.98 (0.38-2.52)	0.88 (0.35-2.21)	1.57 (0.58-4.22)
>High school	1.48 (0.55-3.96)	1.43 (0.54-3.78)	2.38 (0.86-6.57)
Employment[c]			
Not working (reference)	1.00	1.00	1.00
Working	1.62 (0.79-3.32)	1.85 (0.90-3.78)	1.06 (0.52-2.13)
Income			
<US $25,000 (reference)	1.00	1.00	1.00
≥US $25,000	4.00 (1.88-8.51)[d]	2.87 (1.34-6.15)[d]	1.26 (0.62-2.59)

[a]OR: odds ratio.

[b]GED: General Equivalency Diploma.

[c]Working defined as employed for wages and self-employed; not working defined as out of work, homemaker, student, retired, or disabled.

[d]Statistically significant.

Discussion

Principal Results and Comparison with Prior Work

In this sample of predominantly black women with low socioeconomic status recruited from urban hair salons and social service centers, we found high rates of ICT access and use, particularly with regard to smartphone use. Additionally, we found a moderate use of ICT to access health and safety information and moderate interest in receipt of behavioral health information on alcohol or stress management via ICT. Younger age was associated with smartphone ownership, daily internet access, and weekly access of health and safety information. We found that employment was associated with smartphone

XSL•FO
RenderX

ownership, and higher educational attainment was associated with daily internet access via cell phones. Our findings are consistent with other studies demonstrating associations between technology use, age [5,21,25], and educational attainment [5,21,22,25]. Higher income was associated with daily email messaging and internet access via computer and inversely associated with weekly access of health and safety information by cell phone. Notably, smartphone ownership was not associated with income. Our results suggest that ICT health interventions would be accessible and may be of interest as a modality for receipt of behavioral health information in women recruited from urban community sites. These findings support consideration and exploration of the use of ICT, particularly smartphones, as a tool to educate women outside of traditional healthcare delivery settings.

With the evolution of cell technology, portable ICT devices have many of the same functionalities as computers, potentially decreasing the need for a computer. The ICT modality most frequently used by the women in our sample was cell phones, particularly smartphones. While 62.2% of women in our sample reported at least weekly computer use, 96.4% were using a cell phone at least weekly. The near universal report of cell phone ownership and use among our community-based sample is consistent with that of the general US population [1] and clinic-based samples of urban predominantly lower socioeconomic status patients [22,24-28]. Yet with respect to smartphone ownership, our findings differed from previous analyses. A larger proportion of women in our sample reported smartphone use (89%) compared to the most recent general population survey of US adults (68%) [1]. In our sample, cell phones were the principal ICT means for communicating which is consistent with other studies of persons with lower socioeconomic status [22,24,29]. These results are also consistent with national data indicating that racial and ethnic minorities, specifically black and Latino Americans, lead the way with respect to use of cell devices for accessing the internet, social media sites, health information, and tracking or managing health with specialized apps [20,30]. Reasons for these differences are likely multifactorial including decreased rates of home broadband access and tablet ownership for black Americans compared to white Americans, making cell phones the only device available for internet use [31,32], as well as racial differences in attitudes about information exchange via cell devices [33-35].

More than half of the women in the sample used ICT to access health and safety information at least weekly. About half of the participants in this study reported interest in receiving information about stress management and/or alcohol and health via ICT. These rates are somewhat lower than other studies among low socioeconomic samples conducted in clinical settings which focused on the receipt of general medical information rather than behavioral health information. For example, Mitchell and colleagues found that the majority of parents (84%) in pediatric clinics were willing to receive health information through ICT [22]. Studies specifically examining interest in ICT for behavioral health education or intervention have reported variable results. Torous and colleagues found that approximately half of the patients (49%) recruited from 4 psychiatric clinics across the US reported accessing general health information via smartphone in the previous 6 months, but approximately 71% reported interest in using a smartphone to monitor mental health symptoms [36]. In contrast, Sharpe and colleagues found that only 29% of persons living with HIV, who were recruited from 5clinical sites in Florida, reported interest in using a cell phone app to self-manage drinking, though there were significant differences based on respondents' drinking levels, with those with hazardous drinking being more likely to express interest [37]. Our results highlight that interest in and use of ICT for accessing health-related information extend beyond clinic-based samples and support the acceptability of ICT-based behavioral health educational applications, including those focused on stress management or alcohol use.

Limitations

These findings should be interpreted in the context of the study's limitations. The sample size was small and not nationally representative, which limits generalization of our results to men, nonblack racial groups, or nonurban populations. Analyses relied on self-reported data, which introduced potential response bias, as participants may be more likely to provide socially desirable answers, and we did not have procedures to confirm ICT use. This may have been mitigated by our use of ACASI. We did not inquire about use of tablet computers or downloadable apps for cell phones, which limited our ability to report on use or acceptability of these ICT formats for delivering health information and interventions.

Despite these limitations, this study extends the existing literature on ICT use. This study advances the literature by surveying a population that been underrepresented in previous surveys on ICT access or interests. The unique method of sampling from urban hair salons and CAPCs provides valuable new information about ICT use and the potential for health-related education among a community sample of predominantly black women with low socioeconomic status. Additionally, this study used a computer-assisted survey, which is a unique method among studies of populations with low socioeconomic statuses. Our results suggest that smartphone behavioral health education programs would be both accessible and of interest to black women in urban areas with low socioeconomic status.

Implications for Practice and Policy

There are several potential advantages of ICT-delivered health education and intervention applications. ICT interventions allow for increased intervention fidelity as unlike humans the software's intervention delivery will be standardized [7,38]. ICTs offer a more cost-effective approach to education and or intervention delivery as compared to employing a full-time health educator or interventionist [39-41]. ICTs allow for increased confidentiality and elimination of barriers to care associated with traditional interventions, such as discomfort discussing sensitive topics [6,38]. Finally, the widespread adoption of ICT use offers increased portability, reach and convenience, allowing patients to complete the education and/or intervention at times most suitable to them, as well as eliminating potential barriers of lack of transportation and lapses in health insurance [6,38,42,43]. Future research should explore

XSL•FO
RenderX

ICT app use, preferred design and content features as well as the development, piloting, and assessment of utility and cost-effectiveness of different ICT health-related modalities in this population.

Conclusions

ICT, particularly smartphones, were widely available and used in our sample of urban, predominantly black women with low socioeconomic status. These devices offer a promising vehicle to study the delivery of behavioral health education and intervention in this population. Given these findings, developers of ICT-based behavioral health programs should ensure cell platforms are as robustly developed and investigated as computer-based modalities.

Acknowledgments

The authors wish to thank Amanda Federline and Neha “Jia” Ahmad for their hard work in survey programming and administration. We thank Dr Betsy McCaul for her insightful feedback. We thank the directors and staff of the Baltimore City Community Action Partnership Centers as well as the owners and stylists of the participating salons and Salon Advisory Council: Dreadz N’ Headz, Miracle’s Hair Salon, Salon 1713, Salon Infiniti, Ultimate Performance Beauty Salon, and Tameka and Company Salon. Funding for this study was provided to AAHA from the National Institute on Alcohol Abuse and Alcoholism (K23: 4K23AA020316-05).

Conflicts of Interest

AAHA has served as a consultant for Emmi Solutions and Indivior, Inc. All other authors report no conflicts of interest.

References

1. Anderson M. Pew Research Center. 2015 Oct 29. Technology Device Ownership: 2015 URL: http://www.pewinternet.org/2015/10/29/technology-device-ownership-2015 [accessed 2017-11-09] [WebCite Cache ID 6woWExzH2]
2. Haughton LT, Kreuter M, Hall J, Holt CL, Wheetley E. Digital divide and stability of access in African American women visiting urban public health centers. J Health Care Poor Underserved 2005 May;16(2):362-374. [doi: 10.1353/hpu.2005.0034] [Medline: 15937398]
3. Laz TH, Berenson AB. Racial and ethnic disparities in internet use for seeking health information among young women. J Health Commun 2013;18(2):250-260 [FREE Full text] [doi: 10.1080/10810730.2012.707292] [Medline: 23130608]
4. Lorence DP, Park H, Fox S. Racial disparities in health information access: resilience of the Digital Divide. J Med Syst 2006 Aug;30(4):241-249. [Medline: 16978003]
5. Zickuhr K, Smith A. Pew Research Center. 2012 Apr 13. Digital Differences URL: http://www.pewinternet.org/2012/04/13/digital-differences [accessed 2018-07-23] [WebCite Cache ID 718GWLtrw]
6. Johnson K, Isham A, Shah DV, Gustafson DH. Potential roles for new communication technologies in treatment of addiction. Curr Psychiatry Rep 2011 Oct;13(5):390-397 [FREE Full text] [doi: 10.1007/s11920-011-0218-y] [Medline: 21739171]
7. Marsch LA. Leveraging technology to enhance addiction treatment and recovery. J Addict Dis 2012;31(3):313-318 [FREE Full text] [doi: 10.1080/10550887.2012.694606] [Medline: 22873192]
8. Des JDC, Paone D, Milliken J, Turner CF, Miller H, Gribble J, et al. Audio-computer interviewing to measure risk behaviour for HIV among injecting drug users: a quasi-randomised trial. Lancet 1999 May 15;353(9165):1657-1661. [Medline: 10335785]
9. Turner CF, Ku L, Rogers SM, Lindberg LD, Pleck JH, Sonenstein FL. Adolescent sexual behavior, drug use, and violence: increased reporting with computer survey technology. Science 1998 May 8;280(5365):867-873 [FREE Full text] [Medline: 9572724]
10. Wright D, Aquilino W, Supple A. A comparison of computer-assistedpaper-and-pencil self-administered questionnaires in a survey on smoking, alcohol,drug use. Public Opin Q 1998;62(3):331-353 [FREE Full text] [doi: 10.1086/297849]
11. Dulin PL, Gonzalez VM, Campbell K. Results of a pilot test of a self-administered smartphone-based treatment system for alcohol use disorders: usability and early outcomes. Subst Abus 2014;35(2):168-175 [FREE Full text] [doi: 10.1080/08897077.2013.821437] [Medline: 24821354]
12. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
13. Hester RK, Squires DD, Delaney HD. The Drinker's Check-up: 12-month outcomes of a controlled clinical trial of a stand-alone software program for problem drinkers. J Subst Abuse Treat 2005 Mar;28(2):159-169. [doi: 10.1016/j.jsat.2004.12.002] [Medline: 15780546]
14. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]

XSL•FO
RenderX

15. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436 [FREE Full text] [doi: 10.1002/cpp.1855] [Medline: 23918764]
16. Mental Health: Culture, Race, and Ethnicity: A Supplement to Mental Health, A Report of the Surgeon General. In: Substance Abuse and Mental Health Service Administration, Center for Mental Health Services. Rockville, MD: US Department of Health and Human Services; 2001:1-204.
17. Ronzio CR, Guagliardo MF, Persaud N. Disparity in location of urban mental service providers. Am J Orthopsychiatry 2006 Jan;76(1):37-43. [doi: 10.1037/0002-9432.76.1.37] [Medline: 16569125]
18. Guerrero EG, Kao D. Racial/ethnic minority and low-income hotspots and their geographic proximity to integrated care providers. Subst Abuse Treat Prev Policy 2013 Sep 23;8:34 [FREE Full text] [doi: 10.1186/1747-597X-8-34] [Medline: 24059252]
19. Nam E, Matejkowski J, Lee S. Racial/Ethnic Differences in Contemporaneous Use of Mental Health and Substance Use Treatment Among Individuals Experiencing Both Mental Illness and Substance Use Disorders. Psychiatr Q 2017 Mar;88(1):185-198. [doi: 10.1007/s11126-016-9444-0] [Medline: 27271529]
20. Gibbons MC, Fleisher L, Slamon RE, Bass S, Kandadai V, Beck JR. Exploring the potential of Web 2.0 to address health disparities. J Health Commun 2011;16 Suppl 1:77-89. [doi: 10.1080/10810730.2011.596916] [Medline: 21843097]
21. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
22. Mitchell SJ, Godoy L, Shabazz K, Horn IB. Internet and mobile technology use among urban African American parents: survey study of a clinical population. J Med Internet Res 2014 Jan 13;16(1):e9 [FREE Full text] [doi: 10.2196/jmir.2673] [Medline: 24418967]
23. Neuenschwander LM, Abbott A, Mobley AR. Assessment of low-income adults' access to technology: implications for nutrition education. J Nutr Educ Behav 2012;44(1):60-65. [doi: 10.1016/j.jneb.2011.01.004] [Medline: 21924959]
24. Shields W, McDonald E, Perry E, Aitken M, Gielen AC. Cell phone and computer use among parents visiting an urban pediatric emergency department. In: American Public Health Association. 2013 Presented at: American Public Health Association 141st Annual Meeting and Exposition; November 6, 2013; Boston, MA.
25. Dahne J, Lejuez CW. Smartphone and Mobile Application Utilization Prior to and Following Treatment Among Individuals Enrolled in Residential Substance Use Treatment. J. Subst. Abuse Treat 2015;58:95-99. [Medline: 26231698]
26. Demartini TL, Beck AF, Klein MD, Kahn RS. Access to digital technology among families coming to urban pediatric primary care clinics. Pediatrics 2013 Jul;132(1):e142-e148 [FREE Full text] [doi: 10.1542/peds.2013-0594] [Medline: 23753100]
27. McClure EA, Acquavita SP, Harding E, Stitzer ML. Utilization of communication technology by patients enrolled in substance abuse treatment. Drug Alcohol Depend 2013 Apr 01;129(1-2):145-150 [FREE Full text] [doi: 10.1016/j.drugalcdep.2012.10.003] [Medline: 23107600]
28. Ryan MH, Yoder J, Flores SK, Soh J, Vanderbilt AA. Using Health Information Technology to Reach Patients in Underserved Communities: A Pilot Study to Help Close the Gap With Health Disparities. Glob J Health Sci 2015 Oct 20;8(6):86-94 [FREE Full text] [doi: 10.5539/gjhs.v8n6p86] [Medline: 26755484]
29. Bender MS, Choi J, Arai S, Paul SM, Gonzalez P, Fukuoka Y. Digital technology ownership, usage, and factors predicting downloading health apps among caucasian, filipino, korean, and latino americans: the digital link to health survey. JMIR Mhealth Uhealth 2014;2(4):e43 [FREE Full text] [doi: 10.2196/mhealth.3710] [Medline: 25339246]
30. Smith A. Smartphone Adoption and Usage. 2011 URL: http://www.pewinternet.org/2011/07/11/smartphone-adoption-and-usage/ [accessed 2018-01-28] [WebCite Cache ID 6woWlyRoj]
31. Smith A. Mobile Access. 2010. URL: http://www.pewinternet.org/files/old-media/Files/Reports/2010/PIP_Mobile_Access_2010.pdf [accessed 2018-01-28] [WebCite Cache ID 6woX9XkpN]
32. Smith A. Pew Research Center. 2014 Jan 06. African Americans and Technology Use URL: http://www.pewinternet.org/2014/01/06/african-americans-and-technology-use/ [WebCite Cache ID 6woXDc3d5]
33. Anderson M. Racial and ethnic differences in how people use mobile technology. 2015 URL: http://www.pewresearch.org/fact-tank/2015/04/30/racial-and-ethnic-differences-in-how-people-use-mobile-technology/ [accessed 2018-01-28] [WebCite Cache ID 6woXXD8Pq]
34. McGillicuddy JW, Weiland AK, Frenzel RM, Mueller M, Brunner-Jackson BM, Taber DJ, et al. Patient attitudes toward mobile phone-based health monitoring: questionnaire study among kidney transplant recipients. J Med Internet Res 2013;15(1):e6 [FREE Full text] [doi: 10.2196/jmir.2284] [Medline: 23305649]
35. Smith A. Government Online. 2010 URL: http://www.pewinternet.org/2010/04/27/government-online/ [accessed 2018-01-28] [WebCite Cache ID 6woXv8Gbx]
36. Torous J, Chan SR, Yee-Marie TS, Behrens J, Mathew I, Conrad EJ, et al. Patient Smartphone Ownership and Interest in Mobile Apps to Monitor Symptoms of Mental Health Conditions: A Survey in Four Geographically Distinct Psychiatric Clinics. JMIR Ment Health 2014;1(1):e5 [FREE Full text] [doi: 10.2196/mental.4004] [Medline: 26543905]

XSL•FO
RenderX

37. Sharpe JD, Zhou Z, Escobar-Viera CG, Morano JP, Lucero RJ, Ibañez GE, et al. Interest in using mobile technology to help self-manage alcohol use among persons living with the human immunodeficiency virus: A Florida Cohort cross-sectional study. Subst Abus 2017 Jul 19:1-6. [doi: 10.1080/08897077.2017.1356793] [Medline: 28723300]
38. Schneider J, Makelarski JA, Van HM, Lipton RB, Abramsohn E, Lauderdale DS, et al. Differential access to digital communication technology: association with health and health survey recruitment within an African-American underserviced urban population. J Urban Health 2011 Jun;88(3):479-492 [FREE Full text] [doi: 10.1007/s11524-010-9533-6] [Medline: 21279451]
39. Donker T, Blankers M, Hedman E, Ljótsson B, Petrie K, Christensen H. Economic evaluations of Internet interventions for mental health: a systematic review. Psychol Med 2015 Dec;45(16):3357-3376. [doi: 10.1017/S0033291715001427] [Medline: 26235445]
40. Murphy SM, Campbell ANC, Ghitza UE, Kyle TL, Bailey GL, Nunes EV, et al. Cost-effectiveness of an internet-delivered treatment for substance abuse: Data from a multisite randomized controlled trial. Drug Alcohol Depend 2016 Apr 01;161:119-126 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.01.021] [Medline: 26880594]
41. Olmstead TA, Ostrow CD, Carroll KM. Cost-effectiveness of computer-assisted training in cognitive-behavioral therapy as an adjunct to standard care for addiction. Drug Alcohol Depend 2010 Aug 1;110(3):200-207 [FREE Full text] [doi: 10.1016/j.drugalcdep.2010.02.022] [Medline: 20392575]
42. Shi L, Chen C, Nie X, Zhu J, Hu R. Racial and socioeconomic disparities in access to primary care among people with chronic conditions. J Am Board Fam Med 2014;27(2):189-198 [FREE Full text] [doi: 10.3122/jabfm.2014.02.130246] [Medline: 24610181]
43. Mulia N, Schmidt LA, Ye Y, Greenfield TK. Preventing disparities in alcohol screening and brief intervention: the need to move beyond primary care. Alcohol Clin Exp Res 2011 Sep;35(9):1557-1560 [FREE Full text] [doi: 10.1111/j.1530-0277.2011.01501.x] [Medline: 21599711]

Abbreviations

ACASI: audio-computer assisted self-interview
CAPC: Community Action Partnership Center
GED: General Equivalency Diploma
ICT: information and communication technology
OR: odds ratio

Edited by G Eysenbach; submitted 01.02.18; peer-reviewed by S Moore, K Aguirre; comments to author 09.04.18; revised version received 25.05.18; accepted 18.06.18; published 14.08.18.

Please cite as:
Jabour SM, Page A, Hall SF, Rodriguez L, Shields WC, Alvanzo AAH
Information and Communication Technologies Interest, Access, and Use: Cross-Sectional Survey of a Community Sample of Urban, Predominantly Black Women
J Med Internet Res 2018;20(8):e248
URL: http://www.jmir.org/2018/8/e248/
doi:10.2196/jmir.9962
PMID:30108036

©Sarah M Jabour, Alexis Page, Seventy F Hall, Lycinda Rodriguez, Wendy C Shields, Anika A H Alvanzo. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 14.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Predicting Therapy Success and Costs for Personalized Treatment Recommendations Using Baseline Characteristics: Data-Driven Analysis

Vincent Bremer[1*], MA; Dennis Becker[1*], MSc; Spyros Kolovos[2,3*], PhD; Burkhardt Funk[1*], PhD; Ward van Breda[4*], MSc; Mark Hoogendoorn[4*], PhD; Heleen Riper[2*], PhD

[1]Institute of Information Systems, Leuphana University, Lüneburg, Germany

[2]Department of Clinical, Neuro- & Developmental Psychology, Vrije University, Amsterdam, Netherlands

[3]Department of Health Sciences, Vrije University, Amsterdam, Netherlands

[4]Department of Computer Science, Vrije University, Amsterdam, Netherlands

[*]all authors contributed equally

Corresponding Author:
Vincent Bremer, MA
Institute of Information Systems
Leuphana University
C4.320
Universitätsallee 1
Lüneburg, 21335
Germany
Phone: 49 41316771157
Fax: 49 41316771749
Email: vincent.bremer@leuphana.de

Abstract

Background: Different treatment alternatives exist for psychological disorders. Both clinical and cost effectiveness of treatment are crucial aspects for policy makers, therapists, and patients and thus play major roles for healthcare decision-making. At the start of an intervention, it is often not clear which specific individuals benefit most from a particular intervention alternative or how costs will be distributed on an individual patient level.

Objective: This study aimed at predicting the individual outcome and costs for patients before the start of an internet-based intervention. Based on these predictions, individualized treatment recommendations can be provided. Thus, we expand the discussion of personalized treatment recommendation.

Methods: Outcomes and costs were predicted based on baseline data of 350 patients from a two-arm randomized controlled trial that compared treatment as usual and blended therapy for depressive disorders. For this purpose, we evaluated various machine learning techniques, compared the predictive accuracy of these techniques, and revealed features that contributed most to the prediction performance. We then combined these predictions and utilized an incremental cost-effectiveness ratio in order to derive individual treatment recommendations before the start of treatment.

Results: Predicting clinical outcomes and costs is a challenging task that comes with high uncertainty when only utilizing baseline information. However, we were able to generate predictions that were more accurate than a predefined reference measure in the shape of mean outcome and cost values. Questionnaires that include anxiety or depression items and questions regarding the mobility of individuals and their energy levels contributed to the prediction performance. We then described how patients can be individually allocated to the most appropriate treatment type. For an incremental cost-effectiveness threshold of 25,000 €/quality-adjusted life year, we demonstrated that our recommendations would have led to slightly worse outcomes (1.98%), but with decreased cost (5.42%).

Conclusions: Our results indicate that it was feasible to provide personalized treatment recommendations at baseline and thus allocate patients to the most beneficial treatment type. This could potentially lead to improved decision-making, better outcomes for individuals, and reduced health care costs.

XSL•FO
RenderX

(J Med Internet Res 2018;20(8):e10275) doi:10.2196/10275

KEYWORDS

treatment recommendation; cost effectiveness; mental health; machine learning

Introduction

In a clinical context, different forms of behavioral interventions such as face-to-face or internet-based treatments exist for patients with depressive disorders. Clinical and cost effectiveness studies provide important knowledge regarding these treatment alternatives [1]. However, questions remain as to which particular individuals prefer particular treatment types or receive an increased benefit from one specific treatment option over another, especially before the treatment begins. Therapists or other clinicians often make decisions based on personal understanding and experience, leading to high uncertainty or nonoptimal decisions [1]. This uncertainty can potentially result in worse treatment outcomes for individuals and increased health care costs. Simultaneously, policy makers and stakeholders increasingly demand cost-effectiveness evidence in order to support their conclusions and decisions [2].

For supporting these admittedly difficult and complex decisions, approaches exist based on cost analysis or decision analysis [1,3]. The incremental cost-effectiveness ratio (ICER) is a widespread indicator for cost effectiveness [4]. The goal is to support the mentioned decisions by identifying actions that, on average, maximize a specific result [1] such as quality-adjusted life years (QALYs). The ICER is applied on a population level, which means that average values of costs and outcomes are considered for population-level decisions [1,5]. This procedure does not consider any heterogeneity among individuals regarding outcomes and costs. Individual patients, for example, respond differently to treatment and have varying mindsets regarding risks [6,7]. Thus, the average outcomes and costs often do not necessarily represent the best decision for an individual [6]. Even though these aspects are well known, cost-effectiveness analyses based on average values are still widely used [6].

Predictive analyses can provide crucial insight into aspects that influence outcomes and costs of interventions and can be beneficial for patients as well as society [8]. Research that seeks to forecast outcomes for patients with depression already exists. One study, for example, predicted treatment success in the domain of depression and showed that baseline data has predictive power in this context [9]. Another study predicted treatment outcomes of treatment-resistant patients with depression and thereby revealed important predictors such as severity and suicidal risk, among others [10]. These types of statistical procedures can ultimately result in the development of decision support systems in the context of health interventions. In the field of depression treatment, these systems often lead to positive effects and even a reduction of symptoms in various situations [11].

This study focused on making personalized treatment recommendations. For this purpose, we predicted the outcomes and costs for different treatment types, at baseline, on an individual patient level. We applied various machine learning techniques, evaluated them based on their predictive performance, and revealed important features that contributed to the prediction. In order to derive personalized treatment recommendations, we applied an individualized cost-effectiveness analysis based on the ICER. Unlike its traditional utilization based on the ratio of average values, we used individual predictions for each treatment type and its alternative. The predictions and their generated information can provide additional knowledge and enable practitioners, as well as researchers, to individually assign patients at baseline to their most appropriate treatment type in terms of outcomes and costs. This approach is applied to data from an internet-based two-arm randomized controlled trial in the domain of depression.

The forecast of individual outcomes and costs is one of the most important aims in clinical research [12], and personalized analyses and illustrations of cost effectiveness in this context are of increased interest and need [6,13]. Thus, we contribute to existing research by attempting to predict these factors at the start of treatment for each individual and by further proposing a conceptual approach for treatment recommendations, as applied to empirical data.

Methods

Data and Preprocessing

The data we utilized originate from the European Union-funded project E-Compared in which the clinical and cost effectiveness of blended treatment (BT) for depression, where internet-based and face-to-face treatments are combined in one integrated treatment protocol, is evaluated and compared with treatment as usual (TAU) in 9 different countries [14]. Participants were aged 18 years or older, met criteria for a major depressive disorder, were not of high suicidal risk, were not being treated for depression, and had access to an internet connection. Table 1 illustrates the different questionnaires used in the study.

The data consisted of individualized information regarding depressive symptoms, medical costs, and other factors. These questionnaires are widely utilized and known and can be found elsewhere [14-18]. The data in the E-Compared project were collected multiple times during the trial: at baseline, 3 months, 6 months, and 12 months. Questionnaires 3, 4, 6, and 7 (according to Table 1) were also available, not only at baseline but also after other times during data acquisition. Because we were interested in recommendations before the start of the actual treatment, we solely used the baseline information as features in this study.

XSL•FO
RenderX

Table 1. Data utilized in this study.

Data	Description
Demographic data	N/A[a]
Current treatment	Current treatment type, medication, provider
MINI International Neuropsychiatric Interview	Structured clinical interview for making diagnoses
Quick Inventory of Depressive Symptomatology (16-Item) (Self-Report)	Quick Inventory of Depressive Symptomatology
Patient Health Questionnaire-9	Questions regarding depressive symptoms
5-level EQ-5D	EuroQol questionnaire; measuring generic health status; for calculation of quality-adjusted life years
Costs Associated with Psychiatric Illness	Measurement of healthcare costs and productivity losses
Treatment preferences	Individual preferences for blended treatment or treatment as usual

[a]N/A: Not applicable.

We used QALY as an outcome, as measured by the EuroQol questionnaire (5-level EQ-5D version). Utility weights were calculated using the Dutch tariffs [19]. These weights are a preference-based measure of quality of life anchored at 0 (worst perceivable health) and 1 (perfect health). QALYs were calculated by multiplying the utility weights with the amount of time a participant spent in a particular health state. Transitions between the health states were linearly interpolated. The costs that we aimed to forecast were measured from the societal perspective (including healthcare utilization and productivity losses) based on the adapted version of the Trimbos and Institute for Medical Technology Assessment questionnaires on Costs Associated with Psychiatric Illness [18]. Dutch unit costs were used to value healthcare utilization and productivity losses [20]. Costs for the online part of BT included maintenance and hosting of the treatment and costs that occurred for a therapist to provide feedback to participants. We decided to use costs from a societal perspective because they represent interests of society and all other stakeholder groups [1]. More information on the calculation of the costs can be found elsewhere [21]. As dependent variables, we utilized QALY and costs that appear after a 6-month period. This allowed for more observations compared with the data at 12 months (350 patients vs 212 patients) because not all patients had already finished the treatment process. Because we focused on the outcome data up to 6 months, QALY could have a maximum value of 0.5 in our analysis.

During the data preprocessing phase, we merged all mentioned data from Table 1. This process led to 309 features that could be utilized for the prediction. We then calculated the costs and QALY for each individual. We only included patients for which both dependent variables were not missing. By splitting the dataset into groups for the different treatment types (TAU and BT), some factor levels of an item or feature can go missing. We removed 97 features that had just one level or were missing. Multimedia Appendix 1 lists the omitted items from the questionnaires. The resulting dataset still contained 29,568 missing values. Disregarding these values, and thus deleting them, would lead to a substantial decrease in observations. We therefore utilized two different methods for handling them in order to evaluate which method would perform better regarding the predictive performance. We first imputed the numeric values by sampling from a normal distribution based on the mean value and SD of the corresponding feature. We imputed the categorical predictors by sampling from the categorical distribution of those features. As a second approach, we imputed the missing values by the median (numeric variable) and mode (categorical variable). Finally, we ended up with a dataset of 350 observations (1 for each patient) and 212 features. In the following, we have reported only the results for the latter imputation procedure. In Multimedia Appendix 2, we have also demonstrated the final performances for the first imputation method. However, we decided to utilize the latter method because it led to the best performance in terms of prediction.

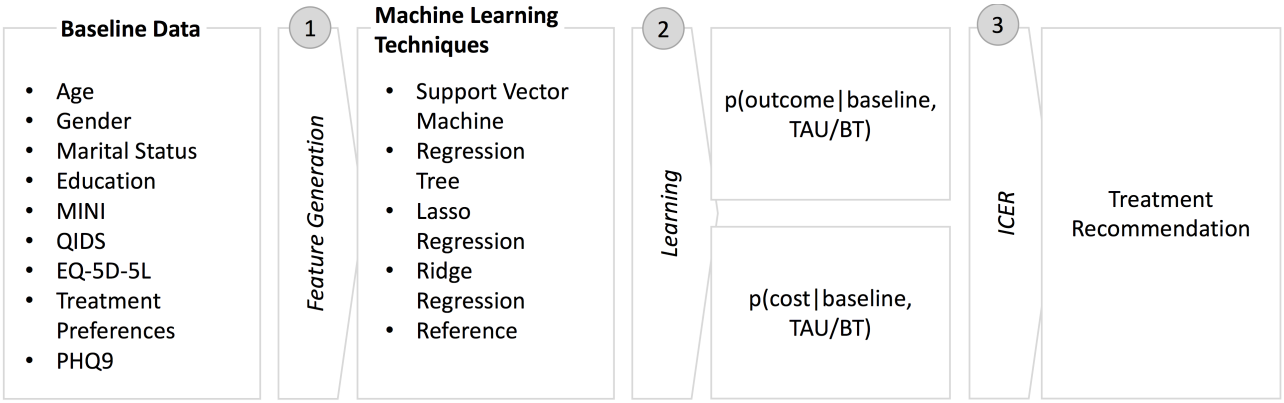
Approach & Statistical Analysis

In order to derive individual treatment recommendations, we utilized the baseline features as input for predicting individual level outcome and costs based on the treatment type, as seen in Figure 1. We applied various machine learning techniques to evaluate which yielded the highest prediction performance. As mentioned by several studies, it is beneficial to compare different statistical procedures in order to eventually find the most precise model, especially when predicting costs due to the challenging nature of this activity [8,22,23]. Because the data consist of numerous features, we applied a feature selection method to reveal variables that contributed to the prediction performance. To demonstrate how the forecasts can be beneficial in recommending treatment types on an individual patient level, we applied the ICER to the predictions.

Specifically, we estimated the conditional probability $p(o,c \mid b,tt)$ for each treatment type, where *o* is the outcome, *c* is the costs, *b* reflects the baseline features, and *tt* is one of the 2 treatment types. Given the limited amount of data, we assumed that the conditional probability could be factorized as follows: $p(o,c \mid b,tt)=p(o \mid b,tt)p(c \mid b,tt)$.

XSL•FO
RenderX

Figure 1. Process for deriving treatment recommendations for individuals. BT: blended treatment; ICER: incremental cost-effectiveness ratio; TAU: treatment as usual.



For the prediction of outcome and costs, we used linear regression and support vector regression (SVR). The latter method has shown good predictive capabilities in various fields [24]. We further utilized regression trees and ridge regression. For finding the optimal parameters, we applied a grid-based search and cross-validation. Additionally, we defined the mean of all outcomes or costs as a reference measure. If unable to achieve a better prediction performance compared with the reference measure, it is questionable if the application of more advanced statistical methods is appropriate in this context. For finding the model that achieves the highest prediction performance, we used leave-one-out cross-validation. That is, one observation is utilized as the test set and the remaining observations are used for training the model. This procedure is repeated for every single observation in the dataset. The error measures we used were root mean square error (RMSE) and mean absolute error (MAE). We have presented both error measures because debate exists as to which measure is more appropriate for the demonstration of predictive performance [25,26].

When utilizing a vast number of features, overfitting presumably occurs. Thus, we used Lasso regression to select features that contributed to the predictive performance. Lasso is a linear regression that introduces a penalty term called regularizer [27]. The error function of the regression, which is to be optimized, consists of the mean square error of the misclassified samples and a term that penalizes the absolute value of the sum of regression coefficients. This linear penalty enforces useless coefficients to shrink toward zero in order to produce a sparse solution. The corresponding optimization problem is illustrated below, where X is the baseline feature, Y is the outcome or costs, and β is the coefficient:

The parameter λ influences the strength of the penalty. Specifically, the higher the value of λ, the higher the penalty. A higher penalty leads to sparser solutions (more coefficients are shrunk to zero). The optimal λ's are found by utilizing cross-validation. After obtaining the specific features that appear to add to the predictive accuracy, we again predicted the outcome values and costs based on the aforementioned machine learning techniques. This time, however, we only utilized the features that were identified by the Lasso regression. Finally, we selected the algorithm that produced the smallest error and therefore performed best for the outcome and cost predictions. Based on these individual predictions, we calculated the ICER, as seen in the equation:

The ICER was then visualized in the cost-effectiveness plane [28]. By predicting the costs and outcomes at baseline and utilizing the ICER, we could then make recommendations about individual patient allocation. We implemented the mentioned models and processes in R (R Core Team; Vienna, Austria) [29].

Results

Overall Findings

Before we focused on the outcome and cost predictions, we illustrated the general improvements of the patients for TAU and BT. The E-Compared project hypothesized noninferiority between both treatment types (ie, BT is not less effective) [14]. Improvement was defined as the difference of the start and end value of the cumulated PHQ9 values. The PHQ9 questionnaire is a reliable measure for depression severity [16]. Because we only investigated the improvements for a 6-month period, these results are not final; however, they can indicate a trend. Table 2 shows that the mean baseline score for PHQ9 was 15.35 for BT and 15.42 for TAU. At the 6-month measurement, the scores were 7.85 and 9.49, respectively. Furthermore, 154 patients in the BT group and 140 patients in the TAU group showed improvement. Therefore, we can see that the PHQ9 value decreased more strongly for BT and that the number of improvements for BT exceeded the outcome of TAU. Applying a *t* test for the comparison of the mean end values resulted in the rejection of the hypothesis that both samples had the same mean (P=.006).

XSL•FO
RenderX

Table 2. Mean of Patient Health Questionnaire-9 scores at baseline and end for treatment as usual and blended treatment as well as the numbers of patients in each condition that improved (N=350).

Measures	Treatment as usual	Blended treatment
Start Patient Health Questionnaire-9, mean	15.42	15.35
End Patient Health Questionnaire-9, mean	9.49	7.85
Patients with improvement, n	140	154
Patients without improvement, n	38	18

Table 3. Results for prediction performance based on all baseline features for varying machine learning approaches.

Model	Outcome		Costs in €	
	MAE_O[a]	$RMSE_O$[b]	MAE_C[c]	$RMSE_C$[d]
Support vector regression	0.0714	0.0997	6299.63	9360.50
Regression tree	0.0698	0.0992	6573.94	9406.11
Ridge regression	0.0711	0.1000	6557.69	9187.78
Reference measure	0.0770	0.1017	7024.11	9539.54

[a]MAE_O: mean absolute error in outcome.

[b]$RMSE_O$: root mean square error in outcome.

[c]MAE_C: mean absolute error in cost.

[d]$RMSE_C$: root mean square error in cost.

Outcome and Cost Prediction

Table 3 illustrates the prediction performance for all utilized machine learning techniques and all baseline features. Overall, the SVR and regression tree had the smallest errors for performance measures. The ridge regression also performed better than the reference measure. Based on a Wilcoxon test, MAEs differed significantly (SVR: P_O=.030, P_C<.001; Tree: P_O=.001, P_C<.001; Ridge: P_O=.049, P_C<.023). Since we had more features than observations, we did not apply ordinary least squares regression when utilizing all baseline features.

We then performed Lasso regression in order to select the important features that contributed to the prediction performance. The tables in Multimedia Appendix 3 show the important features that were utilized and their corresponding coefficient. By applying cross-validation, we chose specific λ values that minimized the mean cross-validated error. For TAU and BT, we used all features up to a λ value of 0.01485 and 0.01479, respectively (433.83 and 651.14 for the cost prediction).

Multiple features appeared repeatedly. Various questions regarding the medication use and the amount of consultations of some kind of therapist, practitioner, or treatment program occurred most often (24 and 16 times, respectively). Furthermore, the anxiety or depression items (6 times), mobility (5 times), origin of the patient (7 times), and energy level questions (4 times) appeared to have an influence on the prediction performance. Using the selected features, we then repeatedly applied the above specified statistical methods in order to achieve a better accuracy.

We observed a general increase in performance (Table 4). All statistical methods performed better than the reference measure (except for RMSE for linear regression and cost prediction), which was again confirmed by a significant Wilcoxon test for MAEs (SVR: P_O<.001, P_C<.001; Regression: P_O<.001, P_C<.001; Tree: P_O=.002, P_C<.001; Ridge: P_O<.001, P_C<.001). This suggested that feature selection resulted in more accurate predictions in this context. The overall results demonstrate *that some machine learning approaches are beneficial when predicting the outcomes and costs*. Since ridge regression predicted the outcome and costs best, we utilized this model in the following analysis.

Figure 2 illustrates the predicted and observed values for each treatment type and dependent variable (QALY/costs). For estimating the ridge regression penalty term, we implemented 100 cross-validation runs and utilized the parameter that minimized the mean cross-validated error among these runs. The predictions were sorted in an ascending order. The blue markers or lines are the predictions and the black markers are the observed values where the y-axis demonstrates the value of the QALY/costs and the x-axis represents the corresponding patient. We observed that the predicted outcome and costs showed high uncertainty. The broader range of the actual observations around the blue markers for the cost predictions indicated that these were more difficult to achieve than outcome predictions in this context. Visually, however, the trend of the predictions appeared to be as expected, and as illustrated by the increased performance compared with the reference measure; this result indicates a step in the right direction.

XSL•FO
RenderX

Table 4. Results for prediction performance based on selected baseline features for varying machine learning approaches.

Model	Outcome		Costs in €	
	MAE_O[a]	$RMSE_O$[b]	MAE_C[c]	$RMSE_C$[d]
Support vector regression	0.0575	0.0812	5164.22	8026.46
Regression	0.0590	0.0793	6436.63	15319.89
Regression tree	0.0684	0.0952	6573.94	9406.11
Ridge regression	0.0553	0.0747	4590.00	6607.31
Reference measure	0.0770	0.1017	7024.11	9539.54

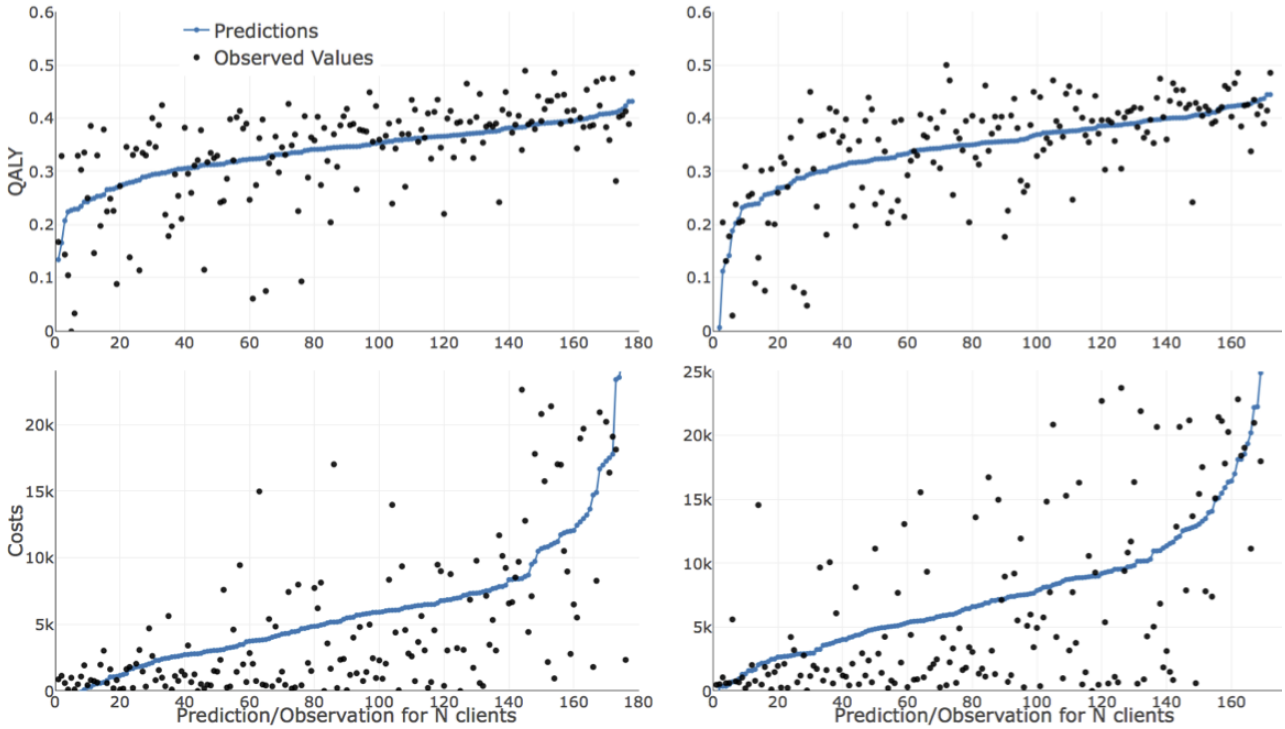
[a]MAE_O: mean absolute error in outcome.

[b]$RMSE_O$: root mean square error in outcome.

[c]MAE_C: mean absolute error in cost.

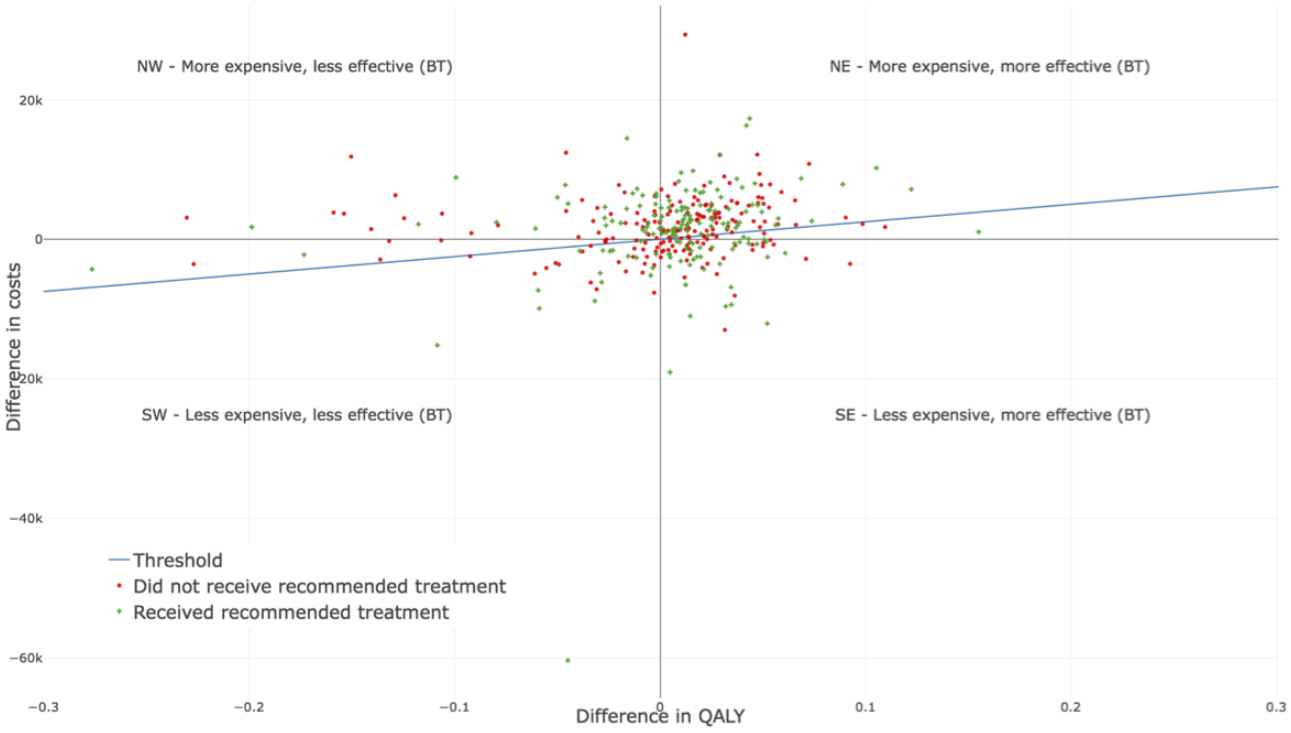
[d]$RMSE_C$: root mean square error in cost.

Figure 2. Predicted and observed values for quality-adjusted life years and costs and both treatment types (left panels for treatment as usual and right panels for blended treatment).



XSL•FO
RenderX

Figure 3. Expected improvement for all patients in relation to costs. The x-axis illustrates the difference in quality-adjusted life years (blended treatment-treatment as usual) and the y-axis the difference in costs (blended treatment- treatment as usual).



Treatment Recommendation

In order to derive individual treatment recommendations, we represent the differential outcomes and costs in the cost-effectiveness plane, where the y-axis is the difference between the costs of each treatment type and the x-axis is the difference between the clinical effects, as seen in Figure 3 [28]. Each quadrant has a different meaning. In our context, the NE quadrant represents higher costs and positive effects for BT; the SE quadrant indicates that BT is less expensive and more effective (BT dominates); the SW quadrant demonstrates the case where BT is less expensive but less effective; and the NW quadrant displays the situation where BT is more expensive and less effective (TAU dominates) [30]. As a first step, a threshold had to be defined that specified up to which point an additional improvement was worth the costs. In the context of this study, the monetary amount or willingness to pay for gaining one QALY differed by country [30]; the commonly used UK WTP thresholds for QALYs are between 25,000 and 35,000 €/QALY [31]. For this study, we used the conservative estimation of 25,000 €/QALY. A value above this threshold indicated that the treatment type was too expensive. Each patient represented by a green cross received the treatment type we would have recommended based on the prediction.

On the contrary, each patient that had a red circle should have received the other treatment type based on the forecasts. Questionnaire items that deviate tremendously for either TAU or BT create high differences when calculating the ICER. The point for the participant at the bottom of Figure 3 at (−0.04, −60.420), for example, is due to the fact that this patient reported a large number of hospital admissions. Since these are very expensive, it led to very high costs for this particular patient, and thus, the difference in costs between BT and TAU was high. Following this process, it is possible to recommend the likely most beneficial treatment type, on an individual level, at baseline.

Table 5 is a contingency table consisting of the patients for whom we recommended a specific treatment type. Only 46.57% (163/350) of all patients were treated using the treatment type we would recommend based on our models and the particular ICER threshold.

We then calculated potential outcomes and costs on a population level assuming the patients would have been allocated according to the predictions. For patients who had already received the recommended treatment type, we utilized the observed outcomes and costs. For patients for whom the actual treatment type was not recommended, we utilized the predictions of the model. Then, QALYs would have decreased by 1.98%, while at the same time, a reduction in costs of 5.42% could have been achieved.

XSL•FO
RenderX

Table 5. Treatment recommendation for all patients (N=350).

Treatment type	Recommended blended treatment, n (%)	Recommended treatment as usual, n (%)
Received blended treatment	70 (20)	102 (29.14)
Received treatment as usual	85 (24.29)	93 (26.57)

Discussion

Principal Findings

Given the growth in demand for personalized treatments and the need for a reduction in costs, predictions of outcomes and costs, in the context of mental health, are increasingly important [3]. In this study, we proposed an approach for personalized treatment recommendations at baseline. Here, individuals are assigned to the most beneficial treatment *before treatment*, which can, if desired, even be automated. We derived these recommendations by predicting patient individual QALYs and costs based on data from a European Union-funded project. We then used the ICER and the cost-effectiveness plane as an individualized treatment recommendation tool. Nowadays, decisions are often made based on the ICER; we proposed a feasible path that allows the individualization and tailoring of this process.

We illustrated that the utilization of all baseline features is not necessarily appropriate in this context. Taking advantage of feature selection techniques can increase prediction performance. As a result, we found that consultations with some kind of therapist, medication usage, anxiety or depression information (severity), mobility items (ie, "I have no problems in walking about"), and origin of the patient play an important role when predicting outcomes and costs in the context of digital health interventions. Therefore, including questionnaires that contain these factors and subsequently utilizing these features in statistical analyses when predicting outcomes and costs can be beneficial. We further illustrated that experimentation with different statistical methods benefits the final results since considerable varying performances occurred among the methods.

However, we demonstrated that prediction is a challenging task. Even though the results suggest that predictive power exists in the baseline features, our analyses indicated that the predictions, and thus the recommendations, come with uncertainty when only baseline information is available. In general, the predictive uncertainty is due to two sources. The first source is the uncertainty in the estimated parameters. With an increased amount of data, the uncertainty in parameter estimation reduces. This does not mean that we would achieve perfect predictions because the second source is related to the variance of treatments that cannot be explained by the model. More specifically, the models do not fully represent the reality and all its complexity. Hence, although the estimation of the model parameters improves with more data, the uncertainty that results from the model specifications and inability of the baseline information to precisely predict results remains. Nevertheless, we showed that we were able to predict the outcomes and costs better, compared with using the mean of the dependent variables as prediction (reference measure). Therefore, we are convinced that the baseline features do include some information regarding the forecast of outcomes and costs and can support practitioners in their decision-making process. Thus, combining these results with the ICER enabled us to provide treatment recommendations on an individual level.

As mentioned earlier, if the patients would have been allocated according to our predictions, QALYs would have decreased by 1.98% and a simultaneous reduction in costs of 5.42% could have been achieved. These results are based on a specific ICER threshold. When applying this procedure in a real-world setting, this threshold can be adjusted to values set by experts or policy makers or available budgets. These experts must make decisions regarding the monetary resources they would want to spend on a specific QALY gain. Thus, the outcome and costs can be controlled by setting this threshold. As suggested by a previous study [32], the cost-effectiveness decision rule might be modeled in a nonlinear form. For example, the value of improvements may vary among the outcome levels. Particularly, a difference between 0.1 and 0.2 on the scale might be more important than a difference between 0.8 and 0.9, even though the absolute difference is the same. The absolute severity of the symptoms can also play an additional role in this context. It might not be justifiable to spend additional monetary effort if a specific patient already does not suffer from severe symptoms. Therefore, experts in the field need to choose appropriate values for the ICER threshold based on their experiences and knowledge and even consider a nonlinear specification.

Even though these results are preliminary, the implementation of such predictive models in clinical decision support systems for usage in interventions can be beneficial. We envision developing a system that incorporates these models and provides treatment recommendations for individuals. However, investment into other aspects is necessary for the realization of such support systems. Besides the technical implementation, the creation of information systems in this context also requires interdisciplinary collaboration among clinicians, computer scientists, and other decision makers [33]. Future users, for example decision makers or therapists, need to be educated appropriately and also be involved in the design phase of the system and its requirements and development, while at the same time, the IT specialists need to be confronted with content-related issues of the user [34,35]. Thus, implementation should be carefully planned and considered as organizational development [36]. Furthermore, a vast amount of financial and organizational resources can be required for the implementation [33], and clinical decision makers need to understand the value and limitations of such decision support systems. Additionally, we need to be cautious with the interpretability of the results because in individual cases, recommendations might lead to suboptimal outcomes and high uncertainty depending on the particular context. Overall, these systems may be used in the future to support the decision-making process of clinicians and therapists and not to replace their treatment recommendations.

XSL•FO
RenderX

Limitations

This study has certain limitations. One limitation is the fact that we utilized data after a 6-month period. Usually, the preferred outcome for cost-effectiveness analysis is based on 12 months. Another limitation, which is closely associated with the previous aspect, is the size of the dataset we used. Given the complexity of the problem, it is inevitable that variations in performance occur when predicting other datasets. Thus, for achieving higher accuracy in predictions, obtaining more data is crucial. Even though our results are promising, more data and evaluations are needed in order to investigate the generalizability of these outcomes and improve the predictive accuracy of statistical techniques. Besides the size of the dataset, the data are heterogeneous in different ways. For example, the data were collected from 9 different European countries, with each having their own country-specific conditions [14]. This can result in country-specific patterns in the data. Given the limited amount of observations on a national level, we have not explored this multi-level structure. Additionally, the dataset consists of a large amount of missing values that needed imputation. Making all baseline questions mandatory for the patients can lead to an increased performance of the statistical procedures and can therefore lower uncertainty.

Conclusions

This study investigated how patients can be allocated to different treatment types in order to increase clinical and cost effectiveness. We demonstrated how to predict outcomes and costs in this context and proposed an approach for individualized treatment recommendations by utilizing the ICER. Simultaneously, we evaluated a variety of machine learning techniques and demonstrated specific features that contribute to the prediction performance. The results are indicative of progress. We hope that policy makers increasingly understand the benefit of predictive modeling in this context and apply these types of models to make better and simultaneously more personalized treatment choices. We further hope that we can contribute to the decision-making process in this field by providing a path that allows the prediction of eventual outcomes and costs on an individual basis before the onset of treatment.

Acknowledgments

This study has been conducted in the context of the European Union FP7 project E-COMPARED (project number 603098) and is based on the collected dataset in February 2017. We therefore thank the European Union for funding and the E-COMPARED consortium for the fantastic cooperation.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Omitted items from analysis.

[PDF File (Adobe PDF File), 40KB - jmir_v20i8e10275_app1.pdf]

Multimedia Appendix 2

Results for prediction performance based on sampling from normal and categorical distribution for varying machine learning approaches.

[PDF File (Adobe PDF File), 36KB - jmir_v20i8e10275_app2.pdf]

Multimedia Appendix 3

Important baseline features based on Lasso regression for quality-adjusted life years and cost prediction for treatment as usual and blended treatment.

[PDF File (Adobe PDF File), 93KB - jmir_v20i8e10275_app3.pdf]

References

1. Ryder HF, McDonough C, Tosteson ANA, Lurie JD. Decision Analysis and Cost-effectiveness Analysis. Semin Spine Surg 2009 Dec;21(4):216-222 [FREE Full text] [doi: 10.1053/j.semss.2009.08.003] [Medline: 23966758]
2. Knapp M. Economic evaluation and mental health: sparse past. fertile future? J Ment Health Policy Econ 1999 Dec 01;2(4):163-167. [Medline: 11967427]
3. van Breda W, Hoogendoorn M, Eiben AE, Andersson G, Riper H, Ruwaard J, et al. A feature representation learning method for temporal datasets. 2016 Presented at: IEEE Symposium Series on Computational Intelligence; 2016; Athens, Greece p. 1-8. [doi: 10.1109/SSCI.2016.7849890]
4. Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC. The role of cost-effectiveness analysis in health and medicine. Panel on Cost-Effectiveness in Health and Medicine. JAMA 1996 Oct 09;276(14):1172-1177. [Medline: 8827972]

XSL•FO
RenderX

5. Sculpher M. Clinical trials provide essential evidence, but rarely offer a vehicle for cost-effectiveness analysis. Value Health 2015 Mar;18(2):141-142 [FREE Full text] [doi: 10.1016/j.jval.2015.02.005] [Medline: 25773547]
6. Ioannidis JPA, Garber AM. Individualized cost-effectiveness analysis. PLoS Med 2011 Jul;8(7):e1001058 [FREE Full text] [doi: 10.1371/journal.pmed.1001058] [Medline: 21765810]
7. Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. Milbank Q 2004;82(4):661-687 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00327.x] [Medline: 15595946]
8. Jones J, Amaddeo F, Barbui C, Tansella M. Predicting costs of mental health care: a critical literature review. Psychol Med 2007 Apr;37(4):467-477. [doi: 10.1017/S0033291706009676] [Medline: 17202004]
9. van Breda W, Bremer V, Becker D, Hoogendoorn M, Funk B, Ruwaard J, et al. Predicting therapy success for treatment as usual and blended treatment in the domain of depression. Internet Interv 2018 Jun;12:100-104 [FREE Full text] [doi: 10.1016/j.invent.2017.08.003] [Medline: 29862165]
10. Kautzky A, Dold M, Bartova L, Spies M, Vanicek T, Souery D, et al. Refining Prediction in Treatment-Resistant Depression: Results of Machine Learning Analyses in the TRD III Sample. J Clin Psychiatry 2018;79(1):16m11385. [doi: 10.4088/JCP.16m11385] [Medline: 29228516]
11. Triñanes Y, Atienza G, Louro-González A, de-las-Heras-Liñero E, Alvarez-Ariza M, Palao DJ. Development and impact of computerised decision support systems for clinical management of depression: A systematic review. Rev Psiquiatr Salud Ment 2015;8(3):157-166 [FREE Full text] [doi: 10.1016/j.rpsm.2014.10.004] [Medline: 25500093]
12. Dunlop BW. Prediction of treatment outcomes in major depressive disorder. Expert Review of Clinical Pharmacology 2015 Aug 02;8(6):669-672. [doi: 10.1586/17512433.2015.1075390] [Medline: 26289221]
13. O'Hagan A, Stevens JW. The probability of cost-effectiveness. BMC Med Res Methodol 2002;2:5 [FREE Full text] [Medline: 11914138]
14. Kleiboer A, Smit J, Bosmans J, Ruwaard J, Andersson G, Topooco N, et al. European COMPARative Effectiveness research on blended Depression treatment versus treatment-as-usual (E-COMPARED): study protocol for a randomized controlled, non-inferiority trial in eight European countries. Trials 2016 Dec 03;17(1):387 [FREE Full text] [doi: 10.1186/s13063-016-1511-1] [Medline: 27488181]
15. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biol Psychiatry 2003 Sep 01;54(5):573-583. [Medline: 12946886]
16. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
17. EuroQol G. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy 1990 Dec;16(3):199-208. [Medline: 10109801]
18. Hakkaart-van Roijen L, van Straten A, Donker M, Tiemens B. Trimbos/iMTA questionnaire for costs associated with psychiatric illness (TiC-P). Institute for Medical Technology Assessment, Erasmus University Rotterdam, Trimbos 2002.
19. M Versteegh M, M Vermeulen K, M A A Evers S, de Wit G, Prenger R, A Stolk E. Dutch Tariff for the Five-Level Version of EQ-5D. Value Health 2016 Dec;19(4):343-352 [FREE Full text] [doi: 10.1016/j.jval.2016.01.003] [Medline: 27325326]
20. Hakkaart-van Roijen L, van der Linden N, Bouwmans C, Kanters T, Tan S. Kostenhandleiding: Methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg. In opdracht van Zorginstituut Nederland 2015.
21. Kolovos S, van Dongen JM, Riper H, Buntrock C, Cuijpers P, Ebert DD, et al. Cost effectiveness of guided Internet-based interventions for depression in comparison with control conditions: An individual-participant data meta-analysis. Depress Anxiety 2018 Mar;35(3):209-219 [FREE Full text] [doi: 10.1002/da.22714] [Medline: 29329486]
22. 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]
23. Manning WG, Mullahy J. Estimating log models: to transform or not to transform? J Health Econ 2001 Jul;20(4):461-494. [Medline: 11469231]
24. Burges CJC. A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery 1998;2(2):121-167. [doi: 10.1023/A:1009715923555]
25. Willmott CJ, Matsuura K, Robeson SM. Ambiguities inherent in sums-of-squares-based error statistics. Atmospheric Environment 2009 Jan;43(3):749-752. [doi: 10.1016/j.atmosenv.2008.10.005]
26. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. Geosci. Model Dev 2014 Jun 30;7(3):1247-1250. [doi: 10.5194/gmd-7-1247-2014]
27. Tibshirani R. Regression SelectionShrinkage via the Lasso. Journal of the Royal Statistical Society Series B 1996;58(1):267-288.
28. Black WC. The CE plane: a graphic representation of cost-effectiveness. Med Decis Making 1990;10(3):212-214. [doi: 10.1177/0272989X9001000308] [Medline: 2115096]
29. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2015. URL: https://www.r-project.org/ [accessed 2018-07-21] [WebCite Cache ID 714xWEfju]

XSL•FO
RenderX

30. Klok RM, Postma MJ. Four quadrants of the cost-effectiveness plane: some considerations on the south-west quadrant. Expert Rev Pharmacoecon Outcomes Res 2004 Dec;4(6):599-601. [doi: 10.1586/14737167.4.6.599] [Medline: 19807531]
31. National Institute for Health and Care Excellence (NICE). Guide to the methods of technology appraisal. 2013. Process and methods guides URL: https://www.ncbi.nlm.nih.gov/pubmed/27905712 [accessed 2018-07-21] [WebCite Cache ID 714xsKuRW]
32. Lord J, Laking G, Fischer A. Non-linearity in the cost-effectiveness frontier. Health Econ 2006 Jun;15(6):565-577. [doi: 10.1002/hec.1083] [Medline: 16416416]
33. Sim I, Gorman P, Greenes RA, Haynes RB, Kaplan B, Lehmann H, et al. Clinical decision support systems for the practice of evidence-based medicine. J Am Med Inform Assoc 2001;8(6):527-534 [FREE Full text] [Medline: 11687560]
34. Berg M. Implementing information systems in health care organizations: myths and challenges. Int J Med Inform 2001 Dec;64(2-3):143-156. [Medline: 11734382]
35. Hartswood M, Procter R, Rouncefield M, Sharpe M. Being there and doing IT in the workplace: A case study of a co-development approach in healthcare. In: Proceedings of the participatory design conference. 2000 Presented at: 6th Biennial Participatory Design Conference; 28 Nov-1 Dec; New York p. 96-105.
36. Atkinson CJ, Peel VJ. Transforming a hospital through growing, not building, an electronic patient record system. Methods Inf Med 1998 Sep;37(3):285-293. [Medline: 9787630]

Abbreviations

BT: blended treatment
ICER: incremental cost-effectiveness ratio
MAE: mean absolute error
QALY: quality-adjusted life years
RMSE: root mean square error
SVR: support vector regression
TAU: treatment as usual

Edited by G Eysenbach; submitted 01.03.18; peer-reviewed by JC Ateca, H Zhu, G Strudwick; comments to author 23.05.18; revised version received 05.06.18; accepted 16.06.18; published 21.08.18.

Please cite as:
Bremer V, Becker D, Kolovos S, Funk B, van Breda W, Hoogendoorn M, Riper H
Predicting Therapy Success and Costs for Personalized Treatment Recommendations Using Baseline Characteristics: Data-Driven Analysis
J Med Internet Res 2018;20(8):e10275
URL: http://www.jmir.org/2018/8/e10275/
doi:10.2196/10275
PMID:30131318

©Vincent Bremer, Dennis Becker, Spyros Kolovos, Burkhardt Funk, Ward van Breda, Mark Hoogendoorn, Heleen Riper. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 21.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Improving Consumer Understanding of Medical Text: Development and Validation of a New SubSimplify Algorithm to Automatically Generate Term Explanations in English and Spanish

Nicholas Kloehn[1], MA, PhD; Gondy Leroy[1], MS, PhD; David Kauchak[2], PhD; Yang Gu[1], BA; Sonia Colina[3], PhD; Nicole P Yuan[4], MA, MPH, PhD; Debra Revere[5], MA, MLIS

[1]Department of Linguistics, University of Arizona, Tucson, AZ, United States

[2]Computer Science Department, Pomona College, Claremont, CA, United States

[3]Department of Spanish and Portuguese, University of Arizona, Tucson, AZ, United States

[4]Health Promotion Sciences Division, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, United States

[5]Department of Health Services, School of Public Health, University of Washington, Seattle, WA, United States

Corresponding Author:
Nicholas Kloehn, MA, PhD
Department of Linguistics
University of Arizona
Communications Building, Room 109
PO Box 210025
Tucson, AZ, 85721
United States
Phone: 1 904 314 4654
Fax: 1 520 621 8105
Email: nick.kloehn@gmail.com

Abstract

Background: While health literacy is important for people to maintain good health and manage diseases, medical educational texts are often written beyond the reading level of the average individual. To mitigate this disconnect, text simplification research provides methods to increase readability and, therefore, comprehension. One method of text simplification is to isolate particularly difficult terms within a document and replace them with easier synonyms (lexical simplification) or an explanation in plain language (semantic simplification). Unfortunately, existing dictionaries are seldom complete, and consequently, resources for many difficult terms are unavailable. This is the case for English and Spanish resources.

Objective: Our objective was to automatically generate explanations for difficult terms in both English and Spanish when they are not covered by existing resources. The system we present combines existing resources for explanation generation using a novel algorithm (SubSimplify) to create additional explanations.

Methods: SubSimplify uses word-level parsing techniques and specialized medical affix dictionaries to identify the morphological units of a term and then source their definitions. While the underlying resources are different, SubSimplify applies the same principles in both languages. To evaluate our approach, we used term familiarity to identify difficult terms in English and Spanish and then generated explanations for them. For each language, we extracted 400 difficult terms from two different article types (General and Medical topics) balanced for frequency. For English terms, we compared SubSimplify's explanation with the explanations from the Consumer Health Vocabulary, WordNet Synonyms and Summaries, as well as Word Embedding Vector (WEV) synonyms. For Spanish terms, we compared the explanation to WordNet Summaries and WEV Embedding synonyms. We evaluated quality, coverage, and usefulness for the simplification provided for each term. Quality is the average score from two subject experts on a 1-4 Likert scale (two per language) for the synonyms or explanations provided by the source. Coverage is the number of terms for which a source could provide an explanation. Usefulness is the same expert score, however, with a 0 assigned when no explanations or synonyms were available for a term.

Results: SubSimplify resulted in quality scores of 1.64 for English (P<.001) and 1.49 for Spanish (P<.001), which were lower than those of existing resources (Consumer Health Vocabulary [CHV]=2.81). However, in coverage, SubSimplify outperforms all existing written resources, increasing the coverage from 53.0% to 80.5% in English and from 20.8% to 90.8% in Spanish

XSL•FO
RenderX

(*P*<.001). This result means that the usefulness score of SubSimplify (1.32; *P*<.001) is greater than that of most existing resources (eg, CHV=0.169).

Conclusions: Our approach is intended as an additional resource to existing, manually created resources. It greatly increases the number of difficult terms for which an easier alternative can be made available, resulting in greater actual usefulness.

(J Med Internet Res 2018;20(8):e10779) doi:10.2196/10779

KEYWORDS

text simplification; health literacy; natural language processing; terminology

Introduction

Background and Significance

Text is an important resource for health-related information as it is easy to create and distribute. Furthermore, health literature is widely available in the form of web-based resources for people to obtain information on medical conditions, diseases, and modalities [1]. However, these documents are often written at a level beyond the comprehension of the average reader [2]. This disconnect reflects an overall trend in misinformation regarding health conditions [3,4].

To mitigate this problem, researchers have sought automatic ways to improve the readability of these texts and the resulting reader comprehension. This natural language programming (NLP) task is known as *text simplification* [5] and has been used to create supervised [6], semisupervized [7], and fully automatic tools [8] to make texts easier for consumers to digest by increasing readability [9]. A central challenge for this research is to develop resources and techniques that enhance the quality and accuracy of these systems. Even though deep neural network approaches and other automated translation algorithms are increasingly being developed, it will take time before they can be applied with sufficient impact and precise simplifications. We intend for our algorithm to supplement existing resources as well as generate useful inputs for other algorithms.

The first step is identifying what makes text difficult. Some of the previous studies have focused on simplifying individual terms, while others have focused on grammatical structures. To identify the difficulty of individual terms, we use *term familiarity*. For a given term, this measure can be calculated by extracting the likelihood that a term occurs in common language usage [10], which we estimate according to the term's frequency in the Google Web Corpus [11]. In this work, we add to the body of research that identifies these terms and replaces them with easier synonyms [12]. However, we went beyond existing approaches by generating new explanations for terms that do not exist in the available resources. To do this, we developed and evaluated a new algorithm to generate new explanations. We generated explanations of terms in plain language using word internal parsing and affix dictionaries with SubSimplify.

Resources for Finding Explanations for Difficult Terms

Ideally, there would be an endless resource of expert-written explanations for difficult terms, optimized for the general public in multiple languages. However, few resources are able to provide appropriate explanations at all and even fewer are able to automatically or semiautomatically produce such explanations.

The resource closest to ideal is the English Consumer Health Vocabulary (CHV) [13], which is included in the Unified Medical Language System (UMLS) [14]. This resource was manually created and provides synonyms as well as definitions for medical terms in a consumer-friendly language. For the purposes of text simplification, these plain language definitions and simple synonyms double as ready-made explanations for difficult terms. However, the number of explanations is low relative to the overall number of difficult terms that occur in a given medical text. The CHV contains 2567 unique definitions and 88,529 synonyms for concepts found in the UMLS. We did not employ the UMLS as a resource because this system focuses on mapping complex medical concepts onto ontologies and is not designed to relate health information to patients or any other person outside the medical domain.

Previous research has shown that the CHV can be used to simplify texts [15-17], but it has also been shown to contain jargon words and not enough consumer-friendly vocabulary when providing summaries for specialized research [18]. Furthermore, while this resource is well tailored to text simplification, it is limited to English terms and explanations. In summary, the CHV provides explanations that can be automatically sourced in a given simplification system. However, CHV is only in English, is for relatively few terms, and can at times contain jargon beyond the reading level of the average reader.

While not being medically focused, WordNet is a useful resource for text simplification. It is an online lexical database containing terms and definitions, as well as interword semantic relations such as hypernyms, hyponyms, synonyms, and antonyms [19]. WordNet provides 128,391 word-sense definitions in English and is also available in Spanish, albeit in a less complete form [20]. Since WordNet is not a medical resource, many of its explanations are not optimal for medical text simplification, and when several senses are provided for a word, it is not always clear which best suits the medical sense. Previously, WordNet has been used to provide synonyms for lexical simplification [21]. For example, hyponym-hypernym relations have been used to generate synonyms that are simpler (more general) for text simplification [22]. In other areas, this resource has been used to simplify texts in the domain of biomolecules [23] and in texts written for non-native English speakers [24]. In summary, while WordNet is larger than CHV and also available in Spanish, the resource is not always optimal for giving the definition for medical terms.

XSL•FO
RenderX

Recent developments using neural networks trained on large bodies of text have produced larger resources such as word embeddings, where words are represented by multidimensional word vectors. The resulting vectors position the word relative to each other in a multidimensional space and have been shown to possess semantic and syntactic relations that allow us to automatically find synonyms and semantically related terms [12]. Given a word, we can use its vector representation to find the word whose vector is nearest to this word. Often, this nearest vector is a synonymous word. One freely available version of this resource is the pretrained Global Vectors for Word Representation (also known as GLoVe) [25]. Prior work has shown that these vectors can prove to be useful to isolate simple yet more frequent terms in the areas of text simplification [26]. However, they can include spurious matches because the approach cannot differentiate antonyms from synonyms. Given that this resource is totally automated, a word vector model can be produced from any language given a relatively large body of text. This means that this resource is also available for Spanish, with pretrained vectors available online [27]. In our study, we employed the GLoVe pretrained vectors for English and for Spanish [27], labeling the approach more generally as Word Embedding Vectors (WEV).

In all, the methods that exist for explanation generation range from specific, and precise, with low coverage to high coverage, with a much lower relative accuracy. In the next section, we describe our approach, which exists on the spectrum between these resources.

Methods

Using Morphological Information to Generate Explanations

We first describe the role that morphological units play in medical terminology and then our algorithm, which extracts information and generates explanations using these morphological units.

The resources described above make use of a word's definition in isolation without reference to the internal characteristics of that word, (ie, the morphology of the word). While it is not always the case, often romance languages contain morphological units that contain relatively clear semantics, such as the case for the prefix *anti-* ("against"), or the suffix *-s* (indicating plural). In certain words in English and in Spanish, these can help one to decipher the meaning of a word. In medicine, many terms, both in English and Spanish, originate from Greek and Latin [28,29]. Greek and Latin affixes have meanings commonly unknown to the average reader, but they nevertheless reflect the overall meaning of a word. While at times the meaning of a word is a direct function of the composition of the meaning of these morphological units, to a large degree in English and Spanish, terms composed of these units tend to have a gestalt effect. On the extreme end, a term may completely differ from the meaning of its morphological units (eg, "ledger" does not mean "ledge"+*er*). However, this problem of semantic drift is small for medical terms, seemingly because medical terms are less affected by semantic drift than more nonmedical, frequent terms.

Affixes that compose medical terms commonly have clear definitions that reflect a word's meaning. For example, given the prefix *cardio-* we know that this term's meaning relates to "the heart." Several resources containing these affixes and their definitions are freely available online [30-33]. From these, we created a unique dictionary of affixes along with their definitions for each language. We extracted 586 unique affixes for English and 498 affixes for Spanish. We define an affix as any morphological unit that has some denotation apart from the word itself. Affixes are categorized by their position, with prefixes occurring at the beginning of the words and suffixes occurring at the end of words. A root is any morphological unit that can stand alone as a single word. For example, the term *cardiovascular* contains the prefix *cardio-*, and the root *vascular*. Independently, these morphological units denote *the heart* and *consisting of a vessel or vessels*, respectively. Although many resources may not contain a definition for *cardiovascular*, by parsing these morphological units, we can automatically generate an explanation that reflects the actual denotation of the term: *relating to the heart and blood vessels*.

In both Spanish and English, words may be composed of multiple suffixes, roots, and prefixes. SubSimplify exploits this fact to generate an explanation for a term. Table 1 shows the examples of affixes and their definitions in both Spanish and English.

In addition to these affix dictionaries, we use word stemming [34] to isolate stemmed, or lemmatized, versions of terms. Stemming and lemmatization are two different methods of reducing a term to something similar to its root, but in a way that does not always reflect the actual root. For example, a resource like WordNet may have a definition for *Gastrointestine*, but not *Gastroinstestinal*. By stemming and stripping the affix *-al*, we increase the ability to find explanations using all resources.

Figure 1 provides an overview of our SubSimplify algorithm. The input to SubSimplify is a term we assume to be difficult, and we recursively lookup affixes and generate an explanation by accumulating the definitions of each affix and root identified. When finished, we align these definitions to provide an explanation of the term.

We use affix dictionaries to identify morphological units programmatically. First, the system identifies affixes and then takes the part of the word that is not an affix and performs a database lookup on stemmed variants of the term. To avoid spurious matches, we work from larger to smaller suffixes, and thus, *anti-* as in *anti-hero* would match before *a-* as in *a-symmetry*. This process occurs iteratively until no affixes are matched, or until there is no root left. In order to describe this process in sequence, Textbox 1 gives a detailed description of each step.

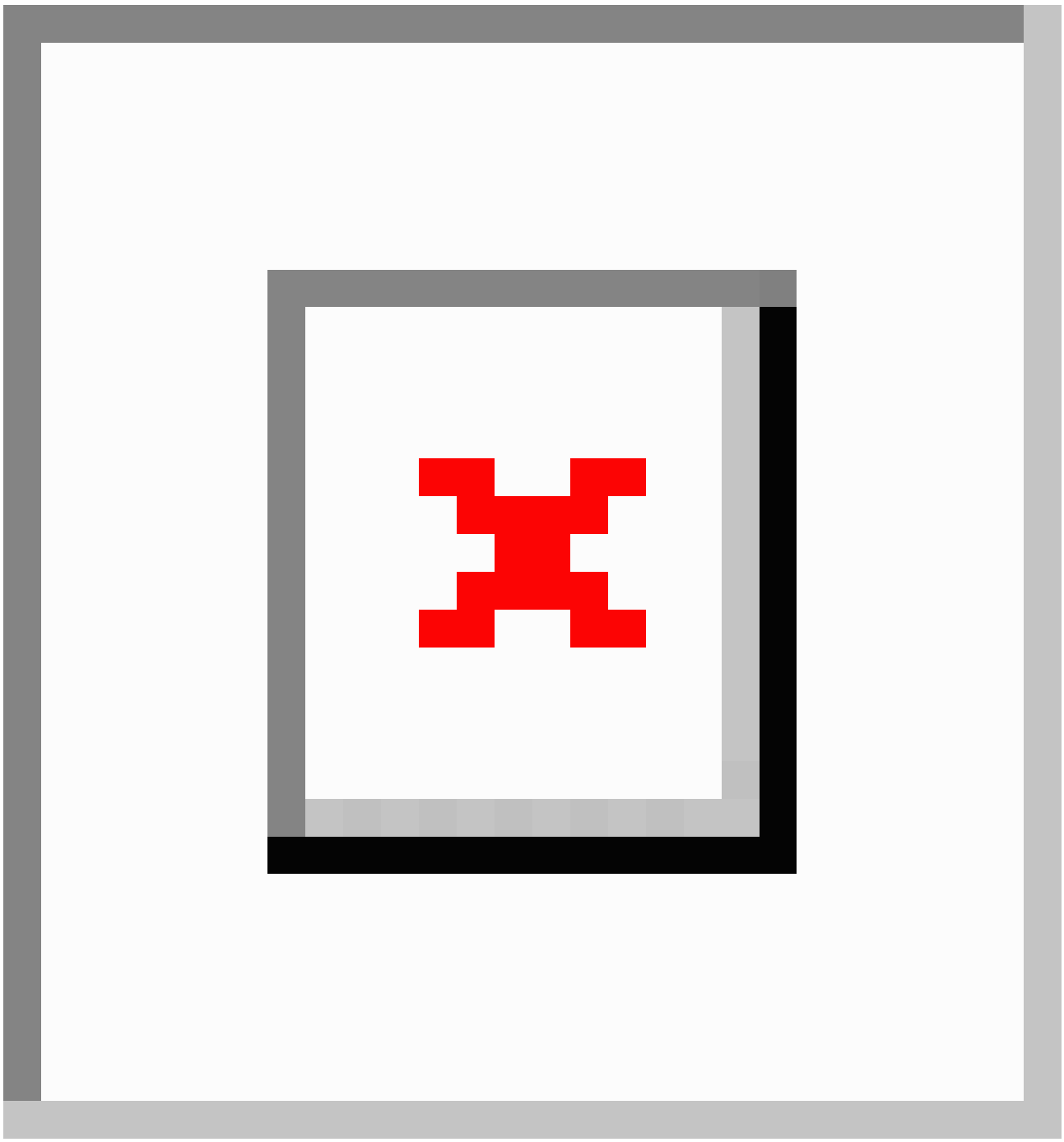
Since words may contain multiple suffixes, the process occurs multiple times where possible. That is, when we extract a root, it is possible that that root may yet contain another suffix or prefix. To highlight this, we provide an example with the term *hyperglycemic* in Textbox 2.

XSL•FO
RenderX

Table 1. Examples of affixes and corresponding definitions in English and Spanish.

Language and affix	Type	Definition	Origin
English			
adip-	prefix	*Of or relating to fat or fatty tissue*	Latin
-dipsia	suffix	*(condition of) thirst*	Greek
Spanish			
pireto-	prefix	*Forma prefija que significa fiebre*	Latin or Greek
-opsia	suffix	*Forma sufija que significa visión*	Greek

Figure 1. SubSimplify flow diagram.



XSL•FO
RenderX

Textbox 1. A description of each step in the SubSimplify algorithm.

Affix Identification:

All affixes in the affix dictionary are compared to the term from the longest to shortest length. If the term contains the affix characters at the beginning (for prefixes) or ending (for suffixes), the system considers this an affix match.

Affix Definition:

For each affix match, the affix dictionary definition is added to the newly constructed explanation.

Root Extraction:

The root of the term is extracted by removing the prefix or suffix. Since this can remove some of the characters of the root unintentionally, we consider the root the remaining characters plus single character variations of the root at the edge where the term was matched.

Search Resources:

The extracted root is then searched in WordNet and in CHV. If it is not found, we reintroduce the root to this same process until no matches are found.

Textbox 2. SubSimplify algorithm application to hyperglycemic.

Affix Identification:

We iteratively go through the affix dictionary and match the prefix *hyper-* in *hyperglycemic*.

Affix Definition:

The definition for *hyper* —"*denotes something as extreme or beyond normal"* —is added to the explanation for the term *hyperglycemic*.

Root Extraction:

We extract *glycemic* from *hyperglycemic.*

Search Resources:

WordNet and CHV are searched for *glycemic* and all single character variants of *glycemic* (eg, *aglycemic*). When not found, we rerun this entire process on *glycemic*, saving the explanation so far.

Table 2. Example English explanations.

Explanation resource	Example term	Explanation
CHV[a]	*pheochromocytoma*	A usually benign, well-encapsulated, lobular, vascular tumor of chromaffin tissue of the Adrenal Medulla
WordNet Summary	*Coryza*	an inflammation of the mucous membrane lining the nose (usually associated with nasal discharge)
WordNet Synonym	*attenuated*	rarefy
SubSimplify	*hyperglycemic*	hyper-glyc-em-ic, "extreme" or "beyond normal"-sugar-em-pertaining to
Word Vector Nearest Neighbor	*toxoplasma*	gondii

[a]CHV: Consumer Health Vocabulary.

This process repeats until there is either no root left, or until the remaining root fails to be identified by any resource. For *glycemic*, the system will identify *-ic* and subsequently *glyc-* before halting at *-em-*.

If the term contains "-" or any other *nonword* characters, we split these as well. The parsed affixes and roots are then aligned with their explanations to provide an affix-by-affix breakdown of the term. For any affix that is not identified in the system, as is the case with *-em-* in *hyperglycemic*, the definition of the root remains the root itself. Upon presenting the term, these affixes are matched with their definition both by order and by color in order to make identification as easy as possible for a writer. An example explanation for *hyperglycemic* is shown in Table 2. This table contains explanations for a few different difficult terms to highlight their quality when present. Note that not all resources contain explanations for all terms, so it is extremely rare that all resources can provide an explanation for a single term.

While the CHV [35] and WordNet Summary resources provided full-sentence explanations (semantic simplification), the WEV and WordNet Synonym provide single-word explanations of each term (lexical simplification). SubSimplify provides a hybrid of the two: for the individual parsed subword units, either a synonym or brief description is presented.

Next, we describe 2 studies designed to evaluate the quality, coverage, and usefulness of these explanations in English and Spanish.

Studies

To evaluate the quality, coverage, and usefulness of the newly generated explanations and how they compare to existing

XSL•FO
RenderX

resources, we conducted two studies: one in English and one in Spanish.

Study 1: English Term Explanation Generation

Study Stimuli

Stimuli

To obtain a range of medical terms that occur in common texts, we extracted 20 documents from Wikipedia written on a medical topic and 100 PubMed abstracts. From these documents, we extracted the difficult terms using term familiarity. For the purposes of this study, we identified difficult terms as those having a frequency less than the 5000th ranked term in the Google Web Corpus, which previous work showed to be a reasonable criterion [7]. Given these difficult terms, we selected 200 terms from each resource type (PubMed and Wikipedia) balanced across all documents (100 and 20, respectively). To investigate the effect of frequency, we also balanced each set of 200 difficult terms by frequency. Two groups were extracted based upon high and low frequencies. High-frequency terms were those which had frequencies in the upper most tertile, and low-frequency terms were those which had frequencies in the lowest tertile. In all, the study contained 400 total terms that were evenly split across high and low frequency, document source, and the documents themselves.

Explanation Generation

We compared our approach to four previous approaches: CHV, WordNet Synonyms and Summaries, and WEV. These resources provided explanations when an exact match could be found for the term in their database.

Metrics

For each of the 400 terms, we calculated 3 metrics: quality, coverage, and usefulness. Quality was judged by subject experts (SEs). The SEs in this study were required to (1) be a native speaker of the language and (2) have at least a master's degree in a public health or a medical-related field. The experts typically had experience evaluating the quality of medical resources, and for this study, they were financially compensated for their time.

For quality, the two SEs reviewed each term along with the candidate definitions and explanations. For each definition or explanation, the SEs annotated how useful it was on a 4-point Likert scale. Table 3 provides a description of each rating level. Coverage was measured by calculating the percentage of terms for which an explanation was provided by each source. Usefulness is a broader measure than quality and takes the availability of terms and resources into account. When a term is not found, it receives a score of 0. While quality gives us an idea of how accurate resource explanations are, usefulness tells us how well such a resource would perform if we were to employ it for all terms.

Procedure

The SEs evaluated the 400 terms and the corresponding explanations provided by each resource. The order of the presentation of explanations was randomized for each term. For each of the terms, the SEs scored the term on both the quality and coverage metrics described above. We then calculated usefulness by normalizing quality by coverage.

In order to give a visual idea of how this study was performed, Figure 2 contains a flowchart containing the steps of the study.

Table 3. Likert quality scale.

Rating	Description
1	Explanation *is not useful* to someone annotating the text.
2	Explanation *is a little useful* to someone annotating the text.
3	Explanation *is useful* to someone annotating the text.
4	Explanation *is very useful* to someone annotating the text.

XSL•FO
RenderX

Figure 2. Steps for English term explanation generation study.

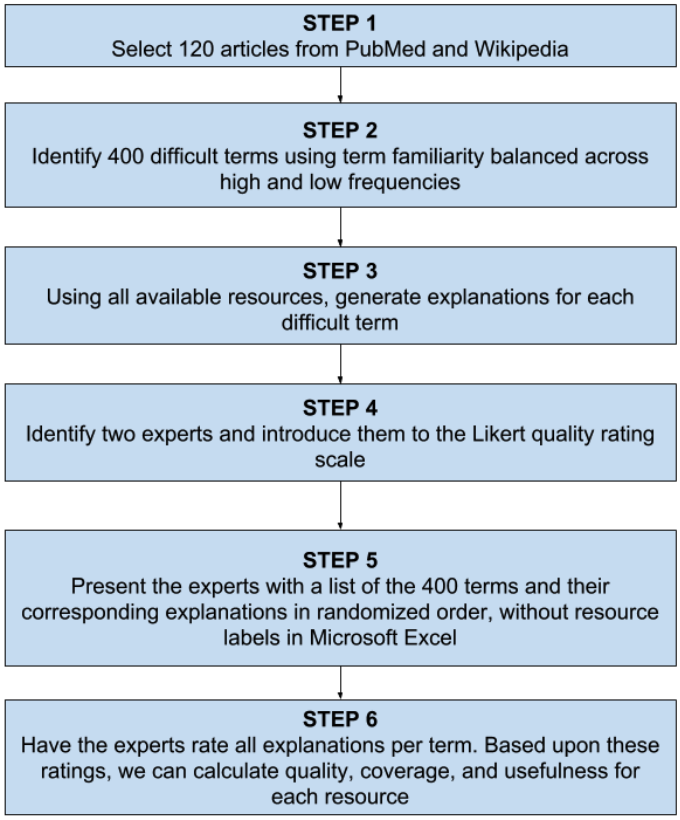


Table 4. English study results.

Metrics	CHV[a]	WordNet Synonym	WordNet Summary	SubSimplify	WEV[b]
Quality (1-4 scale)	2.81	2.09	3.32	1.64	1.64
Coverage (N=400), %	6.0	53.0	53.0	80.5	83.8
Usefulness (0-4 scale)	0.169	1.11	1.76	1.32	1.38

[a]CHV: Consumer Health Vocabulary.

[b]WEV: Word Embedding Vector.

Evaluation Outcomes

Interoperator Variability

To compare the variability in quality scores between each of the SEs, we calculated Crohnbach alpha. Since we did not limit the quality ratings to a rank order, it was possible for each term to have multiple explanations that received the best score per term. Therefore, we calculated Crohnbach alpha in two ways. First, in a conservative version, we calculated whether each SE chose all of the same explanations as the best for each term, and in a more liberal version, whether each SE chose one of the same explanations as the best for each term. For English, the results were 0.69 and 0.90 for the conservative and liberal version, respectively. We, therefore, determined that interoperator reliability was high enough to average their ratings. Table 4 shows the results of the quality, coverage, and usefulness metrics for each explanation source in English.

In Table 4, we see that each column represents explanation sources and the 3 rows give the metrics averaged across SEs. For example, CHV received a mean quality score of 2.81 when present, but could only provide explanations for 24 out of 400 total terms. Subsequently, its usefulness was only 0.169 for the 400 terms. Recall that this resource represents the one that is manually generated to aid lexical simplification in medical documents. As a consequence, the quality rating was relatively high, but the coverage was by far the lowest. Next, we see that WordNet Summaries and Synonyms each provided the same number of explanations. However, the Summaries (Semantic Simplification) scored much higher than the Synonyms (Lexical Simplification) at 3.32 versus 2.09, respectively. Again, given that they only provide explanations for 212 terms, their usefulness was only 1.76 and 1.11, respectively. While SubSimplify had a 1.64 quality score when present, its coverage was 322, whereas that of WordNet was 212, representing an increase from 53.0% to 80.5% in coverage of the 400 *difficult terms*. Consequently, the usefulness of SubSimplify was 1.32, greater than that of WordNet Synonyms and CHV. Last, WEV provided the greatest coverage and performed identically to SubSimplify in quality (1.64), but had greater coverage (335) and quality (1.38). However, as we describe in the next subsection, there was a clear difference between the quality

XSL•FO
RenderX

performance of SubSimplify and WEV. SubSimplify performed better with lower-frequency words and in more technical literature than WEV.

Quality

To evaluate significance, we performed a 2×2×5 analysis of variance (ANOVA), with quality as the dependent variable. The independent measures were document source (Wikipedia or PubMed), frequency (Low or High), and the five explanation sources (CHV, WordNet Synonym, WordNet Summary, SubSimplify, and WEV). There were main effects for frequency ($F_{1,2186}$ = 3.859, P<.02) and explanation type ($F_{4,2186}$ = 260.1, P<.001). This indicates that on average, the resources performed significantly better with lower-frequency terms and that there were significant differences between the resources.

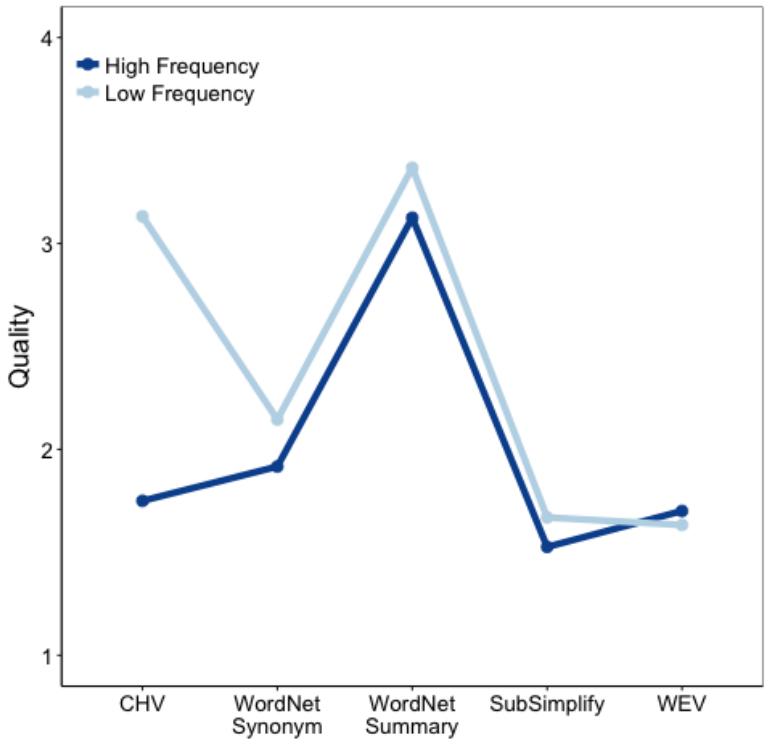
In addition to the main effects, there was a significant two-way interaction between explanation type and frequency ($F_{8,2186}$=2.993, P<.001; Figure 3). Figure 3 contains the mean quality of each resource at low and high frequencies. Given that our documents contained medical terminology, we expected low-frequency words to be the rarest and, therefore, most technical. They presented the hardest target for any system attempting to summarize these documents. For example, CHV, which is written specifically for medical terms, has much greater performance for low-frequency terms than for high-frequency terms (3.12 vs 1.74). Furthermore, WordNet Synonyms and Summaries both performed slightly better for low-frequency terms than for high-frequency ones. Interestingly, SubSimplify also followed this pattern. However, WEV had the opposite trend. Not only did WEV perform better on high-frequency terms than on low-frequency terms but also it performed slightly poorer than SubSimplify for low-frequency terms (compare SubSimplify's 1.67 rating to WEV's 1.63 rating for low-frequency terms).

Coverage

To evaluate the effect of frequency and document source on the coverage of each resource, we performed another 2×2×5 ANOVA with coverage as the dependent variable. There were main effects for frequency ($F_{1,3970}$=3.859, P<.001) and explanation type ($F_{4,3970}$=260.1, P<.001; refer to Table 2, row 2). This indicates that explanations, on average, had significantly greater coverage for high-frequency terms than for low-frequency terms.

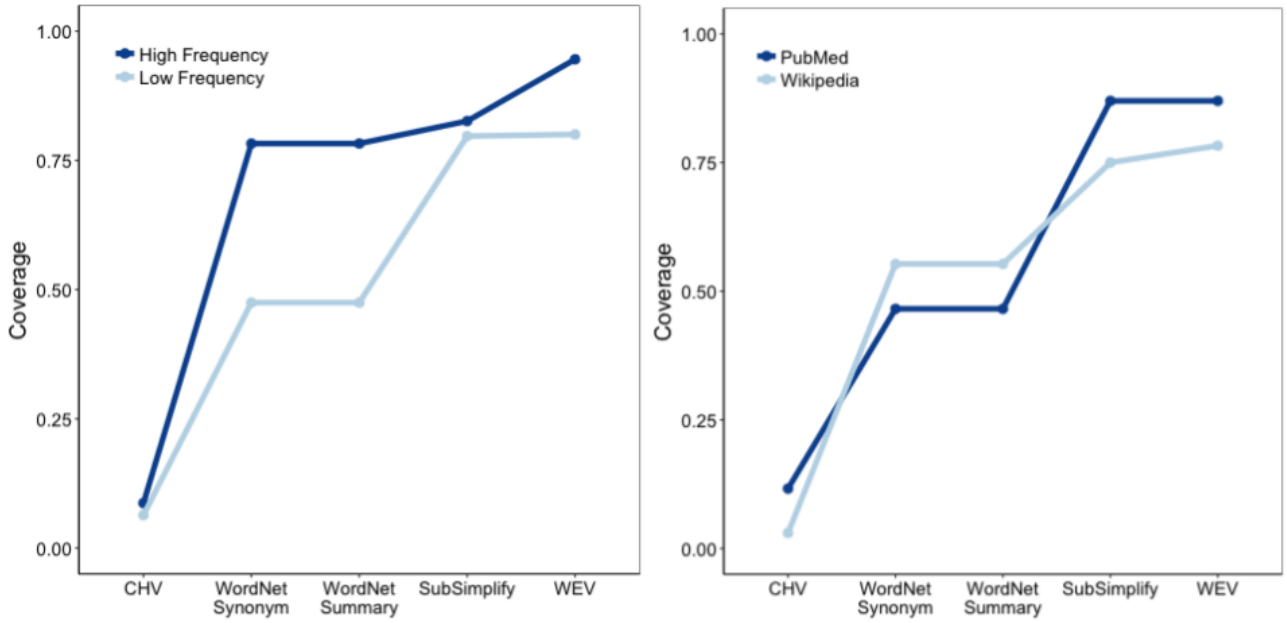
There was a significant two-way interaction between explanation type and frequency ($F_{8,3970}$=6.557, P<.001) and a significant interaction between explanation type and document source ($F_{4,3970}$=11.523, P<.001; Figure 4).

Figure 3. Explanation type-frequency interaction for quality in English. CHV: Consumer Health Vocabulary, WEV: Word Embedding Vector.



XSL•FO
RenderX

Figure 4. Coverage-frequency interaction (left) and coverage-document type interaction (right) in English. CHV: Consumer Health Vocabulary, WEV: Word Embedding Vector.



As seen in the left side of Figure 4, term frequency affects the coverage of each explanation type. Whereas CHV had similar coverage for low- and high-frequency terms, WordNet had much greater coverage for high-frequency terms than for low-frequency terms (0.47 vs 0.77). SubSimplify then increased the coverage of low-frequency terms from 0.47 to 0.78 and of high-frequency terms from 0.77 to 0.79. Last, WEV slightly increased the coverage of low-frequency terms (from 0.77 to 0.78), but increased that of high-frequency terms from 0.79 to 0.94. This indicates that SubSimplify performed quite similar on low- and high-frequency terms, as is the case with CHV, but with an overall much greater coverage.

Next, we look at the interaction of document source and coverage. Recall that PubMed contains more technical medical terms than Wikipedia sources, and therefore, it constitutes terms that should contain more technical jargon. In the right side of Figure 4, we see that CHV, SubSimplify, and WEV each had greater coverage in PubMed than in Wikipedia, whereas WordNet had greater coverage in Wikipedia documents. The x-axis depicts the proportional coverage for each explanation source in PubMed and Wikipedia, and the y-axis includes the change in coverage. WordNet, for example, provides fewer explanations for PubMed (0.48) than for Wikipedia (0.52), whereas SubSimplify provides more for PubMed (0.8) than for Wikipedia (0.75). One critical point to note is that SubSimplify has equivalent coverage to WEV. This indicates that SubSimplify performed as well as the fully automated WEV within technical medical text.

Usefulness

Next, we performed another 2×2×5 ANOVA with usefulness as the dependent variable. There was only a main effect for explanation type ($F_{4,3970}$=95.170, P<.001; refer to Table 2, row 3) and frequency ($F_{1,3970}$=14.663, P<.001). This indicates that the usefulness ratings were significantly different across the different explanations and that usefulness ratings were significantly greater for high-frequency terms on average.

There was a significant two-way interaction between explanation type and frequency ($F_{7,3970}$=5.390, P<.001) and a significant interaction between explanation type and document source ($F_{4,3970}$=6.387, P<.001; Figure 5).

As seen in Figure 5 on the left, term frequency affected the usefulness of each explanation type. While CHV had a greater usefulness for low-frequency versus high-frequency terms, WordNet Synonyms had much greater usefulness for high-frequency than low-frequency terms (1.50 vs 1.00). WordNet Summaries had the greatest usefulness score for high-frequency terms (2.40) and smaller, but still quite high, scores for low-frequency terms (1.62). Meanwhile, SubSimplify had greater usefulness for low-frequency terms than for high-frequency terms (1.43 vs 1.35). Last, WEV had a greater score for high-frequency than for low-frequency terms (1.64 vs 1.42). This affirmed the idea that SubSimplify is a resource that performs best for low-frequency terms.

Next, we look at the interaction of document source and coverage. Recall that PubMed contains more technical medical terms than Wikipedia sources, and, therefore, constitutes terms that should contain more technical jargon. In Figure 5, we see that WordNet performed better in Wikipedia and that SubSimplify and CHV both performed better in usefulness for PubMed. This affirmed that SubSimplify performs best on more technical documents.

Summary

For the English study, we found that SubSimplify performed better than existing medical resources for coverage and had a relatively high quality given its coverage. Furthermore, SubSimplify, much like CHV, performed better for low-frequency and more technical terms, than for high-frequency terms. In addition, this resource had the greatest

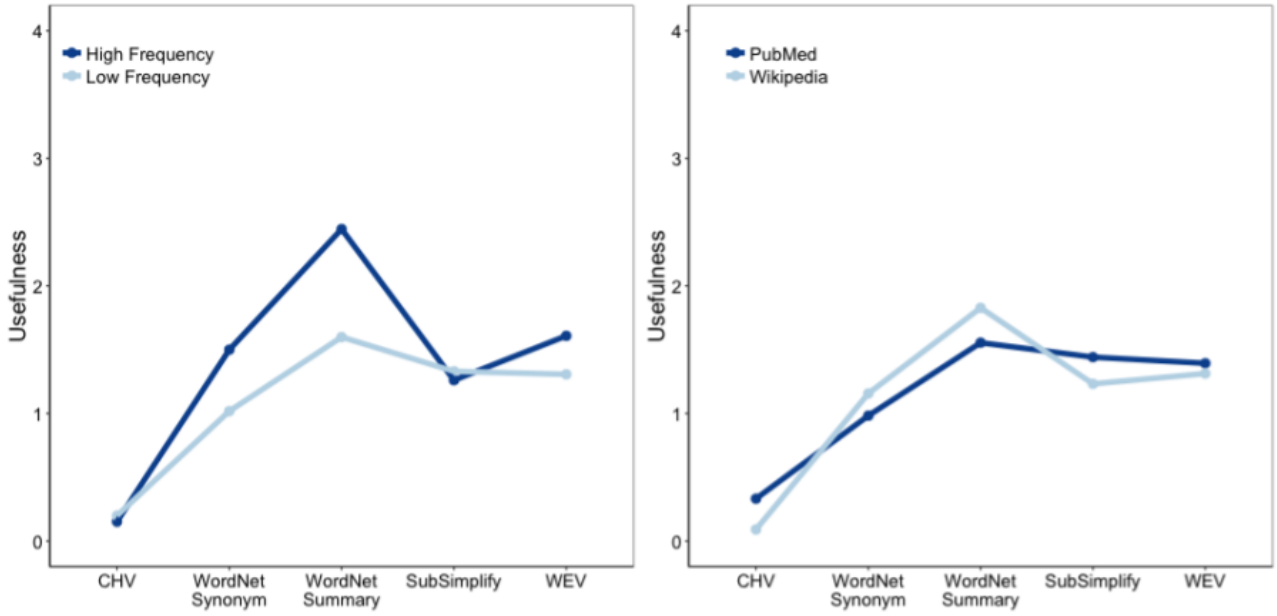
XSL•FO
RenderX

coverage of all the resources for terms found in the PubMed abstracts. These quality, coverage, and usefulness results suggest that SubSimplify is better equipped to generate explanations for low-frequency and technical terms than the other existing resources.

Study 2: Spanish Term Explanation Generation

The second study evaluated our approach in Spanish. This study was identical to the English study, save for two differences. First, there was no Spanish language CHV and WordNet in Spanish only contained summaries (no Spanish synonyms; the version we used contained only Spanish terms mapped to English synonyms). Therefore, we only used WordNet Summaries and compared only three possible explanation resources: our approach (SubSimplify), WordNet, and WEV. Second, since there were no Spanish language PubMed abstracts available, for our second resource, we used Medline Plus [36] instead, which is a resource for medical articles geared toward people interested in health information. Last, all instructions, ratings, and explanations were in Spanish.

Figure 5. Explanation type-frequency interaction (left) and explanation type-document source interaction (right) in English usefulness measures. CHV: Consumer Health Vocabulary, WEV: Word Embedding Vector.



Study Stimuli

Stimuli

We tested 400 medical terms balanced for frequency. To get a range of medical terms that occurred in both common texts and more technical texts, we extracted 20 documents from Wikipedia written on the topic of disease and 20 Medline Plus articles. From these documents, we first extracted difficult terms using the same term familiarity threshold. Our cutoff was a frequency less than the 5000th most common term in the Spanish Google Web Corpus [10]. Within these terms, we balanced the terms across low and high frequency. In all, we split these 400 terms across high and low frequency and document source.

Explanation Generation

We compared our approach to two previous approaches: WordNet and WEV. These resources provided explanations when an exact match could be found for the term in their database.

Metrics

For each of the 400 terms, we extracted explanations. For each explanation, we calculated quality, coverage, and usefulness.

Procedure

The procedure was identical to English except that all instructions and explanations were written in Spanish. The SEs were both bilingual Spanish-English speakers who had nevertheless received Master of Public Health degrees in English.

Evaluation Outcomes

Interoperator Variability

Again, we calculated Crohnbach alpha in both a liberal and conservative version. For Spanish, the results were 0.64 and 0.90 for the conservative and liberal versions, respectively. We, therefore, again determined that the interoperator reliability was high enough to collapse their quality ratings into one group.

Table 5 shows the results of the quality, coverage, and usefulness for each explanation source in Spanish. It can be seen that WordNet had the highest average quality rating of the three resources (2.64), but provided the lowest coverage at 20.5%, with a low resulting usefulness (0.543). The coverage of SubSimplify was much greater at 90%, with a lower average quality rating (1.49). Last, WEV provided a higher quality rating (1.84) but with a lower coverage than SubSimplify (89.75%). Regarding usefulness, WEV outperformed SubSimplify (1.77 vs 1.24).

XSL•FO
RenderX

Given these results, we performed ANOVAs to understand the relationship between frequency and document source for the quality, coverage, and usefulness of each explanation type.

Quality

We performed a 2×2×3 ANOVA to evaluate the effect of document source (Wikipedia or PubMed) and frequency (High or Low) on the quality ratings for each of the three explanation resources (WordNet, SubSimplify, and WEV). There were main effects for frequency ($F_{1,1590}$=13.39, P<.001) and explanation type ($F_{2,1590}$=98.805, P<.001; refer to Table 3 row 1). This indicates that on average, the explanations performed significantly better on high-frequency terms and that there was a significant difference between the average quality of explanations based upon their type.

There was also a significant two-way interaction between explanation type and frequency ($F_{2,1590}$=12.010, P<.001; Figure 6). Figure 6 shows the interaction of frequency with explanation type. The x-axis depicts the mean quality for each explanation source at low and high frequency, and the y-axis includes the change in mean quality ratings as a line. SubSimplify performed better for low-frequency terms (1.54) than for high-frequency terms (1.48), whereas WEV performed worse for low-frequency terms (1.68) than for high-frequency terms (2.03).

Coverage

We performed a 2×2×3 ANOVA to evaluate the effect of document source (Wikipedia or MedlinePlus) and frequency (High or Low) on the coverage ratings for each of the three explanation resources (WordNet, SubSimplify, and WEV). There were also main effects for frequency ($F_{1,2382}$=7.180, P<.001) and explanation type ($F_{2,2382}$=1142.361, P<.001; refer to Table 3). This indicates that on average, the explanations had significantly better coverage on high-frequency terms and that there was a significant difference between the average coverage of explanations based upon their type.

There was a significant two-way interaction between explanation type and frequency ($F_{2,2382}$=4.465, P<.015), and a significant interaction between explanation type and document source ($F_{4,2382}$=6.259, P<.001; Figure 7). For Spanish terms, term frequency affected the coverage of each resource as seen in Figure 7 on the left. For WordNet, there was slightly greater coverage for low-frequency terms (0.22) than for high-frequency terms (0.20), but both were quite low. For SubSimplify, there was greater coverage for high-frequency terms (0.92) than for low-frequency terms (0.87). This was also the case for WEV, with high and low coverages at 0.93 and 0.86, respectively. This indicates that SubSimplify had the greatest coverage for low-frequency terms and WEV had the greatest coverage for high-frequency terms in Spanish.

For document source, WordNet had a greater coverage for MedlinePlus terms (0.25) than for Wikipedia terms (0.20). Likewise, SubSimplify performed better on the more technical terms of MedlinePlus (0.93) than on the more general terms of Wikipedia (0.87). WEV, however, had the opposite effect, with 0.93 for Wikipedia and 0.86 for MedlinePlus. In short, WEV performed better on less technical texts and higher-frequency terms, whereas SubSimplify performed better on low-frequency terms and more technical texts.

Table 5. Spanish study results.

Metrics	WordNet Summary	SubSimplify	WEV[a]
Quality (1-4 scale)	2.64	1.49	1.84
Coverage (N=400), %	20.5	90.0	89.7
Usefulness (0-4 scale)	0.543	1.24	1.77

[a]WEV: Word Embedding Vector.

XSL•FO
RenderX

Figure 6. Explanation type-frequency interaction for quality in Spanish. WEV: Word Embedding Vector.

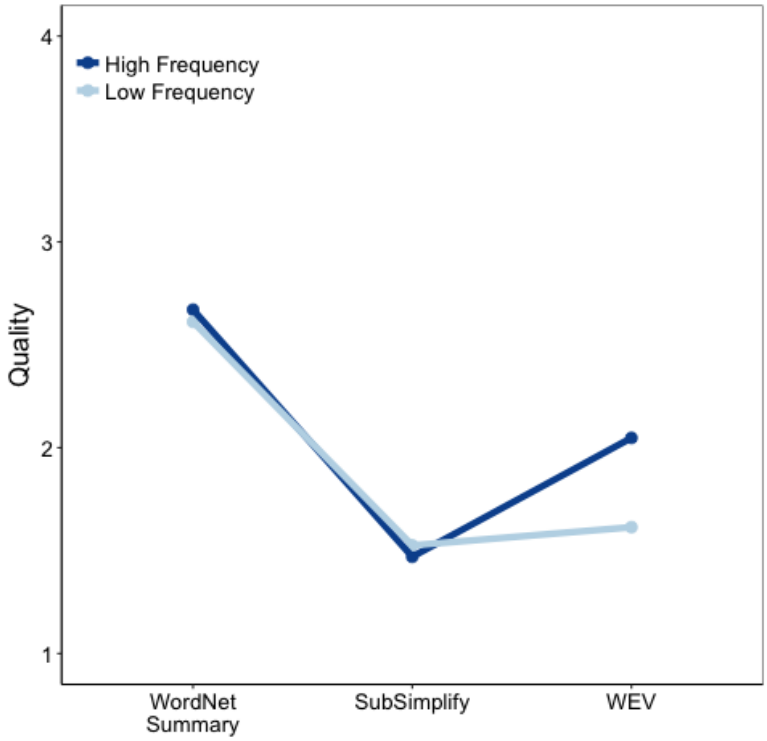
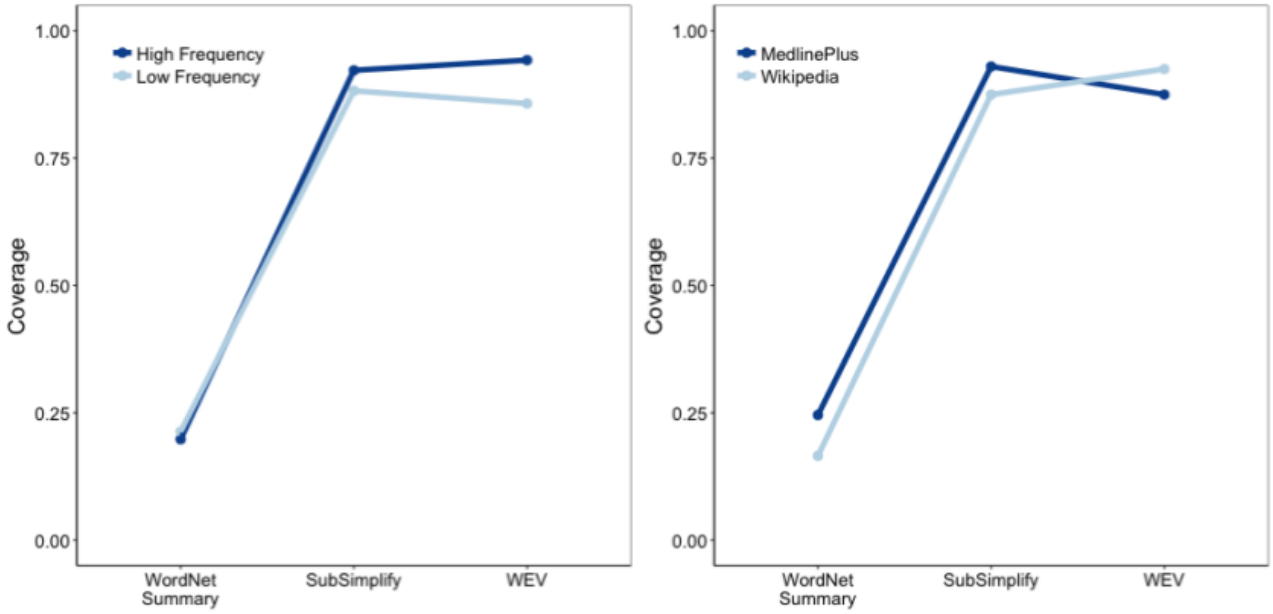
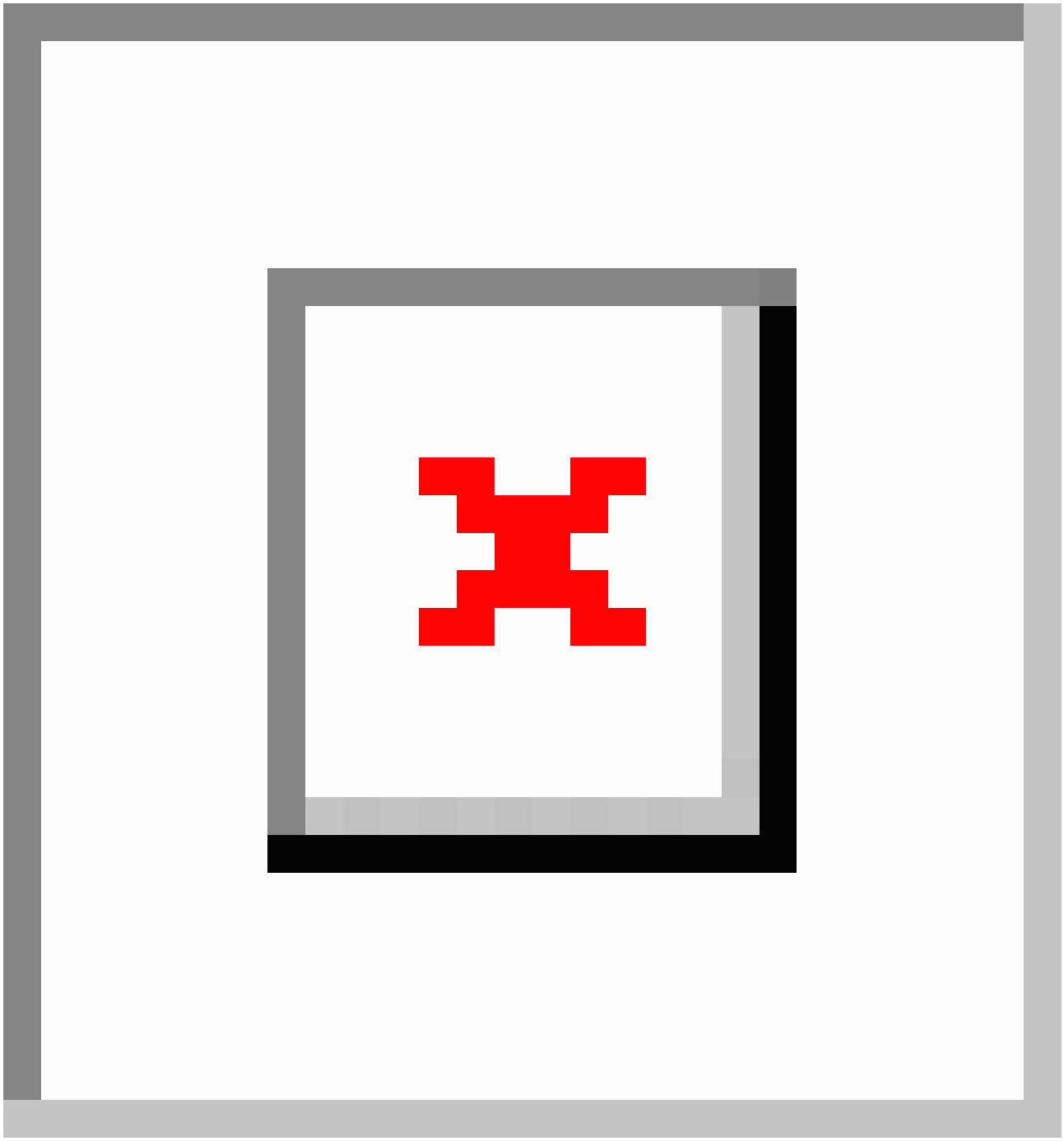


Figure 7. Coverage-frequency interaction (left) and coverage-document type interaction (right) in Spanish. WEV: Word Embedding Vector.



XSL•FO
RenderX

Figure 8. Explanation type-frequency interaction for frequency. WEV: Word Embedding Vector.



Usefulness

We performed the same 2×2×3 ANOVA, but this time with usefulness as the dependent measure. The results indicated main effects for frequency ($F_{1,2382}$=16.197, P<.001), explanation type ($F_{2,2382}$ = 230.268, P<.001), and document source ($F_{1,2382}$=6.737, P<.001). These results indicate that explanations, on average, performed significantly better on high-frequency terms and on Wikipedia documents. This also indicates that there was a significant difference in the usefulness measures for explanation type (refer to Table 3, row 3).

Furthermore, there was a significant two-way interaction between explanation type and frequency ($F_{2,2382}$=17.911, P<.001). Figure 8 depicts this interaction. Notice that WordNet Summaries performed nearly identically on low- and high-frequency terms (0.6). This pattern was also true for SubSimplify (1.4 for both). Last, WEV performed much better in usefulness on high-frequency terms than on low-frequency terms (1.9 vs 1.4).

Summary

For the Spanish study, we found that SubSimplify performed better than WordNet and WEV for coverage. Specifically, it performed best for low-frequency and more technical terms; its average quality was lower than that of the other two resources, but was better at low frequency.

XSL•FO
RenderX

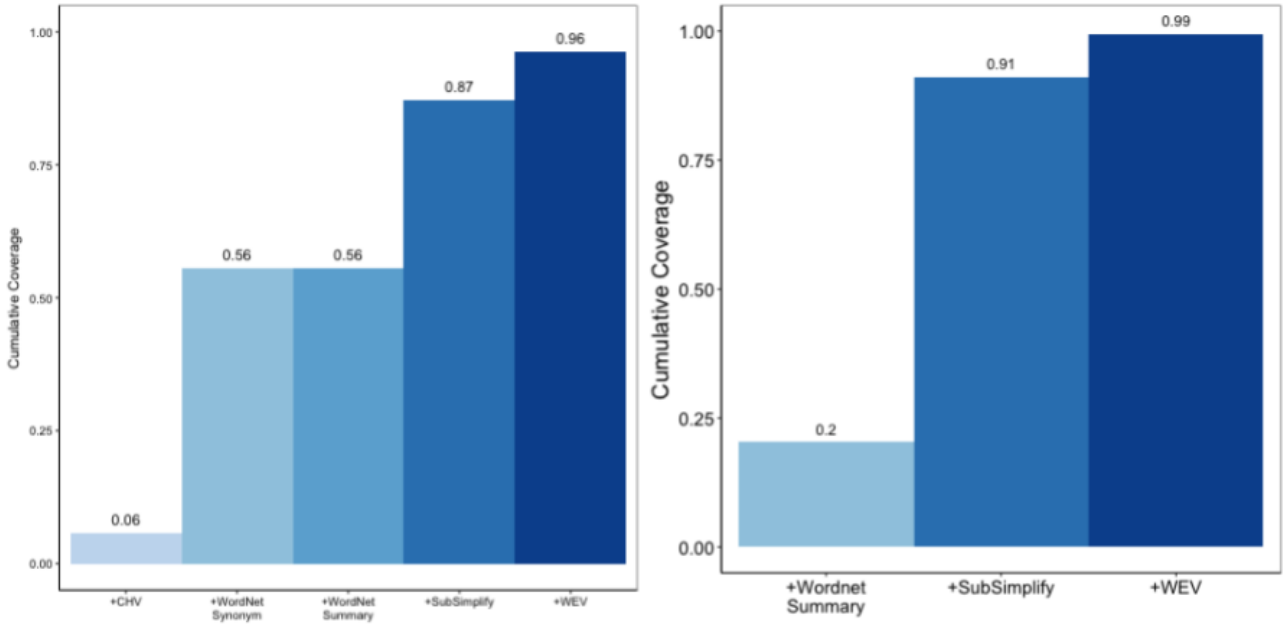
Results

In both English and Spanish, SubSimplify had its best quality ratings at low frequency, which were similar for both languages. Furthermore, in both languages, SubSimplify had similar results regarding coverage. In both the English and Spanish studies, we saw that SubSimplify greatly outperformed the existing resources with its ability to provide multiword explanations for difficult terms. Namely, SubSimplify outperformed CHV and WordNet Summaries in English in quality, and in Spanish, it outperformed WordNet in this same measure. Furthermore, it provided the most explanations at low frequencies and in more technical texts.

At the same time, much of the quality and coverage that we have shown covers overlapping data. Here we have evaluated the coverage of these resources if we were to employ all of them into a single system. Doing so will highlight the role that SubSimplify can play in a larger simplification system. Given the 400 terms in each language, the charts in Figure 7 highlight the cumulative coverage of each resource. Given that not all resources cover the same words, these bar graphs show the coverage of a system that includes each nonoverlapping explanation or synonym from the previous resource.

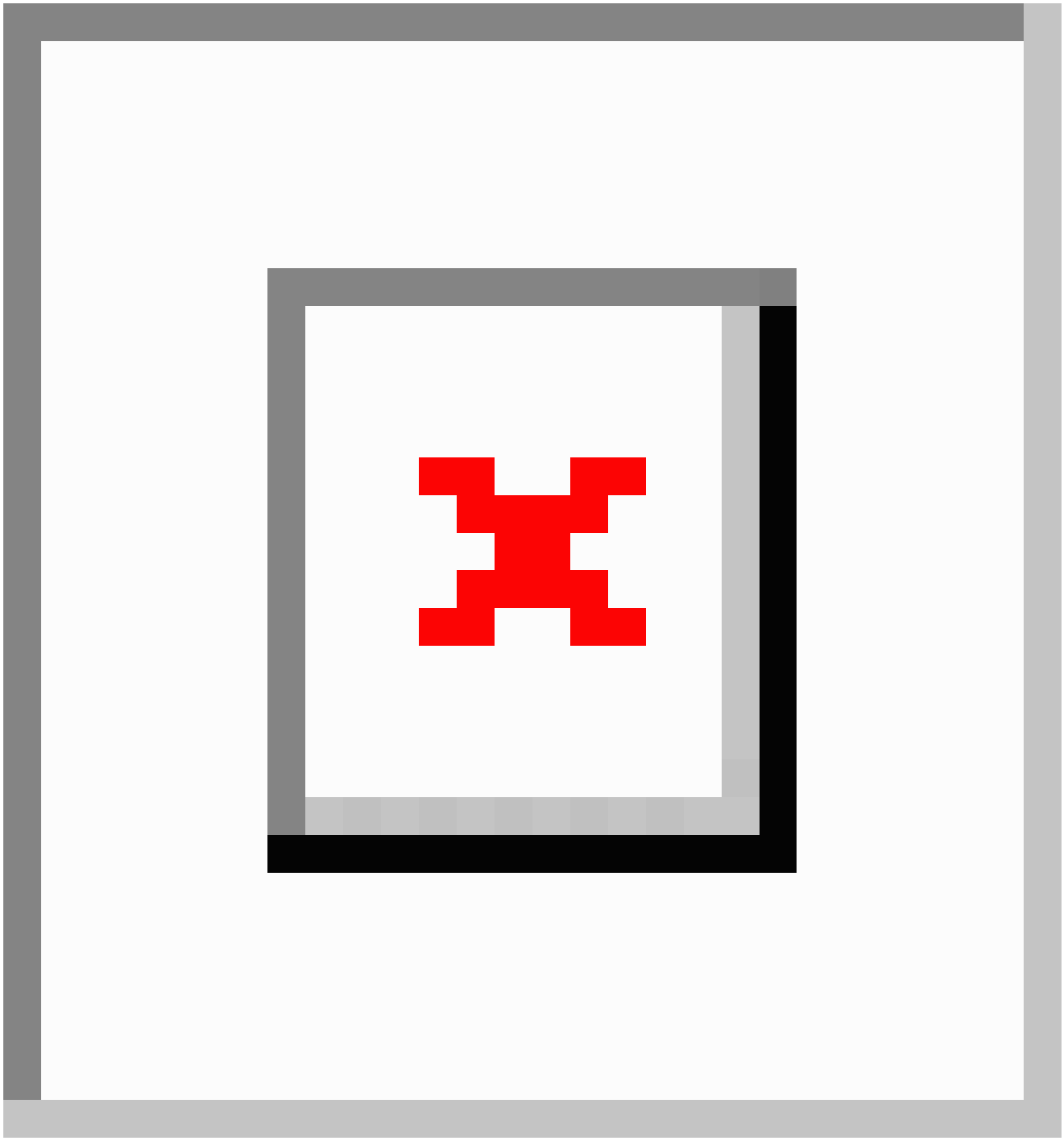
Left, CHV provides the lowest coverage with 23, WordNet then provides 212, SubSimplify provides 322, and, finally, WEV provides 336 out of 400. On the right, we see the number of explanations that our system can provide as we add each resource in Spanish. For example, if we only used CHV, we would only be able to provide explanations for 23 terms. However, as we add each resource, the number of (nonoverlapping) terms for which we can provide explanations increases. As we add WordNet, we can provide explanations for 222 terms; then by adding SubSimplify, we can provide explanations for 349. Last, by adding WEV, we can provide a total of 385 explanations out of the 400 total terms, or 99% of all terms.

Figure 9. Cumulative coverage in English (left) and Spanish (right). CHV: Consumer Health Vocabulary, WEV: Word Embedding Vector.



XSL•FO
RenderX

Figure 10. Cumulative usefulness in English (left) and Spanish (right). CHV: Consumer Health Vocabulary, WEV: Word Embedding Vector.



SubSimplify acts as an intermediate resource between the fully automatic, synonym-providing source of WEV and the annotator-written resources of CHV and WordNet. In its semiautomated approach, it increases the coverage of total terms for which any simplification system can provide multiword explanations. This also is made apparent when we look at the cumulative usefulness of all explanations, which can be found in Figure 9. Here we have provided the cumulative quality rating for all terms additively. For example, in English (left), the CHV quality is the average rating of explanations for all 400 terms, most of them being 0. Then +WordNet Synonyms gives the average quality of these two resources combined. When two resources each provide an explanation, we take the higher rated explanation of the two. The result is that by employing SubSimplify in conjunction with all resources, this simplification system can provide explanations with a 2.64 usefulness rating in English and 2.09 in Spanish (right).

Discussion

Principal Findings

The aim of the English and Spanish studies was to evaluate the efficacy of employing SubSimplify to medical texts, and the results revealed what was expected. Compared with WordNet and CHV, the quality of explanations was, on average, lower. This may be an indication of a few different issues, which we will expand upon in the Limitations.

XSL•FO
RenderX

Surprisingly, the fully automatic system of WEV outperformed our expectations. In creating SubSimplify, we imagined that there would be many spurious matches and synonyms that were unrelated to the difficult terms, but the results showed a better performance than expected. Based upon this, we are motivated to employ WEV as another way to source root synonyms in SubSimplify. That is, in its current form, SubSimplify performs a term lookup in WordNet (and CHV, in English) after parsing each affix. Based upon these results, we are motivated to have the system perform a WEV lookup at this stage as well.

One challenge with SubSimplify is to maximize the understandability of the explanations themselves. While the existing resources contain a single definition of the term, SubSimplify relies upon combining multiple definitions to sew together a single explanation. Currently, the system employs a color-coding scheme that relates the morphological units on one line to their definitions on another. This may make it difficult to read for some people, effectively lowering the quality when people are not used to seeing these definitions. In order to allay this issue, our team plans on implementing a few different formats and testing them out in the near future in an interactive Web program.

Limitations

SubSimplify is naturally limited by two factors. First, not all difficult medical terms contain subword units, and additionally, not all subword units match or are totally accurate. This is because not all terms contain morphological units, and the system has no way of knowing where the characters within a word are actual examples of suffixes. Beyond that, even if they are correct matches of affixes, there is no guarantee that the actual meaning of the terms is directly reflective of the affixes that are found within. For example, it may provide a meaning for *anti-* and *-bodies* in *antibodies*. But the explanation *against bodies* does not reflect the actual meaning of the term.

Another possibility is that many of the explanations generated by SubSimplify can be incomplete. For example, it may provide a meaning for *-al* in *distal* but no meaning for *dist-*. The result would then be difficult for anyone to understand. Nevertheless, the system does provide a bridge between hand annotated and automatic texts and, therefore, should be subject to these sorts of exceptions and problematic cases.

Second, SubSimplify, by virtue of using WordNet and CHV, is limited as well to the coverage of those resources. However, we believe that this work presents useful addition to a system aimed at providing explanations for complex terms.

Conclusions

The niche of SubSimplify is to exploit the regularities of morphological units in medical terminology to provide a window into breaking down the jargon of difficult terms into digestible terms. SubSimplify will improve as the resources used to create it do. Furthermore, we want to look at multiword phrases as oftentimes they reveal the contextual meaning that a single-word context cannot provide alone. This approach is intended as an additional resource that one can add to other methods to automatically provide explanations for difficult texts. The explanations generated by this system greatly increase the number of difficult terms for which an easier alternative can be made available and, thereby, present an advance in the area of text simplification in the medical domain.

Acknowledgments

Research reported in this paper was supported by the National Library of Medicine of the National Institutes of Health under Award Number R01LM011975. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

References

1. Fox S. Pew Research Center. 2011 Feb 01. Health Topics URL: http://www.pewinternet.org/pdfs/PIP_Health_Report_July_2003.pdf [accessed 2006-08-10] [WebCite Cache ID 5I2STSU61]
2. Davis TC, Mayeaux EJ, Fredrickson D, Bocchini JA, Jackson RH, Murphy PW. Reading ability of parents compared with reading level of pediatric patient education materials. Pediatrics 1994 Mar;93(3):460-468. [Medline: 8115206]
3. Kindig D, Panzer A, Nielsen-Bohlman L. Health literacy: a prescription to end confusion. Washington, D.C: National Academies Press; 2004:a.
4. 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]
5. Shardlow M. A survey of automated text simplification. International Journal of Advanced Computer Science and Applications 2014;4(1):58-70.
6. Blum S, Levenston E. Universals of lexical simplification. Language learning 1978 Dec 01;28(2):399-415.
7. Leroy G, Endicott JE, Kauchak D, Mouradi O, Just M. User evaluation of the effects of a text simplification algorithm using term familiarity on perception, understanding, learning, and information retention. J Med Internet Res 2013 Jul 31;15(7):e144 [FREE Full text] [doi: 10.2196/jmir.2569] [Medline: 23903235]

XSL•FO
RenderX

8. Feblowitz D, Kauchak D. Sentence simplification as tree transduction. 2013 Aug 08 Presented at: The Second Workshop on Predicting and Improving Text Readability for Target Reader Populations; 2013; Sofia p. 1-10.
9. Pitler E, Nenkova A. Revisiting readability: A unified framework for predicting text quality. 2008 Oct 25 Presented at: Conference on Empirical Methods in Natural Language Processing; October; University of Edinburgh p. 186-195.
10. Brants T, Franz A. Web 1T 5-gram, 10 European Languages Version 1 LDC2009T25. In: WebT 5-gram Version. Philadelphia: Linguistic Data Consortium; 2009.
11. Leroy G, Kauchak D. The effect of word familiarity on actual and perceived text difficulty. J Am Med Inform Assoc 2014 Feb;21(e1):e169-e172 [FREE Full text] [doi: 10.1136/amiajnl-2013-002172] [Medline: 24100710]
12. Kandula S, Curtis D, Zeng-Treitler Q. A semantic and syntactic text simplification tool for health content. AMIA Annu Symp Proc 2010 Nov 13;2010:366-370 [FREE Full text] [Medline: 21347002]
13. Zeng QT, Tse T, Divita G, Keselman A, Crowell J, Browne AC. Exploring lexical forms: first-generation consumer health vocabularies. AMIA Annu Symp Proc 2006:1155 [FREE Full text] [Medline: 17238774]
14. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004 Jan 01;32(Database issue):D267-D270 [FREE Full text] [doi: 10.1093/nar/gkh061] [Medline: 14681409]
15. Zielstorff RD. Controlled vocabularies for consumer health. J Biomed Inform 2003;36(4-5):326-333 [FREE Full text] [Medline: 14643728]
16. Zeng QT, Tse T, Crowell J, Divita G, Roth L, Browne AC. Identifying consumer-friendly display (CFD) names for health concepts. AMIA Annu Symp Proc 2005:859-863 [FREE Full text] [Medline: 16779162]
17. Zeng-Treitler Q, Goryachev S, Kim H, Keselman A, Rosendale D. Making texts in electronic health records comprehensible to consumers: a prototype translator. AMIA Annu Symp Proc 2007 Oct 11:846-850 [FREE Full text] [Medline: 18693956]
18. Qenam B, Kim TY, Carroll MJ, Hogarth M. Text Simplification Using Consumer Health Vocabulary to Generate Patient-Centered Radiology Reporting: Translation and Evaluation. J Med Internet Res 2017 Dec 18;19(12):e417 [FREE Full text] [doi: 10.2196/jmir.8536] [Medline: 29254915]
19. Fellbaum C, Miller G. A semantic network of English verbs. In: WordNet: An Electronic Lexical Database. Cambridge: MIT Press; 1998.
20. Saggion H, Bott S, Rello L. Comparing resources for Spanish lexical simplification. In: Statistical Language and Speech Processing: First International Conference, SLSP 2013, Tarragona, Spain, July 29-31, 2013, Proceedings (Lecture Notes in Computer Science). Heidelberg Berlin: Springer; 2013:236-247.
21. Shardlow M. A Comparison of Techniques to Automatically Identify Complex Words. 2013 Aug 04 Presented at: In51st Annual Meeting of the Association for Computational Linguistics Proceedings of the Student Research Workshop; 2013; Sofia p. 103-109.
22. Thomas S, Anderson S. WordNet-based lexical simplification of a document. 2012 Sep 19 Presented at: 11th Conference on Natural Language Processing (KOVENS); 2012; Vienna p. 80-88.
23. Shehu A, Plaku E. A Survey of Computational Treatments of Biomolecules by Robotics-Inspired Methods Modeling Equilibrium Structure and Dynamic. jair 2016 Nov 29;57:509-572. [doi: 10.1613/jair.5040]
24. Paetzold G. White Rose eTheses Online.: University of Sheffield Lexical Simplification for Non-Native English Speakers. [Doctoral dissertation] URL: http://etheses.whiterose.ac.uk/15332/1/Final_Version_Thesis.pdf [accessed 2018-04-12] [WebCite Cache ID 6ydMFhua8]
25. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. 2013 Presented at: Advances in neural information processing systems; 2013; Lake Tahoe p. 3111-3119.
26. Glavaš G, Štajner S. Simplifying Lexical Simplification: Do We Need Simplified Corpora? 2015 Presented at: The Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing Short Papers Vol. 2; 2015; Beijing p. 63-68.
27. Cardellino C. Spanish Billion Words Corpus and Embeddings. 2016. URL: http://crscardellino.me/SBWCE/ [accessed 2018-04-12] [WebCite Cache ID 6ydMMtUOT]
28. Tindall A. Medical Terms. In: Medical Terms: Their Roots and Origins. Boca Raton: CRC Press; 1997.
29. Green T. The Greek & Latin Roots of English. Lanham: Rowman & Littlefield; Jul 17, 2014.
30. Wikipedia, The Free Encyclopedia. List of medical roots, suffixes and prefixes URL: https://en.wikipedia.org/wiki/List_of_medical_roots,_suffixes_and_prefixes [accessed 2018-04-12] [WebCite Cache ID 6ydMu8G7g]
31. McAuley D. Global RXPh: The Clinicians Ultimate Reference. 2018 Apr 26. Medical TERMINOLOGY-A thru Z URL: http://www.globalrph.com/medterm.htm [accessed 2018-04-12] [WebCite Cache ID 6ydMsGnXB]
32. Gramáticas. Prefijos y Sufijos Médicos URL: http://www.gramaticas.net/2013/10/prefijos-y-sufijos-medicos.html [accessed 2018-04-12] [WebCite Cache ID 6ydMnxHAa]
33. Salazar M. Enfermería: Blog dedicado a difundir el conocimiento de la Ciencia y Arte de la Enfermería y la Educación. 2013 Feb 26. TERMINOLOGÍA MÉDICA, SUFIJOS, PREFIJOS Y ABREVIATURAS URL: http://msalazar081858.blogspot.com/2013/02/terminologia-medica-sufijos-prefijos-y.html [accessed 2018-04-12] [WebCite Cache ID 6ydMxm4bu]
34. Willett P. The Porter stemming algorithm: then and now. Program 2006 Jul 1;40(3):219-223 [FREE Full text] [doi: 10.1108/00330330610681295]

XSL•FO
RenderX

35. Miller N, Lacroix EM, Backus JE. MEDLINEplus: building and maintaining the National Library of Medicine's consumer health Web service. Bull Med Libr Assoc 2000 Jan;88(1):11-17 [FREE Full text] [Medline: 10658959]
36. Leroy G, Endicott JE. Term Familiarity to Indicate Perceived and Actual Difficulty of Text in Medical Digital Libraries. In: Xing C, Crestani F, Rauber A, editors. Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation. Berlin, Heidelberg: Springer; Oct 2011:307-310.

Abbreviations

ANOVA: analysis of variance
CHV: Consumer Health Vocabulary
NLP: Natural Language Processing
SE: subject expert
UMLS: Unified Medical Language System
WEV: Word Embedding Vector

Edited by G Eysenbach; submitted 13.04.18; peer-reviewed by B Qenam, A Mavragani; comments to author 23.05.18; revised version received 13.06.18; accepted 16.06.18; published 02.08.18.

Please cite as:
Kloehn N, Leroy G, Kauchak D, Gu Y, Colina S, Yuan NP, Revere D
Improving Consumer Understanding of Medical Text: Development and Validation of a New SubSimplify Algorithm to Automatically Generate Term Explanations in English and Spanish
J Med Internet Res 2018;20(8):e10779
URL: http://www.jmir.org/2018/8/e10779/
doi:10.2196/10779
PMID:30072361

©Nicholas Kloehn, Gondy Leroy, David Kauchak, Yang Gu, Sonia Colina, Nicole P Yuan, Debra Revere. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 02.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Abstract Animations for the Communication and Assessment of Pain in Adults: Cross-Sectional Feasibility Study

Charles R Jonassaint[1], MHS, PhD; Nema Rao[2], MDes; Alex Sciuto[2], MDes; Galen E Switzer[3,4,5,6], PhD; Laura De Castro[7], MHSc, MD; Gregory J Kato[7], MD; Jude C Jonassaint[7], RN; Zakia Hammal[8], PhD; Nirmish Shah[9], MD; Ajay Wasan[10], MSc, MD

[1]Center for Behavioral Health and Smart Technology, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, United States
[2]Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA, United States
[3]Department of Medicine, University of Pittsburgh, Pittsburgh, PA, United States
[4]Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, United States
[5]Department of Clinical and Translational Science, University of Pittsburgh, Pittsburgh, PA, United States
[6]Center for Health Equity Research and Promotion, Veterans Affairs Pittsburgh Healthcare System, Pittsburgh, PA, United States
[7]Clinical Protocol and Data Management, University of Pittsburgh Medical Center, Pittsburgh, PA, United States
[8]The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, United States
[9]Division of Hematology, Department of Medicine, Duke University, Durham, NC, United States
[10]Department of Anesthesiology, University of Pittsburgh, Pittsburgh, PA, United States

Corresponding Author:
Charles R Jonassaint, MHS, PhD
Center for Behavioral Health and Smart Technology
Department of Medicine
University of Pittsburgh
Suite 600
230 McKee Place
Pittsburgh, PA, 15213
United States
Phone: 1 412 586 9850
Fax: 1 412 692 4838
Email: jonassaintcr@upmc.edu

Abstract

Background: Pain is the most common physical symptom requiring medical care, yet the current methods for assessing pain are sorely inadequate. Pain assessment tools can be either too simplistic or take too long to complete to be useful for point-of-care diagnosis and treatment.

Objective: The aim was to develop and test Painimation, a novel tool that uses graphic visualizations and animations instead of words or numeric scales to assess pain quality, intensity, and course. This study examines the utility of abstract animations as a measure of pain.

Methods: Painimation was evaluated in a chronic pain medicine clinic. Eligible patients were receiving treatment for pain and reported pain more days than not for at least 3 months. Using a tablet computer, participating patients completed the Painimation instrument, the McGill Pain Questionnaire (MPQ), and the PainDETECT questionnaire for neuropathic symptoms.

Results: Participants (N=170) completed Painimation and indicated it was useful for describing their pain (mean 4.1, SE 0.1 out of 5 on a usefulness scale), and 130 of 162 participants (80.2%) agreed or strongly agreed that they would use Painimation to communicate with their providers. Animations selected corresponded with pain adjectives endorsed on the MPQ. Further, selection of the electrifying animation was associated with self-reported neuropathic pain (r=.16, P=.03), similar to the association between neuropathic pain and PainDETECT (r=.17, P=.03). Painimation was associated with PainDETECT (r=.35, P<.001).

Conclusions: Using animations may be a faster and more patient-centered method for assessing pain and is not limited by age, literacy level, or language; however, more data are needed to assess the validity of this approach. To establish the validity of using abstract animations (“painimations”) for communicating and assessing pain, apps and other digital tools using painimations will need to be tested longitudinally across a larger pain population and also within specific, more homogenous pain conditions.

XSL•FO
RenderX

(J Med Internet Res 2018;20(8):e10056) doi:10.2196/10056

KEYWORDS

pain; pain measurement; chronic pain; medical informatics; mobile apps

Introduction

At least 116 million adults in the United States are affected by chronic pain [1]; that is, pain lasting for more than 3 months [2]. Medical treatment and lost productivity due to pain costs approximately US $635 billion each year, more than the cost of treating cardiovascular disease, cancer, or diabetes [3]. More than 80% of patients presenting for a physician visit report pain as a primary complaint, and a further 80% of these patients receive inadequate treatment for their pain [4]. Accurate assessment and diagnosis of pain is necessary to provide appropriate pain treatment and quality care [5,6].

Despite the development of many validated pain assessment scales, there has been little advancement in pain assessment since the introduction of the McGill Pain Questionnaire (MPQ) in 1970 [7] and the FACES pain scale for pediatric patients in the 1980s [8]. The traditional approach of self-report scales reduces the complexity of pain to a number or to unidimensional statements of pain [9]. Even the most recently developed pain scales or apps still rely on numerical scales and pain adjectives to assess pain, despite evidence that patients struggle with expressing pain to clinicians on static forms that ask them to estimate their pain on a 0 to 10 scale [10]. This oversimplification of pain ignores the sometimes intermittent nature and moment-to-moment variation in key features of the pain experience. Thus, medical providers miss opportunities to relate to their patients and may miss important symptoms and related diagnoses, leading them to intervene on poorly described and ill-defined targets [11,12].

In addition, current pain assessment approaches may perpetuate disparities in health care. Overly complex pain measures that rely on a long list of adjectives may alienate people with low literacy, those with disabilities, seniors with dementia, and many others with communication limitations [13-15]. Racial or ethnic differences in pain perception and expression may not be accurately captured by simplified pain scales [16]. Because the pain report is almost entirely subjective, even white race patients without language limitations can have their pain needs misinterpreted, their symptoms ignored, or their credibility challenged [15].

Advances in technology have made it possible to improve tools that measure patient-reported outcomes and, in turn, allow for a higher quality of data capture [17]. However, the pain assessment scales now being offered in electronic formats are essentially copies of the paper-pencil scales and do not capitalize on the flexibility of electronic media [18]. The introduction of novel, patient-centric pain assessment tools that leverage technology will maximize health care providers' ability to diagnose and treat pain.

This study tested the feasibility and utility of an innovative pain assessment tool that uses graphical illustrations and abstract animations ("painimations") to measure pain quality and intensity. We proposed that using animations to assess pain would overcome the barriers of traditional pain scales, allow patients to more accurately communicate the pain experience, and potentially facilitate pain diagnosis and treatment. Initial development and testing of this concept focused on neuropathic pain because of the well-defined characteristics that differentiate it from other pain types, the high prevalence of neuropathy in chronic pain populations [19], and the availability of validated neuropathic pain scales for comparison [20,21]. This paper describes the design process behind a novel, animation-based pain assessment app called Painimation, as well as the performance characteristics and capabilities of using this app to measure pain and to detect any neuropathic pain component.

Methods

Conceptualization and Development of Painimation

The initial goal of this instrument development effort was to improve our understanding of the patient pain experience and address limitations of current pain assessment and treatment. The first step in this approach was to interview patients who experienced acute and chronic pain. These interviews used principles of contextual inquiry, a common method in the development of interactive applications, to quickly uncover users' perceived needs [22]. In the first set of interviews, 10 patients were asked to recount both successful and unsuccessful encounters with clinicians regarding their pain, using a directed storytelling approach, an ethnographic research method where the participant discusses their past experience, with probes from the interviewer to elicit more detail on the underlying motivations and breakdowns [23]. The most prominent message from both chronic and acute pain patients was that "pain is so hard to describe and explain" and the "exact feeling is impossible to communicate." In the second part of the interviews, participants were given the Brief Pain Inventory [24] to complete in a "think-aloud" protocol [25] where they completed the questionnaire while stating aloud what they were thinking as they worked through each item. Both chronic and acute pain patients expressed confusion around the Brief Pain Inventory 0-10 scale and found the experience of using it "vague" and "ambiguous." Participants were also confused because the concept of "worst possible pain" is different for each person. Participants said, "I have no clue what [10 out of 10 pain] actually means." Additionally, participants said that the scale does not allow for the varied experience of pain; for instance, one might want to say, "It's an 8 when applying pressure, 7 when resting, and 10 early in the morning;" they described this limitation as "a lack of specificity."

A thematic analysis of these interviews indicated that a new pain assessment would need to both accommodate the vague, inexact feelings people often have regarding their pain and enable people to indicate different types of pain at the same time. Taking into account pain literature that suggests pain is more accurately depicted visually [26] and is better

XSL•FO
RenderX

communicated with word pictures, analogies, and metaphors [27,28], we decided to use a highly visual, abstract, and expressive mode of pain communication: animations.

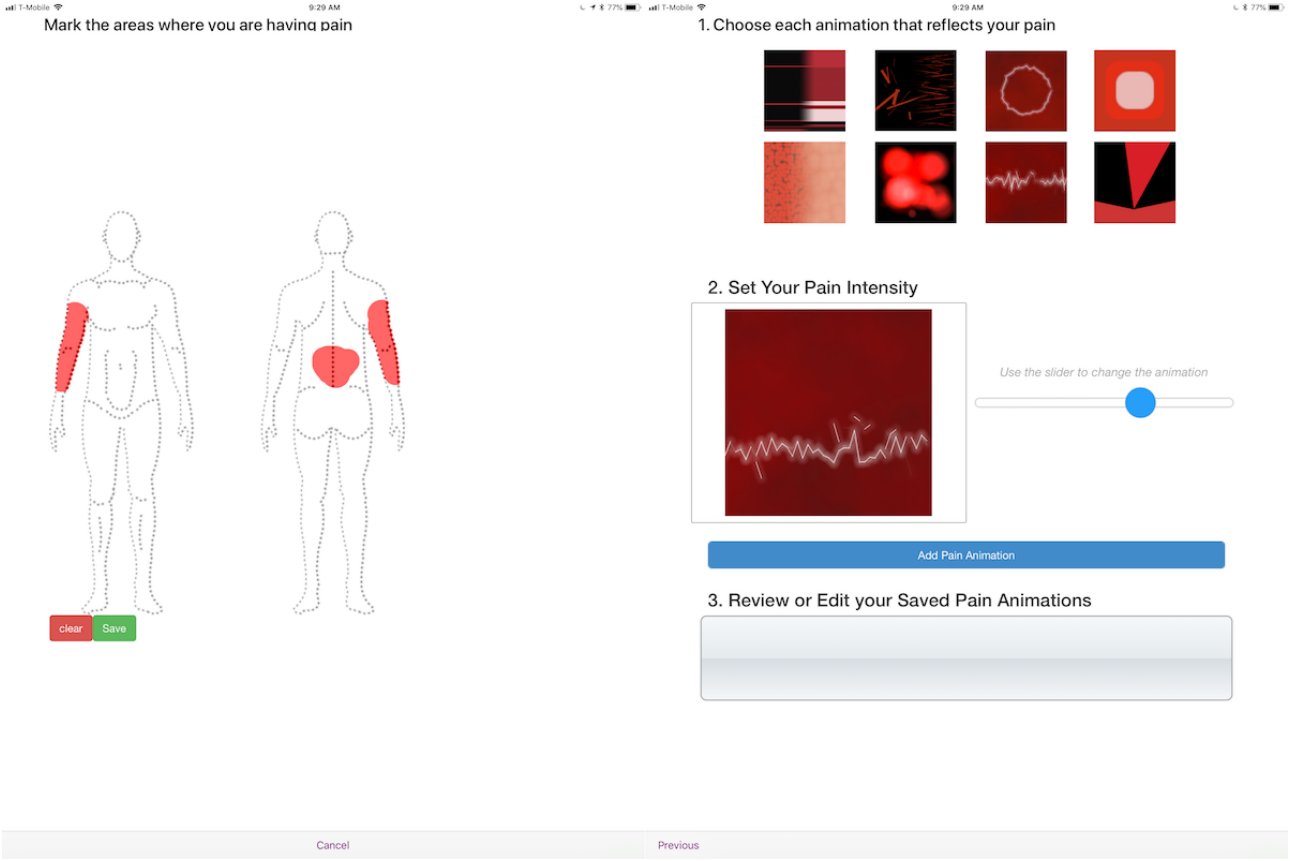
To develop the animations, the investigators started out with a list of words used to describe qualities of pain and reduced this list to several groups. We created animations to represent each group of sensations. Initial animations were tested with a group of 16 participants (see Rao [10] for more description). After some revision and another iteration of design, a final set of eight animations was created and selected for testing. To simplify the reporting of results and identification of animations, the animations were loosely labeled as "pounding," "shooting," "throbbing," "tingling," "cramping," "burning," "stabbing," and "electrifying."

We next developed an iOS app called Painimation that allowed users to select one or more of the eight animations previously listed. Development of Painimation involved an iterative process with three phases. In Phase 1, a group of six patients and family members were shown a demonstration of the initial version (v1.0) of the app and contributed input that was used to revise the app. In Phase 2, a group of five different participants who suffered from chronic pain were given a tablet device with the revised version of the Painimation app (v1.1) for testing. They were asked to enter their current pain symptoms and typical pain over the past 2 weeks. Data from this round of user testing was used to refine the app again. Finally, in Phase 3, all participants from Phase 2 were approached and asked to enter data using the close-to-final version of the Painimation app (v1.2) to confirm that all their concerns expressed in Phase 2 testing had been addressed. Any additional concerns raised during the Phase 3 testing were addressed in the finalized version of the app (v1.3).

App Description

Painimation first shows users a body image and asks them to mark the areas where they are having pain. If they make a mistake, they are able to clear their markings and start again. Once they are satisfied with the selection of pain locations, they save this image and advance to the next screen where they are shown a selection of eight animations, which they then use to describe the quality of their pain (Figure 1; Multimedia Appendix 1). Once the user selects an animation, he or she is asked to indicate the intensity of their pain by using a slider to change the animation "intensity." Moving the slider changes the animation intensity by increasing or decreasing its speed, color saturation, focus, and size. The user can manipulate the animation until they feel it most closely matches or reflects the quality and intensity of their pain experience. The position of the intensity slider (where the lowest position is 0 and the highest position or maximum is 100) is used as the pain intensity score for that specific animation. When the user is satisfied with their selected and customized pain animation, they add the animation to a bin, at which point they can either select another animation or proceed to the next screen. The app allows up to five animations to be added to the bin. Participants can review each of their chosen animations at the final screen and are then presented with the patient satisfaction questionnaire. The backend of the app provides both the qualitative data (pain location and type) and quantitative data (pain intensity and percentage of body covered in pain) on pain symptoms.

Figure 1. Screenshots of the Painimation app that uses images and animations to assess pain location, quality, and intensity.



XSL•FO
RenderX

Proof-of-Concept Study Methods

Study Population

We tested Painimation in the general pain medicine clinic at the University of Pittsburgh Medical Center (Pittsburgh, PA) in adult patients (≥18 years of age) currently receiving treatment for chronic pain. In order to be eligible, patients had to have experienced pain more days than not for at least three consecutive months including the day of study participation. At clinic check-in, eligible patients were given a postcard by the clinic receptionist with a brief description of the study and instructions to approach one of the research assistants in the waiting room if interested. To maintain patient privacy, we did not obtain data on nonparticipants. Those patients who were interested received a tablet computer from a research assistant; the tablet guided the potential participant through the electronic consent process.

Electronic Consent Process and Data Capture

Using the Apple iOS ResearchKit framework facilitated the electronic consent process and completion of study questionnaires on an Apple iPad. Patients who elected to participate in the study and signed the electronic consent form were presented with a brief battery of electronic questionnaires that they were asked to complete while in the waiting room. All data, including a portable digital file of the signed consent, was uploaded directly from the ResearchKit app to a REDCap database [29].

Measures

Participants first completed a basic demographics form and a brief clinical history of their pain, in which they self-reported whether or not they currently had pain, the severity of current pain on a visual analog scale (VAS), the duration of pain condition, and the type of pain condition, choosing from the following: nerve damage, arthritis, sickle cell, fibromyalgia, back pain, neck pain, headache, migraine, joint pain, chronic pain, abdominal pain, or other.

Participants completed the MPQ [7]; the PainDETECT [20] questionnaire, a measure of neuropathic pain; and the experimental Painimation app. Each participant was randomized automatically to complete the Painimation scale first or last. The randomization of questionnaires helped determine if seeing the pain adjectives on the MPQ and/or the pain statements on PainDETECT influenced selection of animations or satisfaction with the measure.

The MPQ consists primarily of three major classes of word descriptors—sensory, affective, and evaluative—that are used by patients to specify subjective pain experience. These word descriptors from the MPQ were correlated with each animation in the Painimation app. Mapping the words participants used to describe their pain to which animation they chose gave a sense of how participants may have been interpreting the animations.

PainDETECT is a screening tool for neuropathic pain, with seven weighted sensory descriptor items and two items relating to the spatial (radiating) and temporal characteristics of the individual pain pattern. Compared to clinical diagnosis, its sensitivity is 85% and its specificity 80%. PainDETECT was initially developed and validated in patients with back pain but also has shown applicability to patients with other types of neuropathic pain. For screening purposes, cut-off scores of 12 or less (a neuropathic component is unlikely) and 19 or greater (a neuropathic component is likely) are recommended [20].

Painimation is a novel app, developed by our group, for pain assessment using graphical illustrations and abstract animations to measure pain location, quality, and intensity. See a more detailed explanation of the app in the “App Description” section earlier in Methods. Data for Painimation were stored both on a backend server and on the iPad device. We confirmed data in both locations for consistency.

Patient satisfaction was evaluated once the user completed entering their pain on Painimation. Users were asked if it (1) was useful for describing their pain, (2) enjoyable to use, and (3) would be useful for communicating their pain to the provider; for each item, users were asked to choose a response on a 1 to 5 Likert scale from 1=strongly disagree to 5=strongly agree. We hypothesized that using the Painimation app would be associated with high patient satisfaction regarding usability of the app and ability to communicate their pain experience.

Statistical Analysis

Distributions and Comparison by Demographics

Descriptive measures of central tendency and dispersion were used to examine distributions of pain scores. The three pain scale scores were evaluated for differences by age (above and below the median), sex, race/ethnicity, and location of pain to determine if there were differences in pain intensity and quality.

Discriminate Ability of Painimation

To examine how well Painimation could accurately discriminate between different pain types, the study compared patients’ animation selection to their pain diagnosis (eg, nerve damage, arthritis, or sickle cell disease), as well as differences in Painimation-recorded pain quality (ie, the painimation[s] they chose) by self-report pain type. The primary analysis for this study was the one-way comparison between neuropathic pain, self-report nerve damage, and nonneuropathic pain. We characterized the association between specific Painimation scores and pain diagnosis with means and Pearson correlations.

Next, to compare Painimation to more traditional pain measures (PainDETECT, the MPQ, and the VAS), we examined distributions of pain scores across all measures using descriptive measures of central tendency, and the associations between measures using Pearson correlation coefficients for continuous measures and phi correlation coefficients when comparing two dichotomous outcomes. The discriminate ability of Painimation versus PainDETECT measures was measured by first using chi-square analysis to compare the association of the PainDETECT-recommended cut-off score (scores ≤12 suggest neuropathic component is unlikely) with self-report nerve damage, versus selection of the electrifying animation with nerve damage.

To test sensitivity and specificity for detecting a neuropathic pain component for Painimation and PainDETECT we

XSL•FO
RenderX

calculated receiver operating characteristic (ROC) curves derived using logistic regression analyses, quantified by area under the curve (AUC). To accomplish this for Painimation, the analysis took into account animation intensity (speed/saturation) and transformed the electrifying animation into a continuous 0 to 100 score by modeling nonselection of electrifying as "0" and, for those who selected "electrifying," using the intensity value they selected for that animation. Using the continuous PainDETECT score and "electrifying" animation score (ie, participant-selected intensity of the animation), we tested confidence intervals of the two AUCs to determine whether the confidence intervals for AUC overlapped between the two measures. We performed linear regression with the response variable being the pain type and the type of measure (Painimation or PainDETECT) as the independent variable.

We also examined the correlation between the electrifying animation and specific qualitative descriptors on the PainDETECT scale by assessing the association of the animation with each questionnaire item. The PainDETECT questionnaire uses an 11-point ordinal scale for intensity of each qualitative descriptor. For each number, we calculated the Painimation intensity mean with 95% confidence intervals, and the median with interquartile ranges. We tested whether 1-point change in the PainDETECT questionnaire was associated with increased probability of a participant's selecting the "electrifying" animation.

Results

Characteristics of Study Sample

The study obtained consent from 202 participants. We removed from analyses 13 participants who were missing data on more than one of the full questionnaires. The excluded sample did not differ from the other participants on any variables of interest. Reasons for missingness were random (ie, ran out of time, being called to clinic room for their physician visit, or poor Internet connection). The analyzed sample (N=189) had a mean age of 51.55 (SD 13.86) years, with 124 of 189 (65.6%) reporting female gender, 123 of 189 (65.1%) reporting white race, and 45 of 189 (23.8%) reporting black or African American race. Of the 189 participants, 66 (34.9%) participants had a college degree or higher, whereas only 12 (6.3%) had less than a high school/GED education. A majority of patients presented with back pain (123/189, 65.1%), chronic pain syndrome (108/189, 57.1%), arthritis (78/189, 41.3%), and nerve damage (67/189, 35.4%).

Due to data errors specific to the Painimation app, we lost study data for n=19 participants. These participants did not differ on any variables of interest from the n=170 participants with complete Painimation data. For the individuals with complete data (n=170), the animations most frequently selected were electrifying (66/170, 38.8%), throbbing (54/170, 31.8%), cramping (51/170, 30.0%), burning (51/170, 30.0%), shooting (47/170, 27.6%), and stabbing (35/170, 20.6%). The tingling (10/170, 5.9%) and pounding (21/170, 12.3%) animations were selected least frequently.

Associations of Painimation Selections With Clinical Features

The animations selected by pain patients showed several associations with clinical features of the patients. Mean current pain level on the VAS was 6.8 (SE 0.2, range 0-10). Those who chose the "electrifying" animation had a significantly higher VAS pain level (mean 7.2, SE 0.2) than those who did not (mean 6.3, SE 0.3; P=.02); those who chose the "stabbing" animation had a marginally higher VAS pain level (mean 7.2, SE 0.2) than those who did not (mean 6.6, SE 0.2; P=.15), but this was not statistically significant. There were no differences on the VAS pain scale for any other animations.

Several animations were associated with different types of pain (Table 1). Neuropathic pain was associated with the "electrifying" animation and marginally associated with the "shooting" animation. Fibromyalgia was associated with the "pounding" and "tingling" animations. Headache-type pain was associated with the "pounding" and "tingling" animations and marginally with the "electricity" animation. Abdominal pain was associated with the "pounding" animation. There were no other notable associations. The PainDETECT score was associated with all pain types except abdominal pain and "other" pain types. PainDETECT score showed the strongest correlation with the "electrifying" and "shooting" animations, and was also correlated to a lesser extent with the "pounding" and "stabbing" animations.

Tables 2 and 3 show associations between pain diagnosis, the "electrifying" animation, and the dichotomized PainDETECT score. The "electrifying" animation was associated with nerve damage and marginally associated with general headache. The dichotomized PainDETECT score was not associated with nerve damage, but was associated with fibromyalgia, back pain, neck pain, headache pain, and chronic pain syndrome.

Each painimation that was associated with specific MPQ pain quality descriptors by at least 10 participants is presented in Table 4. The "electrifying" painimation was associated with the MPQ descriptors sharp, burning, and tingling. The "pounding" painimation was associated with stabbing, cramping, and sore. "Shooting" was associated with pulsing and sharp; "cramping" with stabbing; "throbbing" with pulsing, sore, and hurting; "tingling" with throbbing; "burning" with cramping and sore; and "stabbing" with pulsing, stabbing, and burning. PainDETECT score was correlated with each of the top selected MPQ descriptors except throbbing, sore, hurting, and aching.

XSL•FO
RenderX

Table 1. Correlation matrix showing associations between self-report pain diagnosis, Painimation, and PainDETECT total score (n=170).

Self-report diagnosis	Expressive Painimation animations, *r*								PainDETECT score, *r*
	Electrifying	Pounding	Shooting	Cramping	Throbbing	Tingling	Burning	Stabbing	
Nerve damage	.159[a]	.017	.141	.019	–.010	.021	.046	.135	.165[a]
Arthritis	.084	.045	–.017	–.008	–.040	–.009	–.008	–.077	.149[a]
Fibromyalgia	.054	.170[a]	–.013	.026	.133	.188[a]	.026	.109	.283[c]
Back pain	.083	–.059	–.011	.027	–.051	–.025	–.027	.072	.208[b]
Neck pain	–.047	–.012	.083	–.036	–.033	.000	.104	–.016	.293[c]
Headache	.138	.154[a]	.093	.067	.016	.212[b]	–.067	.070	.274[c]
Migraine	.120	.067	.069	.078	–.007	.086	.010	.001	.249[c]
Joint pain	.069	.049	.017	–.052	–.006	.045	.026	.047	.192[a]
Chronic pain	.112	.100	.150	.033	–.012	.009	.086	.048	.336[c]
Abdominal pain	–.054	.206[b]	–.133	.051	.058	.093	.051	–.049	.079
Other	–.055	.068	.036	.015	.038	.127	.130	–.066	–.025
PainDETECT score	.353[c]	.160[a]	.259[c]	.025	.004	.015	.007	.187[a]	—

[a]*P*<.05.

[b]*P*<.01.

[c]*P*<.001.

Table 2. Association between pain diagnosis and selection of the “electrifying” animation among adults with chronic pain (n=170).

Diagnosis	Selected electrifying animation, n (%)			*P* value
	No	Yes	Total	
Nerve damage	31 (29.8)	30 (45.5)	61 (35.9)	.04
Arthritis	40 (38.5)	31 (47.0)	71 (41.8)	.27
Sickle cell	4 (3.8)	1 (1.5)	5 (2.9)	.38
Fibromyalgia	19 (18.3)	15 (22.7)	34 (20.0)	.49
Back pain	64 (61.5)	46 (69.7)	110 (64.7)	.28
Neck pain	33 (31.7)	18 (27.3)	51 (30.0)	.54
General headache	14 (13.5)	16 (24.2)	30 (17.6)	.07
Migraine headache	14 (13.5)	15 (22.7)	29 (17.1)	.12
Joint pain	40 (38.5)	30 (45.5)	70 (41.2)	.37
Chronic pain syndrome	56 (53.8)	43 (65.2)	99 (58.2)	.15
Abdominal pain	27 (26.0)	14 (21.2)	41 (24.1)	.48
Other	15 (14.4)	7 (10.6)	22 (12.9)	.47

XSL•FO
RenderX

Table 3. Association between pain diagnosis and PainDETECT score among adults with chronic pain (N=186).

Diagnosis	PainDETECT score[a], n (%)			*P* value
	Low	High	Total	
Nerve damage	28 (31.1)	39 (39.4)	67 (35.4)	.23
Arthritis	33 (36.7)	45 (45.5)	78 (41.3)	.22
Sickle cell	1 (1.1)	4 (4.0)	5 (2.6)	.21
Fibromyalgia	7 (7.8)	29 (29.3)	36 (19.0)	<.01
Back pain	51 (56.7)	72 (72.7)	123 (65.1)	.02
Neck pain	17 (18.9)	40 (40.4)	57 (30.2)	.01
General headache	7 (7.8)	26 (26.3)	33 (17.5)	.01
Migraine headache	8 (8.9)	25 (25.3)	33 (17.5)	.01
Joint pain	30 (33.3)	46 (46.5)	76 (40.2)	.07
Chronic pain syndrome	36 (40.0)	72 (72.7)	108 (57.1)	<.01
Abdominal pain	19 (21.1)	26 (26.3)	45 (23.8)	.41
Other	12 (13.3)	13 (13.1)	25 (13.2)	.97

[a]PainDETECT scores can range from 0 to 38. Scores of 0-12 suggest nociceptive pain or a neuropathic pain component is unlikely (<15% likelihood), scores of 13-18 suggest an unclear or ambiguous pain type, and scores of 19-38 suggests neuropathic pain component (>90% likelihood).

Table 4. Correlation matrix showing associations between McGill Pain Questionnaire (MPQ) descriptors, Painimation, and PainDETECT total score (n=170).

MPQ	Expressive Painimation animations, *r*								PainDETECT score, *r*
	Electrifying	Pounding	Shooting	Cramping	Throbbing	Tingling	Burning	Stabbing	
Pulsing	.063	.105	–.182[a]	.098	.236[b]	–.009	–.005	.166[a]	.237[b]
Throbbing	–.079	.035	.036	.040	.149	.163[a]	–.001	.097	.052
Pounding	–.041	.025	.001	.068	.015	.108	.022	.052	.197[b]
Shooting	.083	.124	.109	.023	.033	–.017	.021	.116	.288[c]
Stabbing	.066	.168[a]	.086	.208[b]	–.076	.064	–.061	.224[b]	.208[b]
Sharp	.216[b]	.097	.250[b]	.068	–.064	.040	–.020	.087	.275[c]
Cramping	–.107	.185[a]	–.012	–.044	.135	.070	.173[a]	.021	.189[a]
Burning	.254[c]	.007	.047	.146	–.020	–.022	.044	.169[a]	.375[c]
Tingling	.204[b]	–.006	.074	.042	–.100	.024	–.004	.144	.269[c]
Sore	.073	.162[a]	–.039	.063	.170[a]	.031	.153[a]	.075	.138
Hurting	–.002	.037	–.002	.009	.162[a]	–.052	–.065	.102	.082
Aching	.134	.065	.073	–.069	.071	–.032	.034	.086	.110

[a]*P*<.05.
[b]*P*<.01.
[c]*P*<.001.

Sensitivity and Specificity of Measures to Predict Neuropathic Pain Component

We performed a ROC analysis to determine the ability of Painimation to discriminate a neuropathic pain component in this population, then compared to PainDETECT, a measure already validated for identifying neuropathic pain. In this proof-of-concept study, Painimation simply used selection of the “electrifying” animation and the selected intensity (0-100) as the correlate for neuropathic pain component. The AUC, relating to the sensitivity and specificity, was low for both PainDETECT (AUC=0.60, 95% CI 0.52-0.69) and Painimation (AUC=0.59, 95% CI 0.51-0.67). The comparison of the AUC for the two measures showed no significant difference in their

XSL•FO
RenderX

ability to detect nerve damage in this population (χ^2_1=0.01, *P*=.90).

Evaluation of Patient Satisfaction

At completion, 141 of 162 participants (87.0%) agreed or strongly agreed that Painimation was useful for describing their pain, 137 of 162 (84.6%) agreed or strongly agreed that Painimation was enjoyable to use, and 130 of 162 (80.2%) agreed or strongly agreed they would use Painimation to communicate their pain to providers.

Discussion

Data Support Use of Animations For Communicating Pain

This study explored the use of abstract animations and graphical illustrations as a novel method for assessing pain quality, intensity, and location in adult patients with chronic pain. Although preliminary, these data suggest that a technology-based approach that allows patients to express their pain experience using animations that they can adjust and customize is acceptable, and potentially more efficient than traditional methods of pain assessment.

We believe using animations to measure pain can not only allow patients to describe pain sensations in a similar manner to how they experience them but, by not relying on words or numeric scales, can remove language and literacy as potential barriers to pain assessment. Further, given the well-documented gender, ethnic, and cultural differences in the experience and expression of pain [16,30], a more nuanced measure that eliminates linguistic and cultural biases may help highlight and elucidate group differences. Painimation has the potential to benefit both researchers and clinicians: for the former, it can improve understanding of gender and ethnic differences in pain and, for the latter, it can decrease the frequency of misunderstandings in pain reporting. To our knowledge, there are currently no other pain assessment tools that address pain location, intensity, and quality in a short assessment format that is not limited by language or literacy. Both the FACES scale and VAS, although short in length, are too simplistic and fail to identify the location and quality of pain. Longer multidimensional pain assessment measures such as the MPQ and PainDETECT are burdensome to complete and are not appropriate for all literacy levels. Other mobile device apps, such as Catch My Pain and My Pain Diary, have pain location, pain intensity, and symptoms tracking; however, these pain apps are not able to capture pain type or quality, and lack diagnostic potential.

Indeed, the use of abstract animations to measure pain is a truly novel approach that has not been previously described. Although significant associations in the expected direction were identified, these associations were not very strong, which may suggest patients were interpreting these animations in very different ways, potentially due to the diversity of pain types included in the sample. More data will be needed to understand how patients interpret the animations and the subsequent implications for pain assessment methodology.

Preliminary Evidence of Diagnostic Properties

Testing of Painimation showed that its “electrifying” and “shooting” pain animations were associated with patient-reported nerve damage. These associations were similar to the PainDETECT score with respect to neuropathic pain (*r*=.165). Further, there was no significant difference in specificity and sensitivity of the two measures in predicting nerve damage. The PainDETECT cut-off score was not associated with patient-reported neuropathic pain, whereas the “electrifying” animation was. This suggests that Painimation performed just as well as PainDETECT as a marker of neuropathic pain and has the added benefits of being a much shorter, more engaging, and less complex assessment tool. In fact, using Painimation to detect nerve damage required consideration of only one item, that being the selection or nonselection of “electrifying” pain, and ignored other data inputs such as location, intensity, and other animations selected. The AUC for both measures was modest, but these data provide a foundation for iterative development of Painimation and analytic approaches that may enhance the diagnostic properties of this assessment tool.

Clinical Implications

The American Pain Society guidelines recommend that a numeric pain intensity rating (0-10) be recorded at every clinical encounter. However, this widespread effort to assess pain with a unidimensional numeric scale did not improve the quality of pain treatment for patients [31]. In fact, this initiative may have contributed to the rise in opioid prescriptions [32]. The 0 to 10 scale has not improved pain assessment or treatment because the scale is unidimensional, uninformative, and lacks utility for both patients and their medical care providers [10]. Technologies such as Painimation can be easily implemented into the clinical setting and provide as much or more information than traditional multidimensional pain assessment tools. Such technology could help reestablish pain as the fifth vital sign, helping to realize the original American Pain Society vision for routine pain assessment in clinical care.

Painimation has the potential to allow patients to express their pain in a way not previously possible and improve their communication with providers. One older patient with shingles said about using Painimation, “This is the first time I’ve been able to describe my pain to someone.” Painimation could also help reduce the time it takes for a comprehensive pain assessment and diagnosis. Current pain assessments rely on long interviews or questionnaires that burden patients and can impede clinic flow. One report found that hospitalized participants with cancer took approximately 24 minutes to complete the paper version of the MPQ [33] (range 12-45 minutes). Painimation may provide the same level of data on pain in a shorter time frame. Unfortunately, the current study was unable to collect exact time-to-completion data for each questionnaire due to having administered the questionnaires in a clinic waiting room setting where unpredictable interruptions at times invalidated timestamp data. However, the available data suggests that participants completed Painimation, on average, in less than 2 minutes. Future studies will administer this set of questionnaires to a subset of the sample in an isolated

XSL•FO
RenderX

room to determine more accurate and valid time-to-completion data.

Study Limitations

Despite the interesting nature of the data, our study has several limitations. First, pain type was assessed via self-report, which may not be as accurate as a medical diagnosis. In future studies, to more accurately and definitively identify a neuropathic pain component, it would be helpful to include objective measures of nerve damage such as electromyogram and nerve conduction studies. However, these measures are imperfect given that electromyogram typically does not detect small fiber neuropathies [34], and nerve conductions studies can have false positive results [35]. Second, this study was cross-sectional and the stability of measurements was not assessed over time. Third, although we assessed participant satisfaction with Painimation, we did not collect this data for the other pain scales; therefore, we could not compare participants' satisfaction with Painimation to their satisfaction with the other pain scales. Finally, the pain population was mixed and may not have been the most appropriate for the first test of Painimation given that each pain disorder may be associated with a very unique pain experience. Our primary comparison measure, PainDETECT, also did not perform well in this population, which provides more evidence of population heterogeneity but also exposes the limitations of existing measures. Despite these shortcomings, however, study data show patterns in the expected direction and suggest that using animations to communicate and assess pain is feasible even in a mixed-pain population. Future studies of this technology may benefit from tailoring animations to specific pain populations.

In the current version of the Painimation app, the user identifies pain location(s) and then selects animations that reflect the quality and intensity of their pain. If multiple pain locations are selected, the app does not allow the user to identify which pain location the animation they selected references. The ability to apply animations to a specific body location will be a feature available on future versions of the Painimation app, but was not a feature of the app version tested in this study. In addition, a significant portion of study data was lost due to app errors. App complications were in part due to the challenge of delivering eight animations simultaneously. The study team worked to correct and prevent technical errors; however, future studies will benefit from rigorously testing the robustness of app functions and data transfer before full launch of the study.

Future Research Directions

This report provides an initial evaluation of the utility of animations and graphical illustrations to describe pain quality, type, and severity. Future studies will further evaluate both qualitative and quantitative data regarding patients' perceptions of Painimation and usability compared with other pain assessment tools. It will be important to increase the specificity and sensitivity of the tool by tailoring and testing it with specific pain types (eg, low back pain), in acute versus chronic pain, and with pain in specific diseases (eg, pancreatitis or sickle cell disease). Future longitudinal assessments will be able to test whether Painimation scores predict clinical outcomes such as likelihood of response to pain treatment and risk of rehospitalization. In addition, we will be testing a mobile-phone version of Painimation that allows patients to assess and report their pain daily.

Conclusions

We have set out to change the pain assessment conversation and forge a new direction for research on patient-reported outcomes. This research contributes a patient-centric, automated, pain assessment method that has the potential to not only improve patient-provider communication, but to also generate higher quality patient-reported pain symptoms data to guide diagnoses and treatment. By using animations to assess pain, we can decrease the burden of long, detailed pain assessments while collecting pertinent information on each patient's pain experience through an easy-to-administer, novel, and engaging medium. Further research with animation-based pain assessment tools is needed to improve this tool and to more definitively determine its validity and utility.

Acknowledgments

This project was supported by grant number K23 HL135396 (PI: CRJ) from the National Heart Lung and Blood Institute and the Virginia Kaufmann Foundation. ZH's effort was supported by the National Institute of Nursing Research of the National Institutes of Health under Award Number R21NR016510. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The attached mp4 file is uploaded to illustrate two of the painimations from the app.

[MP4 File (MP4 Video), 13MB - jmir_v20i8e10056_app1.mp4]

References

1. Simon L. Relieving pain in America: a blueprint for transforming prevention, care, education, and research. J Pain Palliat Care Pharmacotherapy 2012 Jul 05;26(2):197-198. [doi: 10.3109/15360288.2012.678473]

XSL•FO
RenderX

2. Treede R, Rief W, Barke A, Aziz Q, Bennett MI, Benoliel R, et al. A classification of chronic pain for ICD-11. Pain 2015 Jun;156(6):1003-1007 [FREE Full text] [doi: 10.1097/j.pain.0000000000000160] [Medline: 25844555]
3. Gaskin DJ, Richard P. The economic costs of pain in the United States. J Pain 2012 Aug;13(8):715-724. [doi: 10.1016/j.jpain.2012.03.009] [Medline: 22607834]
4. Kirsch B, Berdine H, Zablotsky D, Wenzel G, Meyer C. Implementation of a pain management strategy: identifying pain as the fifth vital sign. VHSJ 2000;5:48-57.
5. Paulson-Conger M, Leske J, Maidl C, Hanson A, Dziadulewicz L. Comparison of two pain assessment tools in nonverbal critical care patients. Pain Manag Nurs 2011 Dec;12(4):218-224. [doi: 10.1016/j.pmn.2010.05.008] [Medline: 22117753]
6. Rahu MA, Grap MJ, Ferguson P, Joseph P, Sherman S, Elswick RK. Validity and sensitivity of 6 pain scales in critically ill, intubated adults. Am J Crit Care 2015 Nov;24(6):514-523 [FREE Full text] [doi: 10.4037/ajcc2015832] [Medline: 26523009]
7. Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. Pain 1975 Sep;1(3):277-299. [Medline: 1235985]
8. Wong DL, Baker CM. Pain in children: comparison of assessment scales. Pediatr Nurs 1988;14(1):9-17. [Medline: 3344163]
9. Stinson JN, Kavanagh T, Yamada J, Gill N, Stevens B. Systematic review of the psychometric properties, interpretability and feasibility of self-report pain intensity measures for use in clinical trials in children and adolescents. Pain 2006 Nov;125(1-2):143-157. [doi: 10.1016/j.pain.2006.05.006] [Medline: 16777328]
10. Rao N. Redesigning the Pain Assessment Conversation [master's thesis]. Pittsburgh, PA: Carnegie Mellon University; 2015.
11. Clark WC, Yang JC, Tsui SL, Ng K, Bennett CS. Unidimensional pain rating scales: a multidimensional affect and pain survey (MAPS) analysis of what they really measure. Pain 2002 Aug;98(3):241-247. [Medline: 12127025]
12. de C Williams AC, Davies HT, Chadury Y. Simple pain rating scales hide complex idiosyncratic meanings. Pain 2000 Apr;85(3):457-463. [Medline: 10781919]
13. Herr K, Coyne PJ, McCaffery M, Manworren R, Merkel S. Pain assessment in the patient unable to self-report: position statement with clinical practice recommendations. Pain Manag Nurs 2011 Dec;12(4):230-250. [doi: 10.1016/j.pmn.2011.10.002] [Medline: 22117755]
14. Zwakhalen SM, Hamers JP, Abu-Saad HH, Berger MP. Pain in elderly people with severe dementia: a systematic review of behavioural pain assessment tools. BMC Geriatr 2006 Jan 27;6:3 [FREE Full text] [doi: 10.1186/1471-2318-6-3] [Medline: 16441889]
15. Hadjistavropoulos T, Craig K. A theoretical framework for understanding self-report and observational measures of pain: a communications model. Behav Res Ther 2002 May;40(5):551-570. [doi: 10.1016/S0005-7967(01)00072-9]
16. Campbell CM, Edwards RR. Ethnic differences in pain and pain management. Pain Manag 2012 May;2(3):219-230 [FREE Full text] [doi: 10.2217/pmt.12.7] [Medline: 23687518]
17. Muehlhausen W, Doll H, Quadri N, Fordham B, O'Donohoe P, Dogar N, et al. Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. Health Qual Life Outcomes 2015 Oct 07;13:167 [FREE Full text] [doi: 10.1186/s12955-015-0362-x] [Medline: 26446159]
18. Wood C, von Baeyer CL, Falinower S, Moyse D, Annequin D, Legout V. Electronic and paper versions of a faces pain intensity scale: concordance and preference in hospitalized children. BMC Pediatr 2011 Oct 12;11:87. [doi: 10.1186/1471-2431-11-87] [Medline: 21989306]
19. Torrance N, Smith BH, Bennett MI, Lee AJ. The epidemiology of chronic pain of predominantly neuropathic origin. Results from a general population survey. J Pain 2006 Apr;7(4):281-289. [doi: 10.1016/j.jpain.2005.11.008] [Medline: 16618472]
20. Freynhagen R, Tölle TR, Gockel U, Baron R. The painDETECT project - far more than a screening tool on neuropathic pain. Curr Med Res Opin 2016 Dec;32(6):1033-1057. [doi: 10.1185/03007995.2016.1157460] [Medline: 26907456]
21. Padua L, Briani C, Truini A, Aprile I, Bouhassirà D, Cruccu G, et al. Consistence and discrepancy of neuropathic pain screening tools DN4 and ID-Pain. Neurol Sci 2013 Mar;34(3):373-377. [doi: 10.1007/s10072-012-1011-3] [Medline: 22434411]
22. Millen DR. Rapid ethnography: time deepening strategies for HCI field research. 2000 Presented at: 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques; Aug 17-19, 2000; New York p. 280-286. [doi: 10.1145/347642.347763]
23. Evenson S. Directed storytelling: interpreting experience for design. In: Bennett A, editor. Design Studies: Theory and Research in Graphic Design. New York: Princeton Architectural Press; 2006:231-240.
24. Tan G, Jensen MP, Thornby JI, Shanti BF. Validation of the Brief Pain Inventory for chronic nonmalignant pain. J Pain 2004 Mar;5(2):133-137. [doi: 10.1016/j.jpain.2003.12.005] [Medline: 15042521]
25. Ericsson K, Simon H. Verbal reports as data. In: Psychological Review. Cambridge, MA: MIT Press; 1980:215-251.
26. Schott GD. The cartography of pain: the evolving contribution of pain maps. Eur J Pain 2010 Sep;14(8):784-791. [doi: 10.1016/j.ejpain.2009.12.005] [Medline: 20079670]
27. Casarett D, Pickard A, Fishman JM, Alexander SC, Arnold RM, Pollak KI, et al. Can metaphors and analogies improve communication with seriously ill patients? J Palliat Med 2010 Mar;13(3):255-260 [FREE Full text] [doi: 10.1089/jpm.2009.0221] [Medline: 19922170]

XSL•FO
RenderX

28. Dudgeon BJ, Ehde DM, Cardenas DD, Engel JM, Hoffman AJ, Jensen MP. Describing pain with physical disability: narrative interviews and the McGill Pain Questionnaire. Arch Phys Med Rehabil 2005 Jan;86(1):109-115. [Medline: 15641000]
29. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
30. Greenspan JD, Craft RM, LeResche L, Arendt-Nielsen L, Berkley KJ, Fillingim RB, Consensus Working Group of the Sex, Gender, Pain SIG of the IASP. Studying sex and gender differences in pain and analgesia: a consensus report. Pain 2007 Nov;132 Suppl 1:S26-S45 [FREE Full text] [doi: 10.1016/j.pain.2007.10.014] [Medline: 17964077]
31. Mularski RA, White-Chu F, Overbay D, Miller L, Asch SM, Ganzini L. Measuring pain as the 5th vital sign does not improve quality of pain management. J Gen Intern Med 2006 Jun;21(6):607-612 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00415.x] [Medline: 16808744]
32. Morone NE, Weiner DK. Pain as the fifth vital sign: exposing the vital need for pain education. Clin Ther 2013 Nov;35(11):1728-1732 [FREE Full text] [doi: 10.1016/j.clinthera.2013.10.001] [Medline: 24145043]
33. McGuire DB. Assessment of pain in cancer inpatients using the McGill Pain Questionnaire. Oncol Nurs Forum 1984;11(6):32-37. [Medline: 6568710]
34. Moulin D, Boulanger A, Clark AJ, Clarke H, Dao T, Finley GA, Canadian Pain Society. Pharmacological management of chronic neuropathic pain: revised consensus statement from the Canadian Pain Society. Pain Res Manag 2014;19(6):328-335 [FREE Full text] [Medline: 25479151]
35. Fuller G. How to get the most out of nerve conduction studies and electromyography. J Neurol Neurosurg Psychiatry 2005 Jun;76 Suppl 2:ii41-ii46 [FREE Full text] [doi: 10.1136/jnnp.2005.067355] [Medline: 15961868]

Abbreviations

AUC: area under the curve
MPQ: McGill Pain Questionnaire
ROC: receiver operating characteristic
VAS: visual analog scale

Edited by G Eysenbach; submitted 07.02.18; peer-reviewed by M Alshehri, B Thompson, R Lee, M Linnett; comments to author 29.03.18; revised version received 19.04.18; accepted 15.05.18; published 03.08.18.

Please cite as:
Jonassaint CR, Rao N, Sciuto A, Switzer GE, De Castro L, Kato GJ, Jonassaint JC, Hammal Z, Shah N, Wasan A
Abstract Animations for the Communication and Assessment of Pain in Adults: Cross-Sectional Feasibility Study
J Med Internet Res 2018;20(8):e10056
URL: http://www.jmir.org/2018/8/e10056/
doi:10.2196/10056
PMID:30076127

©Charles R Jonassaint, Nema Rao, Alex Sciuto, Galen E Switzer, Laura De Castro, Gregory J Kato, Jude C Jonassaint, Zakia Hammal, Nirmish Shah, Ajay Wasan. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 03.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of Serious Gaming During the Multidisciplinary Rehabilitation of Patients With Complex Chronic Pain or Fatigue: Natural Quasi-Experiment

Miel AP Vugts[1], MSc; Margot CW Joosen[1], PhD; Agali Mert[2], MD, PhD; Aglaia ME Zedlitz[3], PhD; Hubertus JM Vrijhoef[4,5,6], PhD

[1]Tranzo Scientific Center for Care and Welfare, Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg, Netherlands
[2]Mert Medical, Utrecht, Netherlands
[3]Leiden Institute for Brain and Cognition, Department of Health Medical and Neuropsychology, Leiden University, Leiden, Netherlands
[4]Department of Family Medicine and Chronic Care, Vrije Universiteit Brussel, Brussels, Belgium
[5]Department of Patient & Care, Maastricht University Medical Center, Maastricht, Netherlands
[6]Panaxea BV, Amsterdam, Netherlands

Corresponding Author:
Miel AP Vugts, MSc
Tranzo Scientific Center for Care and Welfare
Tilburg School of Social and Behavioral Sciences
Tilburg University
PO Box 90153
Tilburg, 5000 LE
Netherlands
Phone: 31 13 466 2969
Email: m.a.p.vugts@uvt.nl

Abstract

Background: Current evidence for the effectiveness of specialist multidisciplinary programs for burdensome chronic pain and functional somatic syndromes drives the effort to improve approaches, strategies, and delivery modes. It remains unknown to what extent and in what respect serious gaming during the regular outpatient rehabilitation can contribute to health outcomes.

Objective: The objectives of our study were to determine the effect of additional serious gaming on (1) physical and emotional functioning in general; (2) particular outcome domains; and (3) patient global impressions of change, general health, and functioning and to determine (4) the dependency of serious gaming effects on adherence.

Methods: We conducted a naturalistic quasi-experiment using embedded qualitative methods. The intervention group patients received an additional guided (mindfulness-based) serious gaming intervention during weeks 9-12 of a 16-week rehabilitation program at 2 sites of a Dutch rehabilitation clinic. Simultaneously, 119 control group patients followed the same program without serious gaming at 2 similar sites of the same clinic. Data consisted of 10 semistructured patient interviews and routinely collected patient self-reported outcomes. First, multivariate linear mixed modeling was used to simultaneously estimate a group effect on the outcome change between weeks 8 and 16 in 4 primary outcomes: current pain intensity, fatigue, pain catastrophizing, and psychological distress. Second, similar univariate linear mixed models were used to estimate effects on particular (unstandardized) outcomes. Third, secondary outcomes (ie, global impression of change, general health, functioning, and treatment satisfaction) were compared between the groups using independent t tests. Finally, subgroups were established according to the levels of adherence using log data. Influences of observed confounding factors were considered throughout analyses.

Results: Of 329 eligible patients, 156 intervention group and 119 control group patients (N=275) with mostly chronic back pain and concomitant psychosocial problems participated in this study. Of all, 119 patients played ≥75% of the game. First, the standardized means across the 4 primary outcomes showed a significantly more favorable degree of change during the second part of the treatment for the intervention group than for the control group (beta=−0.119, SE=0.046, *P*=.009). Second, the intervention group showed a greater outcome change in depressive mood (b=−2.748, SE=1.072, *P*=.011) but not in “insufficiency” or concentration problems. Third, no significant group effects on secondary outcomes were found. Fourth, adherence was generally high and invariant.

XSL•FO
RenderX

Conclusions: The findings of this study suggest a very small favorable average effect on relevant health outcomes of additional serious gaming during multidisciplinary rehabilitation. The indication that serious gaming could be a relatively time-efficient component warrants further research into if, when, how, and for which patients serious gaming could be cost-effective in treatment and why.

Trial Registration: Netherlands Trial Registry NTR6020; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=6020 (Archived by WebCite at http://www.webcitation.org/71IIoTXkj)

(J Med Internet Res 2018;20(8):e250) doi:10.2196/jmir.9739

KEYWORDS

behavioral intervention; chronic pain; serious games; musculoskeletal disease; rehabilitation; therapy

Introduction

Background

In a European survey, it has been estimated that 77% of patients with chronic pain (CP) do not access specialist treatments and 40% cannot effectively control their pain [1]. CP is defined by the presence of pain beyond a usual 3- to 6-month duration of organic recovery that may, but does not have to, have an organic cause [2]. Functional somatic syndromes (FSS) are characterized by a persistent pattern of bodily symptoms (ie, pain, fatigue, tinnitus, bowel complaints, and palpitations) for which adequate examination does not reveal sufficiently explanatory specified pathology [3]. Both classifications include, among others, fibromyalgia, chronic low back pain, and irritable bowel syndromes. Global prevalence estimates vary with location and case criteria (severity and disability) and are generally considered high (7%-64% for CP, 3%-43% for tinnitus, 14%-33% for noncardiac chest pain, and 1% for chronic fatigue syndrome) [4-10]. In the absence of satisfactory biomedical solutions, biopsychosocial interventions are offered for improving the physical and emotional functioning [2,3]. A major challenge is to identify which behavior change intervention approaches [11], techniques [12], and delivery modes (eg, computer based) are most accessible and (cost-)effective for certain patients in certain health care settings [13,14]. Herein, a contribution may be made by the pragmatic effectiveness evaluation of a serious gaming intervention (ie, LAKA) during multidisciplinary rehabilitation for patients with emotional or role dysfunctions in association with CP or fatigue [15]. It is hitherto unknown whether, to what extent, and in what respect serious games complement other treatment modes in facilitating intervention effects that are meaningful for patients when offered in addition to other modes of treatment as in regular outpatient multidisciplinary rehabilitation.

Existing Treatment Gaps

National and international guidelines consider various behavioral interventions to be evidence based, but they change with developing insights into the various CP or FSS conditions [16,17]. Intensive multidisciplinary rehabilitation programs are indicated if locally accessible and (unimodal) conservative medication, minimal self-guided intervention, and physical and psychological therapy do not suffice [3,10,17,18]. Ideally, treatment plans are tailored to individual symptom patterns through interdisciplinary procedures [10]. Supporting evidence from randomized controlled trials (RCTs) consists of, at most, medium-sized effects for biopsychosocial interventions compared with alternative treatments [10,18]. An improvement is sought in the modest additional effects of the multi- or interdisciplinary rehabilitation over other kinds of “unimodal” treatments [10,18-20]. There’s growing evidence for the efficacy of acceptance- and mindfulness-based interventions, which can be included in the multi- or interdisciplinary rehabilitation [10,11]. Rather than addressing certain presumed maladaptive illness beliefs, this sort of approach aims to cultivate self-awareness, self-regulation, and self-transcendence in response to aversive conditions such as CP or FSS [11,21]. Moreover, behavioral interventions have approximately equivalent effects when delivered via computers or internet, but adherence to such interventions can be disappointing [14,22]. Promotion of motivation and adherence (by professionals) may lead to better therapy outcomes [23] and is likely to be of help to patients when using computer-based programs [22]. Few trials have reported mixed results for the efficacy of varied computer-delivered interventions (ie, mobile phone, automated telephone responding, and online support group) offered in addition to face-to-face intervention to patients with CP [24-26]. Generally, effect studies of biopsychosocial interventions may need to improve in their methodological quality (ie, statistical power, risk of selection and reporting bias) and uniformity (ie, definitions of case and recovery, diagnostic methods, subjective and objective outcome criteria, and program description) [27,28].

Why Serious Games May Offer a Potential Solution

Serious games, which primarily aim at health benefits, may take the form of a video game [29]. Indeed, how harmful or conducive video gaming is for behavior and health depends on the content (eg, whether it reinforces aggressive or prosocial actions) and context (ie, players and instructional support) of the game [30,31]. Serious games may combine small behavioral and clinical benefits with independent accessibility and standardized content of computer-based interventions as well as unique qualities for learning such as intrinsic motivation, enjoyment, positive affect, sense of presence, and meta-cognition [29,32-34]. Games are a ubiquitous but undefinable cultural phenomenon described as bounded “spaces” physically, imaginarily, or in time apart from ongoing reality, wherein individuals involve voluntarily, create meanings, and develop adaptive capacities, such as sports and rituals [35,36]. Intrinsic motivation, as in games, is beneficial for learning quality [37]. It has been hypothesized that behavioral change is strengthened by engagement qualities triggered by

XSL•FO
RenderX

storytelling, fantasy, and interactivity in serious games [33]. In the fields of mental health care and rehabilitation, gaming, motion capturing, and virtual reality technologies potentially support the treatment of various well-known conditions such as depression, anxiety, phobias, poststroke, and acute pain [38-42]. After serious gaming, debriefing may be offered to facilitate the transfer of patient experiences into targeted individual learning results [43]. Thus, previous studies have shared the idea of (subtle) positive moderation of treatment effects because of distinctive beneficial motivational qualities triggered by features of serious games. However, adequately powered studies on the comparative effects of games for health are lacking in general, and little to nothing is known about their complementary effectiveness in regular health care contexts, such as multidisciplinary rehabilitation [44]. Moreover, patient adherence to serious gaming when deployed in practice and its influence on outcomes require empirical assessment in effectiveness evaluation [45].

A Mindfulness Approach to Serious Gaming in Multidisciplinary Rehabilitation

Serious gaming can be a complementary modality that strengthens mindfulness-based modules in treatments like multidisciplinary rehabilitation. Adopting a mindfulness approach to serious gaming deviates from an approach wherein particular antecedent cognitions of health behaviors are targeted [33]. Mindfulness approaches offer mental training principles (ie, focused attention, open monitoring, or ethical enhancement) for promoting (1) a temporary state of nonjudgmental, nonreactive, present-centered attention and awareness (self-awareness); (2) a capacity to effectively modulate behavior (self-regulation); and (3) a positive relationship between self and others, transcending self-focused needs and increasing prosocial characteristics (self-transcendence) [21]. A complementary role of mindfulness-based serious gaming might not necessarily be to facilitate mental training in patients over prolonged periods, but rather to promote independent practicing by any (other) means in the context of daily life. Plausibly, mental training objectives can be temporarily achieved in conjunction with gaming [46], but it can contradict an obsessive drive that could characterize long-term and frequent video gaming [47]. Over longer durations, individuals may apply mental training principles independently in various ways, depending on behavioral factors (recollection of instructions, intent, habit) [21]. Change techniques (eg, commitment to change, action planning, drawing attention to discrepancies between behavior and goals, noncontingent praise, performance instructions, self-monitoring, and salient feedback on behavior) as well as emotional and social consequences, reduction of negative emotions, values affirmation, etc [12], for (novice) mental training activities could be integrated (via an "Avatar" role) into a serious game. From this line of argument, it was proposed that a short serious gaming intervention adds to the effectiveness of a mindfulness-based approach during multidisciplinary rehabilitation for (subtly) better effects on relevant health outcomes in patients with CP or FSS.

Objectives

In this study, we investigated the effectiveness of serious gaming as a complement to the multidisciplinary rehabilitation of patients with CP or FSS. The selection of health outcomes was guided by a field consensus on the relevance of physical and emotional functioning, patients' global impression of improvement, and negative effects [48]. The primary objective was to determine the effect of additional serious gaming on multiple domains of physical and emotional functioning simultaneously. Secondary objectives were to understand which outcome domains are particularly affected, positively or negatively, by serious gaming during rehabilitation and whether serious gaming affects patient's global impressions of change, general health, and functioning. The final objective was to determine whether outcomes of serious gaming are dependent on adherence. Adherence is defined as the extent to which patients expose themselves, in terms of content, frequency, and duration, to the "hard core" of a serious gaming intervention—playing a serious game and attend the debriefing. The following were the research questions:

1. To what extent does an additional serious gaming intervention affect a change in patients' physical and emotional functioning during regular multidisciplinary rehabilitation?
2. Regarding which particular domain(s) of physical and emotional functioning does an additional serious gaming intervention affect outcome change during multidisciplinary rehabilitation?
3. To what extent does an additional serious gaming intervention during multidisciplinary rehabilitation affect patients' impressions of change, subjective health and functioning, and satisfaction with treatment?
4. To what extent is the degree of effectiveness dependent on levels of adherence?

Methods

Study Design

A protocol for this embedded experimental mixed-methods study was registered (preresults) in the Dutch trial register (NTR6020), previously published in detail [15], and followed accordingly. General information and important executive details relevant to the present objectives are discussed here. In the absence of a legal obligation for medical ethics review, the protocol was reviewed for the protection of patients' rights in accordance with the letter and reasoning of applicable legislation and research practice and endorsed by the Psychological Ethics Committee of the Tilburg School of Social Sciences (EC-2016.25t). The study design sorts with the nature of multidisciplinary rehabilitation, which is complicated by tailoring, multiple interacting components, and outcome multidimensionality [45]. Quantitative methods were prioritized for assessment purposes. The two-armed naturalistic quasi-experiment was set up pragmatically, comprising an intervention group of patients who received an additional serious gaming intervention offered during weeks 9-12 of a standardized 16-week rehabilitation program at 2 sites of a Dutch rehabilitation clinic. Simultaneously, an approximately equal

XSL•FO
RenderX

number of control group patients followed the same program without serious gaming, as usual, at 2 similar sites of the same clinic (from February 2016 to January 2017). Concurrently collected qualitative data were first used to refine hypotheses blind to trial outcomes and later for triangulation and post hoc explanation of quantitative results.

Setting, Recruitment, and Data Sources

The convenient selection of control sites aimed for homogeneity across the study groups. The 4 participating sites were located in the south of the Netherlands, where multidisciplinary biopsychosocial rehabilitation, but not serious gaming, is covered under basic health care insurance. In view of ecological validity, all patients with a regular physician indication for multidisciplinary rehabilitation who completed the first 8 weeks of rehabilitation were considered eligible for this study. From the beginning of the second half of their rehabilitation program (July-November 2016), patients were consented by their direct care providers. This timing was chosen for patient convenience and optimal response. To lower the risk of selection bias, patient recruitment was closely monitored through regular site visits. Consent was requested for the processing of patients' codified clinical diagnostic and outcome data and, perhaps, being contacted for an interview. Outcome data consisted of patients' routine outcome monitoring administered by the clinic through a standardized Web-surveying procedure at the baseline (t0), intermediate (t1: after 8 weeks of treatment), and posttreatment (t2: after 16 weeks of treatment). Only intervention group candidates were requested to answer feedback questions through the same familiar Web-survey procedure immediately after their debriefing session.

To avoid bias by inflicting outcome expectations in patients as subjects and outcomes assessors, information letters did not contain statements about presumed effects of serious gaming or parallel group comparison. After serious gaming, feedback data were made available to the researchers to support the concurrent qualitative research, but routinely collected clinical (diagnostic and outcome) data were not. In this way, data management served to prevent the risk of biased interpretation through breaching the protocoled sequence of hypotheses refinement and quantitative testing.

Patients

Based on physical and psychological examination results and clinical interviews, physicians indicated eligibility for multidisciplinary rehabilitation treatment based on the following inclusion criteria. The eligible patients were between 18-67 years of age, had pain for more than 6 months or fatigue complaints or musculoskeletal disease for more than 3 months, had no indication for another more cost-effective treatment, and had concomitant psychosocial problems. The exclusion criteria were as follows: patients with psychiatric symptoms that are not adequately controlled, a marked risk of psychological decompensation through a rehabilitation treatment, language or communication problems that make it impossible to follow rehabilitation, or demonstrable inability to change behavior (eg, due to personality disorders, third party liabilities, or otherwise). Notably, no additional computer literacy criteria were applied for participation in this study.

Interventions

Both study groups received an intensive 16-week biopsychosocial multidisciplinary rehabilitation program with a particular focus on well-being and social role participation [49] (Table 1). Under the supervision of a rehabilitation physician, patients received on average 100 hours of treatment in either one-on-one or in group settings from a team of 2 physiotherapists and 2 registered psychologists with a master's degree. Weekly intensity varied between 3 and 7 hours, decreasing with an increase in social role participation throughout the program. Overall, patients received 38% physiotherapy, 30% mindfulness approaches, 23% (other kinds of) psychotherapy, and 9% of activating and counseling in social role participation.

XSL•FO
RenderX

Table 1. Overview of program components offered during the first and second halves of the regular multidisciplinary rehabilitation program.

Components offered	Weeks 1-8	Weeks 9-16
Physical therapy	• Graded activity (group) • Exercise therapy • Physiotherapy[a] • Education (lifestyle, pain physiology)	• Graded activity (group) • Exercise therapy
Psychotherapy approaches	• Coping with stress • Extinction of fear-avoidance beliefs[a] • Cognitive restructuring[a] • Eye movement desensitization[a] • Mentalization techniques[a]	• Coping with stress • Cognitive restructuring
Activating and counseling in social role participation	• Work and health (education and counseling) • Social skills	
Mindfulness interventions	• Rationale • Psychological well-being assessment • Mental training • Additional 2-day course (mental training skills)	• Mental training

[a]Allocation depended on examination results.

The various interventions were centrally assigned, based on individual examination results for physical status, psychological and posttraumatic distress, coping, cognition, and well-being. In this study, the strategies used to promote health behavior were as follows: shaping knowledge about antecedents and health consequences, goal setting and feedback, social support, exposure, behavioral repetition and substitution, skills training (in relaxation, social skills, and mental training), and identity development (ie, cognitive restructuring and values affirmation). Mindfulness interventions already included in the basic program included basic rationales, mental training instructions, and psychological well-being assessment. An intensive 2-day mental training course was offered to all patients, except those with high levels of well-being.

The treatment offered to the intervention group only differed systematically from the control group in the addition of a serious gaming module (the control group did not receive something else instead); this was verified empirically. For the intervention group, the rehabilitation clinic had suitable facilitated rooms with Wi-Fi connections, tablet PCs installed with the serious game "LAKA," and the automated planning of four 1-hour small group sessions (1-6 patients simultaneously) in connection to regular therapy hours (mostly exercise sessions at the beginning or end of a working day) during weeks 9-12. Sessions were planned for patients to have sufficient time for completing the game, at least, once. Patients logged in with their personal identification number and self-chosen password with which they also accessed Web surveys. Experienced therapists (3 psychologists and 1 physiotherapist) were scheduled to provide support during the first (introduction) and fourth (debriefing) sessions. The goal of debriefings was to discuss the experiences of game play and technology acceptance and to transfer learning results to patients' daily lives. Other local staff members managed the accessibility of the game LAKA during sessions 2 and 3. Notably, patients could also download and play LAKA at home. Local therapists and other staff participated in developing their role in the delivery of serious gaming.

The Serious Game LAKA

The serious game LAKA is an adventure game where patients take the role of an Avatar during a virtual trip around the world. The game is easy to control using a touch-screen tablet computer and takes on average 2.5 hours to complete (Multimedia Appendix 1 shows the screenshots).

In LAKA, patient players perform alternate tasks vicariously; they select optional responses in various encounters with other characters, monitor and evaluate satisfaction about selected responses (and their consequences), and meditate (3-minute exercises). First, players select between a male or female Avatar and assign a name. It was prompted that Avatar choices reflect those of the player. A cut-scene sets up the story; the Avatar, who wants change after experiencing a deterioration in physical and social functioning, meets a nonplaying character (NPC) named LAKA. LAKA challenges the Avatar to make "conscious" decisions during 16 "encounters" with other NPCs, for example, when standing in line, on getting invited to someone's home, and at 4 destinations (ie, London, Turkey, Asia, and Africa) on a trip around the world. Each "encounter" is built as a flow of Avatar actions and NPC responses.

For each Avatar action, 1 of 5 options (eg, physically interact, verbally react, or ignore something) can be preselected and confirmed by players. These options are modeled after a set of reference values—generosity, moral discipline, forbearance, and enthusiastic perseverance. NPC responses are unpredictable, for example, a friendly act can result in a kind response or being scammed. At the end of each destination, LAKA asks the Avatar to self-rate the level of "satisfaction" regarding his or her choices. Indirect feedback, in the form of a number of puzzle pieces, is given by (1) the degree of correspondence of Avatar choices with the reference values and (2) the degree to which that correspondence agrees with satisfaction ratings. When the

XSL•FO
RenderX

Avatar is depicted "mind-wandering" when traveling across destinations, instructions are received for a basic meditation exercise (focused attention and open monitoring) [50]. These model-based elements are interspersed with short action games, puzzle games, images, and information associated with the location of the Avatar to be enjoyed or skipped by preference.

Quantitative Measures

Outcomes

Table 2 provides an overview of the outcome variables, surveys with references to instrument validity information, and times of assessment. Available primary outcome measures for operationalizing elements of research questions 1 and 2 included 4 evidently valid numerical rating- and Likert-scale measures that operationalize relevant and plausible targets for mindfulness-based intervention in the target group [11,48]: a numerical rating scale for the current pain intensity, the Checklist Individual Strength (CIS) for fatigue, the catastrophizing subscale of the pain coping and cognitions list, and the Symptoms Checklist (SCL-90) for psychological distress [51-54]. In addition, Likert-scale items on patients' global impression of change (PGIC), general health and functioning, and treatment satisfaction were available to operationalize secondary outcomes. The PGIC was measured using a single ordinal scale item [48]. Three available 0-100 numerically scaled questionnaire items about perceived health and functioning formed an internally consistent scale (Cronbach alpha at t1=.80, at t2=.75) for general subjective health and functioning. Furthermore, 3 different questions assessed the treatment satisfaction, 2 of which were taken from the consumer quality index for rehabilitation centers [55].

Adherence

The operationalization of adherence distinguishes between progress in game play and debriefing attendance. The latter was established from clinical recordings of presence at either the initially planned or a rescheduled debriefing session. Log data from the game, designed to track the progress and give feedback on the performance, were used to determine the categories of completion percentage. These categories (5-1) represent 0%-6% completion when no progression logs were observed and <50%, 50%-75%, 75%-100%, and >100% game completion when patients played the game LAKA more than once.

Table 2. Primary and secondary outcome measures.

Variables	Survey information	Time[a]
Primary outcomes		
Pain intensity	• Current pain intensity Numerical Rating Scale 0-100 [53]	t0, t1, t2
Fatigue	• Checklist Individual Strength [51]	t0, t1, t2
Psychological distress	• Symptom Checklist [52]	t0, t1, t2
Pain coping and cognition	• Pain Coping & Cognitions List; catastrophizing subscale [54]	t0, t1, t2
Secondary outcomes		
Global impression of change, general health, and functioning	• How do you assess your health, compared with the situation at the start of your treatment? (−2, much, or −1, slightly declined; 0, neither declined nor improved; 1, slightly, or 2, much improved) • What do you think of your current health in general? (0, bad-100, excellent) • Please indicate how satisfied you are generally taken with your current level of functioning. (0, not at all satisfied-100, very satisfied) • Please indicate the distance from your "old" level of functioning before the onset of the complaint. (0, very far removed-100, not at all removed)	t1, t2
Treatment satisfaction	• Would you recommend this treatment center to other rehabilitation patients? (1, certainly not; 2, probably not; 3, probably yes; 4, certainly yes); item from the consumer quality (CQ) index [55] • Which grade would you give to the rehabilitation center? (0-10; CQ-index) • Did the treatment meet your expectations? (1, not at all; 2, mostly not; 3, mostly; 4, completely)	t2

[a]t0=baseline; t1=intermediate (after 8 weeks of treatment); t2=posttreatment (after 16 weeks of treatment).

Case Description and Potential Confounders

Data retrieved for describing patients and enabling the optimal control for observed confounding factors consist of demographics, medical history, physical and psychological examination results, and registrations of allocated and attended interventions.

Study Size

Study size was determined by *a priori* power calculation as described in detail in the protocol [15]. G*Power was used to calculate a required sample size of 212 patients. These calculations were based on a multivariate analysis of variance of global effects (1−beta=0.8, f^2=0.0625, alpha=.05, 2 groups, and maximally 5 outcomes). By taking a margin of 20% for

XSL•FO
RenderX

dropout and missing values into account, the minimum number of patients was finally determined at 250.

Statistical Methods

All statistical analyses were performed using SPSS 22 (IBM, New York). Descriptive statistics and chi-square and Student *t* tests were used to summarize demographic, disease-specific, treatment exposure, and baseline outcome characteristics. Variables that may differ per group on the baseline were added as covariates in subsequent analyses. Statistical methods were generally aimed at testing two-sided hypotheses regarding study group differences in differences between intermediate and posttreatment outcome levels as this corresponds with the timing of the additional serious gaming intervention. Furthermore, Sidak-Holm correction was used when controlling for multiple outcome testing [56].

First, a multivariate mixed linear effect model was fitted to estimate a parameter for the study group difference in simultaneous change of the 4 primary outcomes between intermediate and post treatment assessments. For this, the MIXED procedure for defining parallel growth processes was applied on standardized scores of the 4 primary outcome variables (Textbox 1) [57]. The procedure facilitates an intention-to-treat analysis, optimizes statistical power, imposes an equality constraint on parameter estimates across multiple outcome measures, and takes outcome interdependencies into account. Hereby, the parameter estimation is unbiased under the assumption of missing data at random. To correct the parameter estimate for design limitations, it was modeled together with components that are not logically attributable to the serious gaming intervention, including global time and group effects, group differences in outcome changes in time prior to serious gaming, and covariates. Furthermore, selections between nested models, that is, excluding or including factors for time, treatment sites, and covariates were based on the statistical significance of changes in the model fit.

Textbox 1. Linear mixed modeling operations.

Operational details on the (planned or initial) multivariate linear mixed model:

- All models applied the restricted maximum likelihood estimation.
- MIXED requires a vertical (re)structured dataset with all outcome values inserted in one column, nesting primary outcomes (4) and time factors (3) within individuals.
- Indexes were created for individuals (1-275), outcomes over time (1-12), outcomes (pain intensity=1, fatigue=2, catastrophizing=3, and psychological distress=4), time (baseline=1, post=2, and intermediate=3 [reference category]), and group (1=intervention, and 2=control [reference category]).
- The procedure used standardized outcome values, calculated separately within outcomes.
- An unstructured covariance matrix (UN) for the random effects and a heterogeneous autoregressive matrix for the repeated effects were assumed when fitting multivariate models. Use of alternative covariance structures (UN, compound symmetry, autoregressive, Toeplitz, and ante dependent) either disabled convergence or resulted in worse fit.
- Basic model specification: the outcome index was specified as randomly varying for the estimation of intercepts for each of the 4 outcomes. The planned “basic” model contained 11 fixed-effect parameters, including 4 outcome factors, 3 treatment sites, 2 time factors (1=intermediate vs baseline, 2=intermediate vs post), and 2 group × time factors (group × time 1, group × time 2), with random error terms (10) and repeated effects (13), adding up to 34 parameters to be estimated in total.
- Model fit changes, that is, exclusions (ie, site and time factors) or inclusions (covariates), were assessed using (chi-square) tests for differences in the −2 Log Likelihood information criterion.
- Sensitivity analyses revealing similar results included multivariate models run on full cases only, outcome data after outlier removal (*z*-scores above 3 or 5), and alternative *z*-score calculations.

Operational details for univariate mixed linear models:

- An unstructured covariance matrix was assumed for all univariate models.
- Univariate models included the same covariates as the multivariate models.

Subsequently, reliable change indexes (RCIs) were calculated (again based solely on the difference between intermediate and posttreatment scores) to determine within-group proportions of individual patients who reported improvement (RCI<−1.96) or decline (RCI>1.96) [58]. Improvement was defined as a clinically significant decrease (RCI<−1.96) in one or more of the 4 primary outcome variables. Decline was defined as a reliable increase in one or more outcomes (RCI>1.96). When patients did not show decline or improvement, their status was deemed stable. Differences in proportions in these categories were compared between the groups.

Second, effects of serious gaming on particular plausible outcome types identified through qualitative research were estimated using univariate mixed linear effect models of unstandardized outcomes. Third, (changes in) secondary outcomes were compared between the groups. Fourth, the multivariate linear mixed model was rerun after replacing the original group dummy variable by ordinal adherence variables to calculate parameter estimates separately for subgroups of differing rates of serious gaming progress and debriefing the attendance relative to controls (the reference category).

XSL•FO
RenderX

Concurrent Qualitative Methods

Qualitative data consisted of patients' typed responses to an open feedback question and audiorecorded, verbatim transcribed semistructured interviews. The open feedback question was: "How do you think serious gaming will contribute to your daily life (ranging from 0=negatively or nothing to 10=hugely)? And in what way?" We purposively selected 8 patients with varying expected contributions of serious gaming for semistructured face-to-face interviews (lasting 30-60 minutes). Of them, 2 were selected for their high expectations (scoring ≥9), 2 for their low expectations (scoring ≤1), and 4 for their mediocre expectations (4-6) regarding the contribution of serious gaming to their daily living. In addition, telephone interviews were planned with control group subjects who had been matched by direct care providers on case descriptions by gender, age, symptom patterns, and coping style. However, this was stopped after 2 short interviews (lasting <15 minutes) as it was not regarded informative due to case differences beyond the small set of matching variables. All interviews started with a request to patients to talk openly about their health status before rehabilitation and any changes experienced throughout. Subsequently, patients were invited to elaborate on the perceived contribution of serious gaming.

A deductive content analysis approach was performed on the interview transcripts using Atlas.ti. The first 4 interviews were coded independently by MV and a second author (MJ or AZ). Then, because no more differences in coding were observed, MV coded the remaining interviews. First, data were reduced by distinguishing text fragments related to patients' expected health outcomes. To those fragments, labels were attached according to sensitizing concepts about (subjective) health outcomes because existing theoretical frameworks on relevant health outcome dimensions were available and preselected for quantitative operationalizations [51-54]. Those sensitizing concepts covered relevant outcome domains for patients with CP (physical symptoms, physical functioning, and emotional functioning (eg, anxiety and depressed mood) [54] or fatigue (subjective fatigue and motivation and concentration problems) [51].

Results

Participants

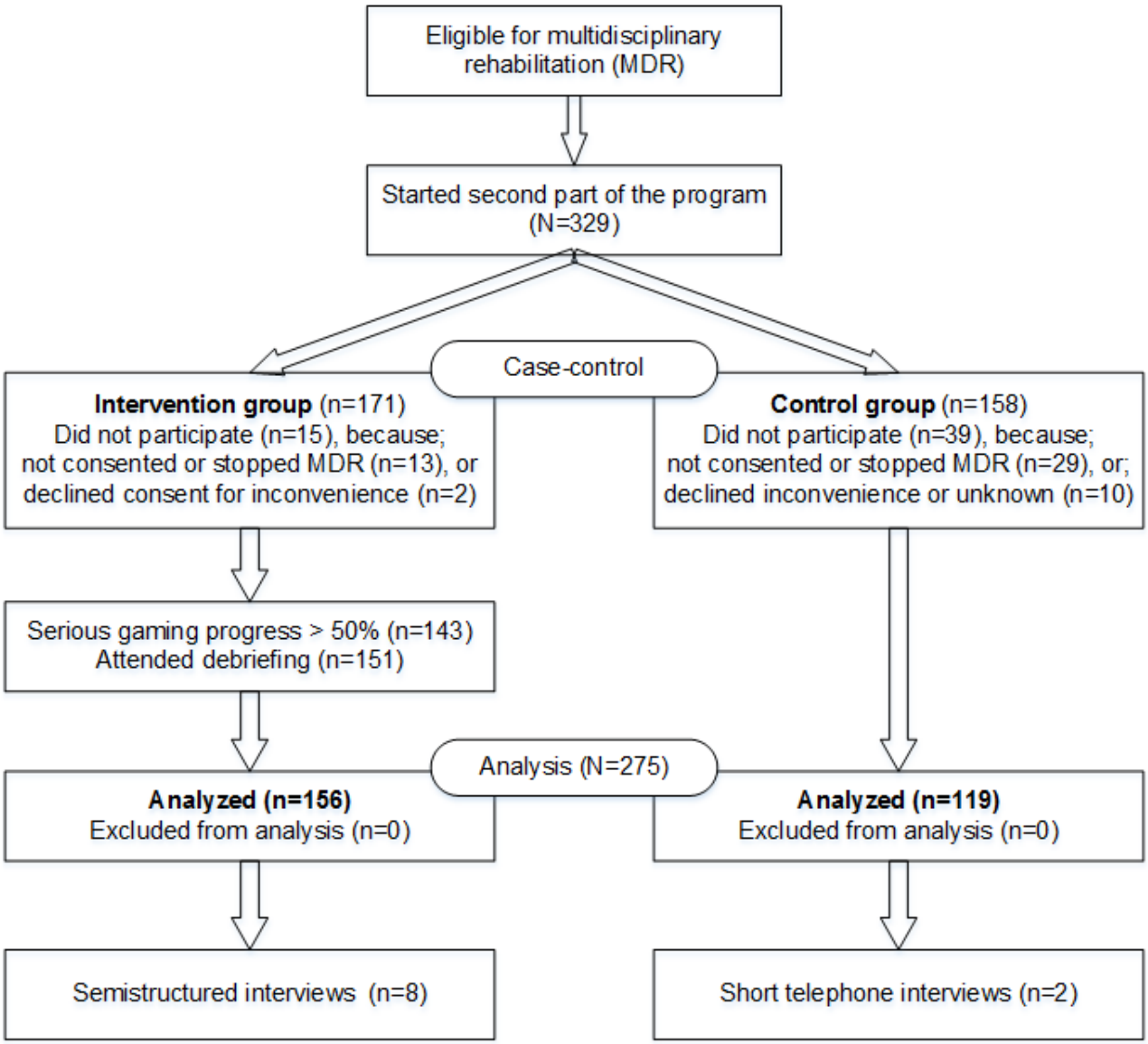
Recruitment was stopped when sufficient numbers of participants were included in both study groups. By then, 83.6% (275/329) of eligible patients who had started the second part of the treatment consented to participate—156 in the intervention and 119 in the control group (Figure 1). A decline in participation due to reasons such as inconvenience (ie, presumed burdens of the treatment or study at the time of consent), delayed consent, and thereby missing patients who stopped participating during the second part of the treatment was observed more often at control sites. Furthermore, posttreatment data were missing for 1 control group patient and 7 intervention group patients.

Descriptive Data

Back pain was the most prevalent physical symptom reported (184/275, 66.9%), but pain symptoms with other origins (headaches, gastrointestinal, fibromyalgia, and osteoarthritis) were reported as well (Table 3). Social problems were also prominent, with 49.8% (137/275) of the patients experiencing problems with family members. The modal norm group categories for the SCL-90 anxiety and depression symptom scales were "very high" relative to "healthy individuals" and "high" relative to patients with CP, but "below average" relative to psychiatric patients [52]. Compared with all 3 norm groups, modal score categories for sleeping problems were high. The study groups were similar regarding most baseline characteristics, but several *P* values found, suggesting differences between the study groups. Marginally higher socioeconomic status (SES) and more comorbid neurological and less cardiac diseases were observed in the intervention group. In the control group, relatively more patients reported back pain, were taking medication at the baseline, and had returned to work at intermediate assessment. In general, more than half of the patients had suffered from their chronic fatigue or pain condition for over 2 years before entering the rehabilitation and most had received prior (specialized) care for this (Table 3).

XSL•FO
RenderX

Figure 1. The CONSORT flowchart of participants.



XSL•FO
RenderX

Table 3. Participants' baseline characteristics.

Variable	Intervention group (n=159)	Control group (n=116)	Full sample (N=275)	Group difference	
				Student t[a] or χ^2 test	*P* value
Age (years), mean (SD)	44.2 (11.55)	44.9 (11.42)	44.5 (11.48)	t_{273}=−0.5	.60
Female gender, n (%)	104 (66.7)	88 (73.9)	192 (69.8)	χ^2_1=1.7	.19
Socioeconomic status score[b], mean (SD)	−.05 (0.95)	−.36 (1.28)	−.18 (1.12)	t_{272}=2.223	.03
Returned to work (at intermediate)	16 (10.3)	24 (20.2)	40 (14.5)	χ^2_1=5.4	.02
Highest educational level (N=132)[c], n (%)				χ^2_3=2.1	.56
Primary school	1 (1.3)	0 (0.0)	1 (0.8)		
Lower secondary education	21 (27.3)	12 (21.8)	33 (25.0)		
Higher secondary education	28 (36.4)	19 (34.5)	47 (35.6)		
Tertiary education	28 (36.4)	23 (41.8)	51 (38.6)		
Work status, n (%)				χ^2_3=1.6	.67
Full-time employed	24 (15.1)	18 (15.5)	42 (15.3)		
Fully absent	55 (34.6)	48 (41.4)	103 (7.5)		
Partially absent	39 (24.5)	31 (26.7)	70 (25.5)		
Unemployed	38 (24.4)	22 (18.5)	60 (21.8)		
Pain locations, n (%)					
Neck or head	77 (49.4)	61 (51.3)	138 (50.2)	χ^2_1=0.1	.76
(Low) back	92 (59.0)	92 (77.3)	184 (66.9)	χ^2_1=10.3	.001
Upper extremities	57 (36.5)	44 (37.0)	101 (36.7)	χ^2_1=0.0	.94
Lower extremities	59 (37.8)	57 (47.9)	116 (42.2)	χ^2_1=2.8	.09
Symptom duration, n (%)				χ^2_3=1.9	.60
3-6 months	13 (8.4)	10 (8.4)	23 (8.4)		
6-12 months	35 (22.6)	20 (16.8)	55 (20.1)		
1-2 years	36 (23.2)	26 (21.8)	62 (22.6)		
>2 years	71 (45.8)	63 (52.9)	134 (48.9)		
Symptom course, n (%)				χ^2_2=0.1	.95
Deteriorating	100 (4.1)	76 (63.9)	176 (4.0)		
Improving	23 (14.7)	19 (16.0)	42 (15.3)		
Presence of comorbid medical diagnoses, n (%)					
Cardiology	19 (12.2)	25 (21.0)	44 (16.0)	χ^2_1=4.0	.048
Neurology	16 (10.3)	0 (0.0)	16 (5.8)	χ^2_1=13.0	<.001
Endocrinology	14 (9.0)	10 (8.4)	24 (8.7)	χ^2_1=0.0	.87
Pulmonology	24 (15.4)	19 (16.0)	43 (15.6)	χ^2_1=0.0	.90
Visits to other health care providers during the program, n (%)				χ^2_4=1.7	.79
Never	54 (36.2)	44 (37.3)	98 (36.7)		
1-2 times	49 (32.9)	37 (31.4)	86 (32.2)		
3 times or more	46 (30.9)	37 (31.3)	83 (31.1)		

XSL•FO
RenderX

Variable	Intervention group (n=159)	Control group (n=116)	Full sample (N=275)	Group difference	
				Student t[a] or χ^2 test	P value
Body mass index (kg/m^2), mean (SD)	27.0, 5.16	27.1, 5.13	27.1, 5.13	t_{270}=−0.181	.86
(Very) low oxygen absorption capacity (Åstrand class)[d], n (%)	44 (38.3)	36 (42.3)	80 (40.1)	χ^2_1=3.0	.81
Symptom recurrence (yes), n (%)	93 (59.6)	73 (61.3)	166 (60.4)	χ^2_1=0.1	.77
Previous specialized medical care received (yes), n (%)	101 (64.7)	85 (71.4)	186 (67.6)	χ^2_1=1.4	.24
Treated elsewhere (baseline), n (%)	81 (52.3)	64 (53.8)	145 (52.9)	χ^2_1=0.1	.80
Medication intake, n (%)	104 (66.7)	93 (78.2)	197 (71.6)	χ^2_1=4.4	.04

[a]If Levene's test for equality of variances was significant, equal variances were not assumed.

[b]The socioeconomic status (SES) index by neighborhood is derived from a number of characteristics of the people living there: their education, income, and position in the labor market. The higher the index, the higher the SES. Nationally, the mean is 0, SD is 1.09, and the minimal and maximal scores are −6.75 and 3.06, respectively.

[c]Highest education data were incomplete because it was not administered for a part of the course of the natural experiment; missing values (N=143) were group independent.

[d]Physical condition: age- and weight-corrected oxygen absorption capacity measured using the submaximal Åstrand performance test. Missing values are because of the exclusion of observations under 120 beats per minute or testing contraindications (ie, high blood pressure).

Qualitative Results

Codes describing intervention group patients' (4 males, 4 females) responses to open questions about expected outcome changed because serious gaming did not contain the domains of physical symptoms, physical functioning, or subjective fatigue. However, in interviews with patients with the highest expectations (score 8 or 9 out of 10; 2/8, 25%), possible benefits in the realms of emotional functioning (ie, depressed mood and obsessive or compulsive behavior) and concentration problems were voiced. These outcome domain labels were attached to patients' expressions about expectations of improved awareness, regulation, or transcendence of negative thought and lack of interest (depressive mood), problems in decision making (obsessive or compulsive behavior), or losing focus on tasks (concentration problems), which is (partly) illustrated by the following quotes:

> *What I gain from it? Yes, maybe that when you're busy with something…that you're really focused on it and not being distracted…Yes, it's clear that I have that focus more.*

> *In your daily life you are confronted with things that you, or I in any case, did not initially see as stress…Well, I often travel by train, and sometimes things annoy me, but I usually ignore it. Now I have something like: I can talk to them…So, you are irritated, and at the moment you notice it you are annoyed, so it's getting worse…Yes, you can just make it go away so that it does not adversely affect your mood.*

Following these perceptions in a minority of patients, it was proposed that serious gaming generally facilitates a small amount of additional change regarding the primary outcomes of fatigue and emotional functioning (see trial registry). Moreover, additional change for patients in the intervention group was expected to be reflected by observing stronger decreases in scores based on depression and insufficiency subscales of SCL-90 and concentration problems subscale of CIS.

Quantitative Outcome Assessment

At the baseline, participants reported on average moderate pain intensity, high fatigue, and high psychological distress levels compared with norm group averages (Table 4). After treatment, average outcome score levels were subsequently mild, higher than average, and "average" (relative to healthy norm groups).

The final multivariate mixed linear model included a study group dummy instead of the site index, SES scores, and intermediate return to work as (potential) confounding variables (Textbox 2). Model fit did not improve by adding pain location or comorbidity factors, medication intake, and amounts of particular kinds of psychotherapy received.

In addition, patterns of change in all 4 primary outcomes taken together throughout the rehabilitation program of each study group were visualized (Figure 2), showing that outcome scores improved in parallel before exposure to serious gaming and improved relatively more for the intervention group between intermediate and posttreatment. The multivariate mixed model, which assumed equivalent changes across the 4 primary outcomes, indicated statistically significant improvement over the first half (beta=−.805, SE=0.042, P<.001) and the second half (beta=−.473, SE=0.034, P<.001) of treatment. The parameter estimate for the interaction effect (simultaneously on the 4 outcomes) of group × time 1-2 (representing the interval between intermediate and posttreatment) favored the intervention group to a very small extent by −.119 (SE=0.046, P=.009); this equals to 8.59% of the total amount of outcome change within the intervention group.

XSL•FO
RenderX

Table 4. Primary outcome scores.

Outcome; measure and time[a]	Intervention group		Control group	
	Number of observations	Mean (SD)	Number of observations	Mean (SD)
Pain intensity (current); Numerical rating scale (0-100) [53]				
t0	156	56.60 (32.27)	119	58.71 (30.92)
t1	156	35.79 (25.80)	119	35.03 (26.31)
t2	150	26.08 (24.07)	118	29.81 (25.56)
Fatigue Checklist Individual Strength (CIS) [51][b]				
t0	154	110.97 (18.63)	118	108.05 (15.88)
t1	154	84.05 (26.51)	118	83.42 (24.57)
t2	147	60.68 (27.04)	116	65.62 (26.07)
Catastrophizing; Pain Coping & Cognitions List (PCCL) [54][c]				
t0	126	3.54 (0.94)	97	3.46 (0.86)
t1	126	2.68 (0.86)	97	2.60 (0.93)
t2	121	2.10 (0.90)	95	2.05 (0.88)
Psychological distress; Symptom Check List (SCL-90) [52][d]				
t0	156	195.25 (50.55)	119	193.21 (49.40)
t1	156	161.93 (43.21)	119	151.18 (35.92)
t2	149	120.45 (32.96)	117	118.26 (28.90)

[a]t0=measurement; t1=intermediate (after 8 weeks of treatment); t2=posttreatment (after 16 weeks of treatment).

[b]Norm information for the CIS: average for healthy controls, mean=41.5, SD, 19.8; average of a norm group of patients with chronic fatigue syndrome, mean=113.3, SD, 14.6.

[c]Some data are missing by the design of routine outcome monitoring; PCCL scores are absent for very low pain intensity scores.

[d]The baseline mean is high compared with a norm group of patients with chronic pain.

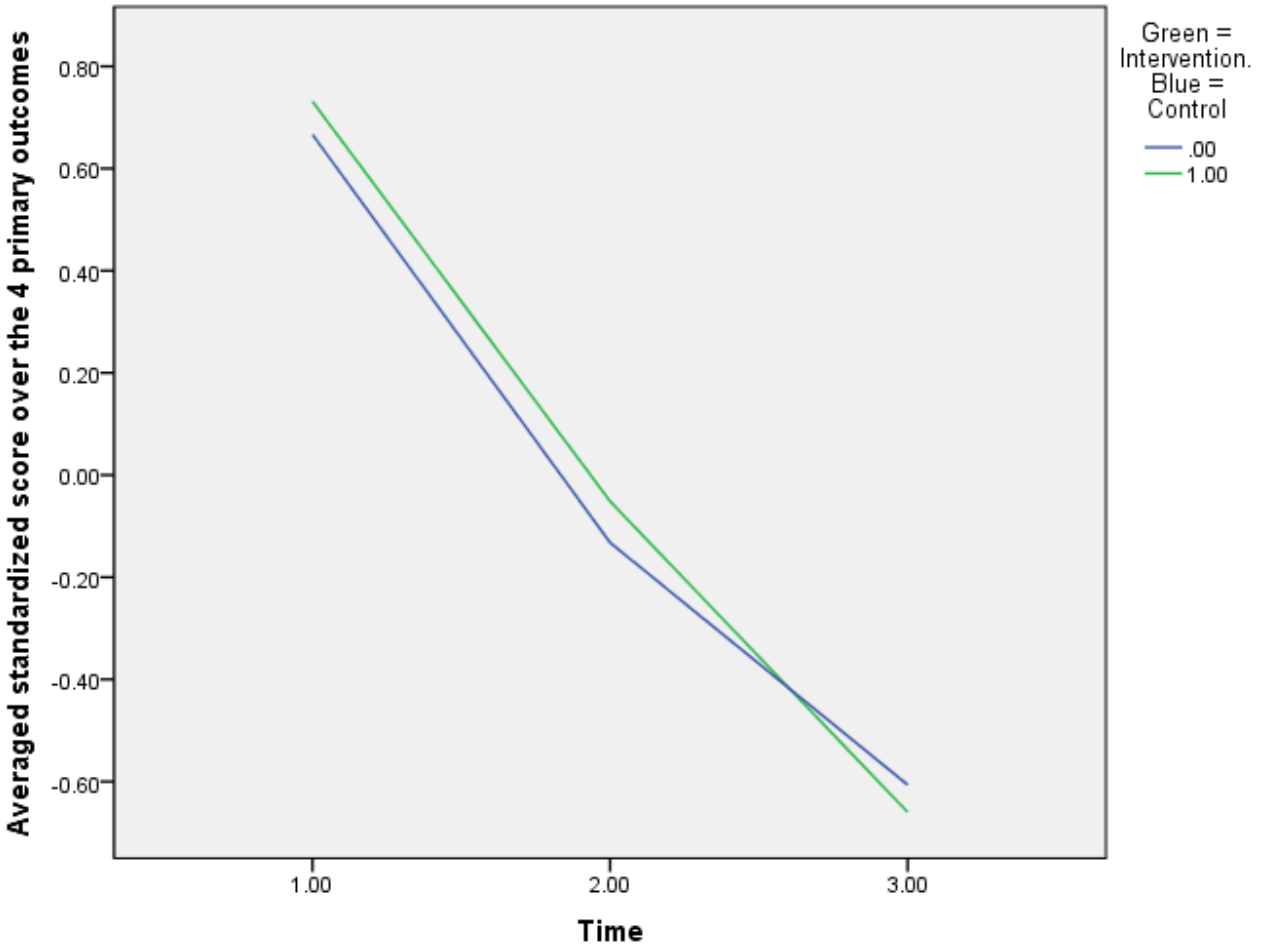
Textbox 2. Specified multivariate linear mixed model of standardized primary outcome scores.

A predicted (standardized) value for an individual patient on any of the 4 primary outcomes at a certain point in time is calculated as the sum of a random intercept regarding the outcome type (1=pain intensity, 2=fatigue, 3=catastrophizing, 4=psychological distress) and fixed-effect parameter estimates for the following:

- intervention group membership (1=intervention, 2=control [reference category]);
- time 0-1 (baseline=1 relative to intermediate=3 [reference category]);
- time 1-2 (post=2, relative to intermediate=3);
- socioeconomic status (SES) multiplied by the SES score;
- being returned to work at intermediate assessment;
- interaction between intervention and time 0-1; and
- interaction between intervention and time 1-2.

XSL•FO
RenderX

Figure 2. Patterns of primary outcome change during rehabilitation between the groups.



From the intermediate to posttreatment assessment, 48.7% (73/150) and 2.7% (4/150) of the patients in the intervention group reported reliable improvement and decline, respectively, in one or more primary outcomes (Figure 3). In the control group, these proportions were 40.7% (48/118) and 7.6% (9/118). Furthermore, proportional distributions of reliable improvement, stability, and deterioration were not different between the groups (χ^2_1=5.677, *P*=.06).

Second, univariate tests for a hypothesized group effect on changes in unstandardized CIS concentration problems and SCL depression and insufficiency scores (Table 5) resulted in a two-sided *P* value below the adjusted Holm-Sidak criterion level (alpha<.017) only for a comparatively stronger decrease in intermediate to post-SCL depressive symptom scores for the intervention group (unstandardized regression coefficient *b*=−2.74, *P*=.011).

Observations on secondary outcomes showed generally high scores for PGIC, general health, functioning (distance perceived relative to before the onset of pain or fatigue complaints and current satisfaction), and treatment satisfaction ratings (Table 6). Moreover, no group differences in secondary outcome variables were observed at posttreatment or in change since the intermediate assessment. A summary of developments in primary and secondary outcomes throughout the second part of the rehabilitation program is presented in Multimedia Appendix 2.

Finally, log data within the intervention group showed that 1 patient logged in but did not play the game, 12 played up to 50% of the game, 24 played 50%-75% of the game, 110 played 75%-100% of the game, and 9 patients continued to play a second time. Of all, 54.7% (87/156) of the patients completed 16 “encounters,” which equals to completing the game precisely once. Notably, completed encounters averaged 14.5 and ranged from 0 to 28. Among patients who did not finish the game (60/156, 38.5%), relatively few had reported completing tertiary education (5/21, 23.5%; χ^2_3=10.075, *P*=.02) or previously receiving specialist care (33/60, 55.0%; χ^2_1=4.23, *P*=.04). A debriefing session was attended by 151 patients. Groups with low adherence were too small to provide valid efficacy estimates within each.

XSL•FO
RenderX

Figure 3. Within-group proportions for reliable improvement or decline in one or more of the 4 primary outcomes between the groups.

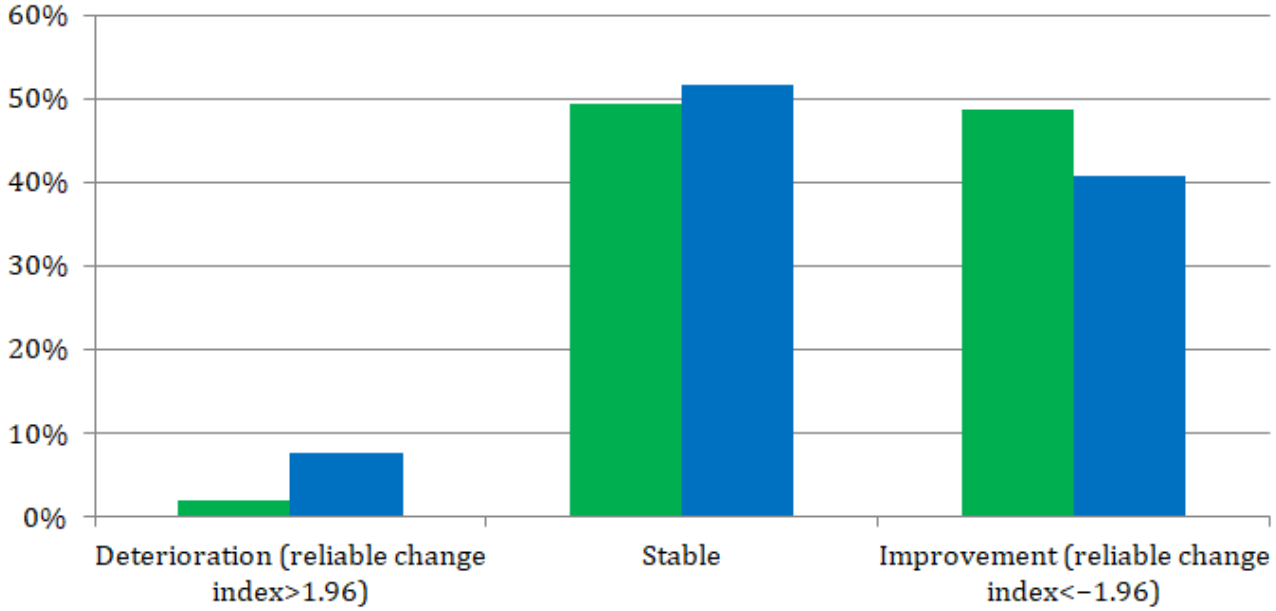


Table 5. Univariate linear mixed modeling results.

Subscale used and time[a]	Intervention group		Control group		Effect	Unstandardized regression coefficient *b* (SE)	*P* value[b]
	Number of observations	Mean (SD)	Number of observations	Mean (SD)			
Symptoms checklist depression subscale							
t0	156	40.49 (12.22)	119	39.87 (12.69)	t0-1[c]	−8.50 (0.92)	<.001
					t1-2	−7.20 (0.71)	<.001
t1	156	31.99 (11.31)	119	28.89 (8.94)	t0-1 × X[d]	2.52 (1.40)	.07
t2	149	24.85 (8.64)	117	24.50 (7.95)	t1-2 × X	−2.75 (1.07)	.01
Symptoms Checklist insufficiency subscale							
t0	156	24.63 (6.68)	119	23.96 (7.15)	t0-1	−4.5 (0.48)	<.001
					t1-2	−4.0 (0.43)	<.001
t1	156	20.1 (6.05)	119	18.62 (5.55)	t0-1 × X	.82 (0.76)	.27
t2	149	16.11 (5.57)	117	15.97 (5.41)	t1-2 × X	−1.36 (0.65)	.04
Checklist Individual Strength concentration problems subscale							
t0	154	26.69 (7.62)	118	24.71 (7.34)	t0-1	−4.84 (0.70)	<.001
					t1-2	−5.96 (0.55)	<.001
t1	154	21.85 (7.80)	118	21.03 (7.20)	t0-1 × X	−1.25 (1.06)	.24
t2	147	15.84 (7.88)	116	16.27, (7.24)	t1-2 × X	−1.17 (0.83)	.16

[a]t0=measurement; t1=intermediate (after 8 weeks of treatment); t2=posttreatment (after 16 weeks of treatment).

[b]Sidak-Holm-corrected alpha criterion levels were applied to the 3 primary outcomes, being .02 for the lowest *P* value, .03 for the second highest *P* value, and .05 for the highest *P* value.

[c]t1 parameters in this table are multiplied by −1 because t1 (index=3) was the reference category.

[e]X: intervention group.

XSL•FO
RenderX

Table 6. Secondary outcomes by group and time (t1=intermediate [after 8 weeks of treatment]; t2=posttreatment [after 16 weeks of treatment]).

Outcome	Intervention group		Control group		Δ Group (by time)[a]	
	t1 (n=156)	t2 (n=150)	t1 (n=119)	t2 (n=118)	Student *t* or χ^2 test	*P* value
Patient global impression of change, n (%)					χ^2_6=3.3	.77
Much deteriorated	1 (0.1)	0 (0.0)	1 (0.1)	1 (0.1)		
Slightly deteriorated	5 (3.2)	4 (2.8)	2 (1.7)	5 (4.3)		
Stable	16 (10.3)	10 (6.9)	25 (21.0)	6 (5.2)		
Slightly improved	97 (62.2)	46 (31.7)	63 (52.9)	41 (35.3)		
Much improved	37 (13.5)	85 (58.6)	28 (23.5)	63 (54.3)		
Subjective health and functioning, mean (SD)					t_{266}=−1.16	.25[b]
General health	55.45 (24.18)	71.23 (22.57)	57.97 (23.17)	71.90 (20.39)		
Functioning “level”	46.05 (25.81)	70.19 (25.22)	50.02 (25.66)	69.78 (24.53)		
Functioning “distance”	40.12 (25.34)	55.47 (32.10)	42.04 (27.65)	52.85 (31.37)		
Treatment satisfaction, mean (SD)						
Rating program	N/A[c]	8.33 (1.20)	N/A	8.06 (1.46)	t_{266}=1.65	.10
Recommend program to other patients, n (%)					χ^2_2=4.8	.09
Certainly not	N/A	0 (0.0)	N/A	0 (0.0)		
Probably not	N/A	4 (2.7)	N/A	9 (7.6)		
Probably yes	N/A	40 (26.7)	N/A	37 (31.4)		
Certainly yes	N/A	106 (70.7)	N/A	72 (61.0)		
Expectations met, n (%)					χ^2_3=4.67	.20
Not at all	N/A	0 (0.0)	N/A	3 (2.5)		
Mostly not	N/A	14 (9.3)	N/A	12 (10.2)		
Mostly	N/A	84 (56.0)	N/A	58 (49.2)		
Completely	N/A	52 (34.7)	N/A	45 (16.8)		

[a]If data were available at intermediate and posttreatment, group differences were assessed in change from intermediate to posttreatment.

[b]Group differences were tested in a change of the average scores over the sums of three items (that together formed an internally consistent scale); similar results were obtained if MIXED or repeated measures analysis of variance was used.

[c]N/A: not applicable.

Discussion

Summary of Evidence

In this study, we aimed to determine to what extent and in what respect a novel 4-hour mindfulness-based serious gaming intervention is effective in facilitating additional change in relevant physical and emotional functioning outcomes during a regular multidisciplinary rehabilitation for patients with CP or FSS. Furthermore, we studied whether such effects have clinical relevance for health improvement as conceived by patients themselves and whether these effects depend on the varying adherence within a regular care setting. Patients with mainly (low) back pain with comorbid psychosocial problems were found to adhere well to additional serious gaming during regular multidisciplinary rehabilitation, resulting in a very small (merely statistical) strengthening effect on the reduction of physical and emotional symptoms, as a whole, and of depressive symptoms, in particular. The effect of serious gaming alone, as a relatively small additional program component, did not reach clinically relevant levels; this was also suggested because patient impressions of health change and treatment satisfaction showed no improvement compared with the regular program, which already showed high satisfaction and treatment success rates. Nonetheless, within this context of multidisciplinary rehabilitation, 4 additional hours planned for serious gaming (4% of therapy time) in small groups, largely without direct professional support, accounted for 8.9% of the total average primary outcome change for the intervention group during rehabilitation.

Several insights arise from relating these results to those of previous studies on similar interventions. First, the very small effect size found in this study suggests a relatively weak effect compared with the small effect sizes found in previous studies. Those studies included evaluations of the effect of exposure to games on health outcomes with pragmatic trial designs [59,60], as well as systematic reviews and meta-analyses of randomized

XSL•FO
RenderX

controlled studies on the efficacy of games for various clinical and behavioral outcomes [29,34]. Plausible explanations for a lower estimate in this study are the relatively low intensity and late supply of serious gaming relative to other efficacious psychotherapy (including mindfulness) interventions offered through other modalities of multidisciplinary rehabilitation. For any such short-term component in multidisciplinary rehabilitation, an effect large enough to be generally noticeable to patients would be extraordinary within the target population, and many patients may already have benefited from "traditional" means to improve. Second, both present and earlier findings suggest that changes with mindfulness approaches occur simultaneously across outcomes [61]. Still, our quantitative and qualitative results combined also add specifically to the anecdotal evidence from previous randomized trials that depressive symptoms are a plausible target for serious gaming [60,62]. Third, our findings indicate a possible relative efficiency of the independent usage and guidance in groups that constitute a "blended" form of serious gaming. To illustrate, the effect size estimate found in this study approaches the estimates found in a previous meta-analysis on the outcomes of computer (internet)-supported therapy across chronic somatic conditions (standardized mean difference ranging between 0.17 and 0.21 across outcomes) [63]. Therefore, the results of this study indicate, but do not prove, that serious gaming could serve as a complement or substitute to (parts of) other sorts of computer-based or blended treatments aimed at allocating scarce professional guidance more efficiently. Finally, previous authors doubted whether the adherence and efficacy of computer-based interventions are readily transferable to contexts, as in this study, wherein patients are recruited from a clinical setting instead of being openly recruited from general populations via the internet or other media [64]. This study sheds light onto this transferability issue by showing that a relatively high level of adherence can be achieved within a regular health care context where self-selection for the modality is limited, when a serious gaming supplement is offered "by default," based on understandings of usage factors [65].

Strengths and Limitations

Strengths of this study relate to the novelty of the serious gaming approach, statistical power, and the apparently favorable conditions for pragmatic research. This evaluation addresses a unique combination of setting, patient, and intervention characteristics (mindfulness approach and blended mode of supply). Achieving the predetermined required sample size for observing a modest effect with reasonable chance responds to previous reviews on the effectiveness of games for health that found promising results for mainly small, underpowered studies [34,39]. Furthermore, this study has taken account of Type I error risk through outcome multiplicity, factors of nonusage, and risk of biased patient expectations through the informed consent procedure. In the execution of the study, we encountered occasional unintended difficulties in accessing the game (ie, forgetting passwords) but did not encountered problems or threats to internal validity, besides those inevitable and known in the protocol phase. The precision of key results was supported by the results of sensitivity analyses after outlier removals, alternate outcome standardization, removal of incomplete cases due to treatment dropout, and extensions and simplifications (eg, exclusion of baseline data, inclusion or exclusion of potential confounding variables) of the prediction model. Mixing quantitative and qualitative outcome data led to unambiguous findings regarding the size, outcome domains, and the clinical relevance of serious game effects. Regarding the external validity, this application of pragmatic methods adds complementary insight into the effectiveness of serious gaming for patients in regular health care settings beyond controlled clinical trial conditions. The inclusive patient recruitment strategy reflects the reality of a regular care setting to which the results are to be generalized.

However, several study weaknesses should be considered, comparing this study with supposedly ideal circumstances for a randomized controlled (multicenter) trial. Not applied for practical reasons were broader recruitment of treatment settings, researcher control on selection procedures, the use of an individual or site-level randomization procedure for balancing unobserved characteristics between study groups, inclusion of certain measures (long-term follow-up, objective outcomes, functional interference, quality of life, and participation), and collections of cost data. In addition, intervention group participants were aware that they received a novel treatment component. However, this is not expected to have influenced the results as an insignificant association was observed between outcome expectations of serious gaming and health outcome change levels (intermediate to posttreatment) within the intervention group. Adding an additional component to an already intensive treatment program has neither been deemed likely nor intended to increase cost-effectiveness at present, but may offer useful insight for achieving this in the future. Although previous studies have suggested that effects of serious games for behavioral change are retained [29], it remains uncertain how a very small reinforcing influence on patterns of outcome change that started earlier during treatment will develop further in time. Furthermore, a lack of more stringent diagnostic methods at inclusion poses an internal validity threat. Data are also missing about characteristics of patients who dropped out during the first part of the program. Moreover, present results suggest that (everything else being equal) additional serious gaming adds very little to the outcome improvement, but intervention group participants did not reach more favorable outcome levels at posttreatment. A possible explanation is that control group symptom levels were slightly lower overall because of an effect of recruitment that was too small to observe. Besides, not all expected outcome domains found through the qualitative research were confirmed with quantitative results; this might also be attributed to a lack of power as precalculations have not been based on an increasing number of statistical tests. Finally, generalizability is limited by the convenient selection of 4 locations from a single Dutch care center.

Suggestions for Research and Practice

In light of previous research, the very small positive effect on relevant outcomes found in this powerful pragmatic study reaffirms that both caution and optimism about the effectiveness of serious games as a treatment facilitator are warranted. Findings imply that serious gaming holds potential, as for the present mindfulness-based approach to it, but requires further

XSL•FO
RenderX

investigation before wider dissemination within multidisciplinary rehabilitation programs or other regular health care settings (eg, psychological therapy). From patients' point, expectations on potential benefits are to be placed in perspective, that is, results merely suggest that multidisciplinary rehabilitation based on a biopsychosocial approach (ie, one that includes mindfulness approaches to learning to live with CP or FSS) with generally modest effects (ie, offer little assurance for recovery) could be somewhat improved (in a slight, merely statistical, degree) by adding serious gaming as a modality. We do not suggest that additional serious gaming causes more patients to experience clinically relevant treatment effects. Nonetheless, the study results do suggest that the delivery of a small part of an evidence-based treatment by means of a serious game can be trusted. Therefore, researchers should continue to pursue adequately powered and, if possible, RCTs when aiming to assess the effects of (mindfulness-based) serious gaming.

As part of a general search for effective combinations of approaches, techniques, and modalities to intensive rehabilitation programs, the serious gaming approach requires further theoretical refinement as to know how and when clinically relevant benefits are achieved by which patients and why. The current state of evidence provides little support as to identify those circumstances in which patients with CP or FSS will likely have best experiences and outcomes from which (computer-based) biopsychosocial or alternative treatments and why [10,18,66,67]. In this regard, our findings specifically point toward very small positive effects when (mindfulness-based) serious gaming is presented later on in a rehabilitation process to patients with chronic back pain and comorbid psychosocial problems. Patients, policy makers, and professionals must be aware of the ongoing developmental stage, wherein the accumulation of knowledge is needed before the full potential of serious gaming can be realized routinely and efficiently into complex health care systems [29,68]. Thus, to achieve the highest potential of serious gaming for health, more theoretically oriented and context-sensitive studies are needed in addition to more powerful outcome assessment trials. To facilitate progress, researchers need to focus on a broad range of research questions about when and which kinds of serious games are (cost-)effective, for whom, compared with other treatment options, and why. This endeavor requires (1) hypotheses-driven process evaluations alongside trials (using quantitative, qualitative, or mixed methods); (2) transparent and universal reporting on the qualities of the methodology (ie, eHealth CONSORT statement additions [69]), serious games for health (rationale, functionality, and data security [70]), and behavioral change content (theoretical approaches, change strategies, and presentation methods [71]) in trials, (3) implementation research investigating organizational, professional, patient, and intervention factors; (4) impact assessment as dependent on actual reach; and (5) health technology assessments.

Conclusions

Based on a powerful natural quasi-experiment, the results of this study suggest that serious gaming, as an additional modality for mindfulness intervention of short duration during regular multidisciplinary rehabilitation, adds very little to reducing physical and psychological symptoms in patients with CP or FSS (ie, indicated with chronic back pain and concomitant psychosocial problems). In addition, the results hint, but cannot yet prove, that these very small benefits are nonetheless relevant in terms of efficiency if one considers how little (extra) time it costs from scarce expert care providers. An effect with respect to depressive mood may exist that a minority of patients conceive as relevant for their daily life. Moreover, the findings clearly support a generally good adherence to a blended form of serious gaming in a regular care setting. Taken into account the conditions of serious gaming in this study (ie, relatively low intensity compared with the complete treatment program that patients received), the results fit the expectations created by previous studies that generally found slightly higher (small) effects on behavioral and clinical outcomes (ie, studies on serious games in various populations or studies on computer-based interventions in patients with CP or FSS). Therefore, the potential of serious games for being effective in changing behavioral and clinical outcomes across targeted populations is reaffirmed and further (theory-driven) research on serious gaming aimed at predictably (cost-)effective applications for individual patients across health care settings encouraged.

Acknowledgments

This work was supported by Ciran (the owner of LAKA) regarding the development and allocation of the serious gaming intervention and provision of raw quantitative data. Several partners made essential contributions to LAKA, as acknowledged in the published evaluation protocol. Qualitative data collection, data management, data analyses, interpretation of results, writing of the report, and publication decisions are authorized by university staff members.

Conflicts of Interest

Ciran was the employer of MAPV and AM during the execution of the evaluation and paid personal fees to HJMV outside the submitted work. MAPV was provided time and occasion to conduct independent doctoral research, via agreement, at Tranzo, Scientific Center for Care and Welfare. The terms of this arrangement have been reviewed and approved by Tranzo in accordance with its policy on objectivity in research. MCWJ and AMEZ have nothing to declare.

Multimedia Appendix 1

Screenshots and trailer.

XSL•FO
RenderX

[PDF File (Adobe PDF File), 1MB - jmir_v20i8e250_app1.pdf]

Multimedia Appendix 2

Change in primary and secondary outcomes throughout the second part of rehabilitation.

[PDF File (Adobe PDF File), 373KB - jmir_v20i8e250_app2.pdf]

References

1. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. Eur J Pain 2006 May;10(4):287-333. [doi: 10.1016/j.ejpain.2005.06.009] [Medline: 16095934]
2. Gatchel RJ, Peng YB, Peters ML, Fuchs PN, Turk DC. The biopsychosocial approach to chronic pain: scientific advances and future directions. Psychol Bull 2007 Jul;133(4):581-624. [doi: 10.1037/0033-2909.133.4.581] [Medline: 17592957]
3. Henningsen P, Zipfel S, Herzog W. Management of functional somatic syndromes. Lancet 2007 Mar 17;369(9565):946-955. [doi: 10.1016/S0140-6736(07)60159-7] [Medline: 17368156]
4. Jackson T, Thomas S, Stabile V, Shotwell M, Han X, McQueen K. A Systematic Review and Meta-Analysis of the Global Burden of Chronic Pain Without Clear Etiology in Low- and Middle-Income Countries: Trends in Heterogeneous Data and a Proposal for New Assessment Methods. Anesth Analg 2016 Sep;123(3):739-748. [doi: 10.1213/ANE.0000000000001389] [Medline: 27537761]
5. Fayaz A, Croft P, Langford RM, Donaldson LJ, Jones GT. Prevalence of chronic pain in the UK: a systematic review and meta-analysis of population studies. BMJ Open 2016 Jun 20;6(6):e010364 [FREE Full text] [doi: 10.1136/bmjopen-2015-010364] [Medline: 27324708]
6. Eslick GD, Jones MP, Talley NJ. Non-cardiac chest pain: prevalence, risk factors, impact and consulting--a population-based study. Aliment Pharmacol Ther 2003 May 01;17(9):1115-1124 [FREE Full text] [Medline: 12752348]
7. McCormack A, Edmondson-Jones M, Somerset S, Hall DA. Corrigendum to “A systematic review of the reporting of tinnitus prevalence and severity” [Hear. Res. 337 (2016) 70-79]. Hear Res 2016 Sep;339:219 [FREE Full text] [doi: 10.1016/j.heares.2016.08.010] [Medline: 27612990]
8. Fass R, Achem SR. Noncardiac chest pain: epidemiology, natural course and pathogenesis. J Neurogastroenterol Motil 2011 Apr;17(2):110-123 [FREE Full text] [doi: 10.5056/jnm.2011.17.2.110] [Medline: 21602987]
9. Johnston S, Brenu EW, Staines D, Marshall-Gradisnik S. The prevalence of chronic fatigue syndrome/ myalgic encephalomyelitis: a meta-analysis. Clin Epidemiol 2013;5:105-110 [FREE Full text] [doi: 10.2147/CLEP.S39876] [Medline: 23576883]
10. Vos-Vromans DCWM, Smeets RJEM, Huijnen IPJ, Köke AJA, Hitters WMGC, Rijnders LJM, et al. Multidisciplinary rehabilitation treatment versus cognitive behavioural therapy for patients with chronic fatigue syndrome: a randomized controlled trial. J Intern Med 2016 Mar;279(3):268-282. [doi: 10.1111/joim.12402] [Medline: 26306716]
11. Veehof MM, Oskam M, Schreurs KMG, Bohlmeijer ET. Acceptance-based interventions for the treatment of chronic pain: a systematic review and meta-analysis. Pain 2011 Mar;152(3):533-542 [FREE Full text] [doi: 10.1016/j.pain.2010.11.002] [Medline: 21251756]
12. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
13. Wentzel J, van der Vaart R, Bohlmeijer ET, van Gemert-Pijnen JE. Mixing Online and Face-to-Face Therapy: How to Benefit From Blended Care in Mental Health Care. JMIR Ment Health 2016;3(1):e9 [FREE Full text] [doi: 10.2196/mental.4534] [Medline: 26860537]
14. Eccleston C, Fisher E, Craig L, Duggan GB, Rosser BA, Keogh E. Psychological therapies (Internet-delivered) for the management of chronic pain in adults. Cochrane Database Syst Rev 2014;2:CD010152. [doi: 10.1002/14651858.CD010152.pub2] [Medline: 24574082]
15. Vugts MAP, Joosen MCW, Mert A, Zedlitz A, Vrijhoef HJM. Serious gaming during multidisciplinary rehabilitation for patients with complex chronic pain or fatigue complaints: study protocol for a controlled trial and process evaluation. BMJ Open 2017 Jun 08;7(6):e016394 [FREE Full text] [doi: 10.1136/bmjopen-2017-016394] [Medline: 28600377]
16. van der Feltz-Cornelis CM, Hoedeman R, Keuter EJW, Swinkels JA. Presentation of the Multidisciplinary Guideline Medically Unexplained Physical Symptoms (MUPS) and Somatoform Disorder in the Netherlands: disease management according to risk profiles. J Psychosom Res 2012 Feb;72(2):168-169. [doi: 10.1016/j.jpsychores.2011.11.007] [Medline: 22281461]
17. Sanders SH, Harden RN, Vicente PJ. Evidence-based clinical practice guidelines for interdisciplinary rehabilitation of chronic nonmalignant pain syndrome patients. Pain Pract 2005 Dec;5(4):303-315. [doi: 10.1111/j.1533-2500.2005.00033.x] [Medline: 17177763]
18. Kamper SJ, Apeldoorn AT, Chiarotto A, Smeets RJEM, Ostelo RWJG, Guzman J, et al. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain. Cochrane Database Syst Rev 2014 Sep 02(9):CD000963. [doi: 10.1002/14651858.CD000963.pub3] [Medline: 25180773]

XSL•FO
RenderX

19. Marques MM, De Gucht V, Gouveia MJ, Leal I, Maes S. Differential effects of behavioral interventions with a graded physical activity component in patients suffering from Chronic Fatigue (Syndrome): An updated systematic review and meta-analysis. Clin Psychol Rev 2015 Aug;40:123-137. [doi: 10.1016/j.cpr.2015.05.009] [Medline: 26112761]
20. Gatchel RJ, McGeary DD, McGeary CA, Lippe B. Interdisciplinary chronic pain management: past, present, and future. Am Psychol 2014;69(2):119-130. [doi: 10.1037/a0035514] [Medline: 24547798]
21. Vago DR, Silbersweig DA. Self-awareness, self-regulation, and self-transcendence (S-ART): a framework for understanding the neurobiological mechanisms of mindfulness. Front Hum Neurosci 2012;6:296 [FREE Full text] [doi: 10.3389/fnhum.2012.00296] [Medline: 23112770]
22. Macea DD, Gajos K, Daglia CYA, Fregni F. The efficacy of Web-based cognitive behavioral interventions for chronic pain: a systematic review and meta-analysis. J Pain 2010 Oct;11(10):917-929. [doi: 10.1016/j.jpain.2010.06.005] [Medline: 20650691]
23. Lundahl B, Moleni T, Burke BL, Butters R, Tollefson D, Butler C, et al. Motivational interviewing in medical care settings: a systematic review and meta-analysis of randomized controlled trials. Patient Educ Couns 2013 Nov;93(2):157-168. [doi: 10.1016/j.pec.2013.07.012] [Medline: 24001658]
24. Naylor MR, Keefe FJ, Brigidi B, Naud S, Helzer JE. Therapeutic Interactive Voice Response for chronic pain reduction and relapse prevention. Pain 2008 Feb;134(3):335-345 [FREE Full text] [doi: 10.1016/j.pain.2007.11.001] [Medline: 18178011]
25. Kristjánsdóttir Ó, Fors EA, Eide E, Finset A, Stensrud TL, van Dulmen S, et al. A smartphone-based intervention with diaries and therapist feedback to reduce catastrophizing and increase functioning in women with chronic widespread pain. part 2: 11-month follow-up results of a randomized trial. J Med Internet Res 2013 Mar 28;15(3):e72 [FREE Full text] [doi: 10.2196/jmir.2442] [Medline: 23538392]
26. Moessner M, Aufdermauer N, Baier C, Göbel H, Kuhnt O, Neubauer E, et al. [Efficacy of an internet-delivered aftercare program for patients with chronic back pain]. Psychother Psychosom Med Psychol 2014 Feb;64(2):47-53. [doi: 10.1055/s-0033-1351266] [Medline: 24101036]
27. Henrich JF, Knittle K, De Gucht V, Warren S, Dombrowski SU, Maes S. Identifying effective techniques within psychological treatments for irritable bowel syndrome: a meta-analysis. J Psychosom Res 2015 Mar;78(3):205-222. [doi: 10.1016/j.jpsychores.2014.12.009] [Medline: 25579201]
28. Wilshire C, Kindlon T, Matthees A, McGrath S. Can patients with chronic fatigue syndrome really recover after graded exercise or cognitive behavioural therapy? A critical commentary and preliminary re-analysis of the PACE trial. Fatigue: Biomedicine, Health & Behavior 2016 Dec 14;5(1):43-56. [doi: 10.1080/21641846.2017.1259724]
29. DeSmet A, Van Ryckeghem D, Compernolle S, Baranowski T, Thompson D, Crombez G, et al. A meta-analysis of serious digital games for healthy lifestyle promotion. Prev Med 2014 Dec;69:95-107. [doi: 10.1016/j.ypmed.2014.08.026] [Medline: 25172024]
30. Wouters P, van Nimwegen C, van Oostendorp H, van der Spek E. A meta-analysis of the cognitive and motivational effects of serious games. Journal of Educational Psychology 2013;105(2):249-265. [doi: 10.1037/a0031311]
31. Gentile DA, Anderson CA, Yukawa S, Ihori N, Saleem M, Ming LK, et al. The effects of prosocial video games on prosocial behaviors: international evidence from correlational, longitudinal, and experimental studies. Pers Soc Psychol Bull 2009 Jun;35(6):752-763 [FREE Full text] [doi: 10.1177/0146167209333045] [Medline: 19321812]
32. Bailey JV, Murray E, Rait G, Mercer CH, Morris RW, Peacock R, et al. Interactive computer-based interventions for sexual health promotion. Cochrane Database Syst Rev 2010(9):CD006483. [doi: 10.1002/14651858.CD006483.pub2] [Medline: 20824850]
33. Baranowski T, Buday R, Thompson DI, Baranowski J. Playing for real: video games and stories for health-related behavior change. Am J Prev Med 2008 Jan;34(1):74-82 [FREE Full text] [doi: 10.1016/j.amepre.2007.09.027] [Medline: 18083454]
34. Primack BA, Carroll MV, McNamara M, Klem ML, King B, Rich M, et al. Role of video games in improving health-related outcomes: a systematic review. Am J Prev Med 2012 Jun;42(6):630-638 [FREE Full text] [doi: 10.1016/j.amepre.2012.02.023] [Medline: 22608382]
35. Huizinga J. Homo Ludens: Proeve Eener Bepaling Van Het Spel-Element Der Cultuur (Athenaeum Boekhandel Canon) (Dutch Edition). Amsterdam: Amsterdam University Press; 2008.
36. Sutton-Smith B. The Ambiguity of Play. Boston MA: Harvard University Press; 1997.
37. Cerasoli CP, Nicklin JM, Ford MT. Intrinsic motivation and extrinsic incentives jointly predict performance: a 40-year meta-analysis. Psychol Bull 2014 Jul;140(4):980-1008. [doi: 10.1037/a0035661] [Medline: 24491020]
38. Lohse KR, Hilderman CG, Cheung KL, Tatla S, Van der Loos HM. Virtual reality therapy for adults post-stroke: a systematic review and meta-analysis exploring virtual environments and commercial games in therapy. PLoS One 2014;9(3):e93318 [FREE Full text] [doi: 10.1371/journal.pone.0093318] [Medline: 24681826]
39. Li J, Theng Y, Foo S. Game-based digital interventions for depression therapy: a systematic review and meta-analysis. Cyberpsychol Behav Soc Netw 2014 Aug;17(8):519-527. [doi: 10.1089/cyber.2013.0481] [Medline: 24810933]
40. Horne-Moyer HL, Moyer BH, Messer DC, Messer ES. The use of electronic games in therapy: a review with clinical implications. Curr Psychiatry Rep 2014 Dec;16(12):520 [FREE Full text] [doi: 10.1007/s11920-014-0520-6] [Medline: 25312026]

XSL•FO
RenderX

41. Malloy KM, Milling LS. The effectiveness of virtual reality distraction for pain reduction: a systematic review. Clin Psychol Rev 2010 Dec;30(8):1011-1018. [doi: 10.1016/j.cpr.2010.07.001] [Medline: 20691523]
42. Lumsden J, Edwards EA, Lawrence NS, Coyle D, Munafò MR. Gamification of Cognitive Assessment and Cognitive Training: A Systematic Review of Applications and Efficacy. JMIR Serious Games 2016;4(2):e11 [FREE Full text] [doi: 10.2196/games.5888] [Medline: 27421244]
43. Crookall D. Serious Games, Debriefing, and Simulation/Gaming as a Discipline. Simulation & Gaming 2011 Jan 06;41(6):898-920 [FREE Full text] [doi: 10.1177/1046878110390784] [Medline: 22973420]
44. Kharrazi H, Lu AS, Gharghabi F, Coleman W. A Scoping Review of Health Game Research: Past, Present, and Future. Games Health J 2012 Apr 18;1(2):153-164 [FREE Full text] [doi: 10.1089/g4h.2012.0011] [Medline: 24416638]
45. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008;337:a1655 [FREE Full text] [Medline: 18824488]
46. Kim B, Park H, Baek Y. Not just fun, but serious strategies: Using meta-cognitive strategies in game-based learning. Computers & Education 2009 May;52(4):800-810 [FREE Full text] [doi: 10.1016/j.compedu.2008.12.004] [Medline: 25904163]
47. Przybylski AK, Weinstein N, Ryan RM, Rigby CS. Having to versus wanting to play: background and consequences of harmonious versus obsessive engagement in video games. Cyberpsychol Behav 2009 Oct;12(5):485-492. [doi: 10.1089/cpb.2009.0083] [Medline: 19772442]
48. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008 Feb;9(2):105-121. [doi: 10.1016/j.jpain.2007.09.005] [Medline: 18055266]
49. Garschagen A, Steegers MA, van Bergen AH, Jochijms JA, Skrabanja TL, Vrijhoef HJ, et al. Is There a Need for Including Spiritual Care in Interdisciplinary Rehabilitation of Chronic Pain Patients? Investigating an Innovative Strategy. Pain Pract 2015 Sep;15(7):671-687. [doi: 10.1111/papr.12234] [Medline: 25229884]
50. Lutz A, Slagter HA, Dunne JD, Davidson RJ. Attention regulation and monitoring in meditation. Trends Cogn Sci 2008 Apr;12(4):163-169 [FREE Full text] [doi: 10.1016/j.tics.2008.01.005] [Medline: 18329323]
51. Dittner AJ, Wessely SC, Brown RG. The assessment of fatigue: a practical guide for clinicians and researchers. J Psychosom Res 2004 Feb;56(2):157-170. [doi: 10.1016/S0022-3999(03)00371-4] [Medline: 15016573]
52. Arrindell W, Ettema J. SCL-90: Handleiding bij een multidimensionele psychopathologie-indicator. Lisse: Swets test publishers; 2003.
53. Price DD, Bush FM, Long S, Harkins SW. A comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales. Pain 1994 Feb;56(2):217-226. [Medline: 8008411]
54. Spinhoven P, Ter KM, Kole-Snijders AMJ, Hutten MM, Den OD, Vlaeyen JWS. Catastrophizing and internal pain control as mediators of outcome in the multidisciplinary treatment of chronic low back pain. Eur J Pain 2004 Jun;8(3):211-219. [doi: 10.1016/j.ejpain.2003.08.003] [Medline: 15109971]
55. Zuidgeest M, Reitsma-van Rooijen R, Sixma H, van der Hoek L, Rademakers J. CQ-index Revalidatiecentra: meetinstrumentontwikkeling en bepaling van het discriminerend vermogen. Utrecht: Nivel; 2010.
56. Holm S. A simple sequentially rejective multiple test procedure. Scandinavian journal of statistics 1979:65-70.
57. Heck R, Thomas S, Tabata L. Multilevel and Longitudinal Modeling with IBM SPSS. London: Taylor & Francis; 2013.
58. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [Medline: 2002127]
59. Maddison R, Foley L, Ni MC, Jiang Y, Jull A, Prapavessis H, et al. Effects of active video games on body composition: a randomized controlled trial. Am J Clin Nutr 2011 Jul;94(1):156-163 [FREE Full text] [doi: 10.3945/ajcn.110.009142] [Medline: 21562081]
60. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MFG. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. BMJ 2012;344:e2598 [FREE Full text] [Medline: 22517917]
61. Turner JA, Anderson ML, Balderson BH, Cook AJ, Sherman KJ, Cherkin DC. Mindfulness-based stress reduction and cognitive behavioral therapy for chronic low back pain: similar effects on mindfulness, catastrophizing, self-efficacy, and acceptance in a randomized controlled trial. Pain 2016 Nov;157(11):2434-2444 [FREE Full text] [doi: 10.1097/j.pain.0000000000000635] [Medline: 27257859]
62. Roepke AM, Jaffee SR, Riffle OM, McGonigal J, Broome R, Maxwell B. Randomized Controlled Trial of SuperBetter, a Smartphone-Based/Internet-Based Self-Help Tool to Reduce Depressive Symptoms. Games Health J 2015 Jun;4(3):235-246. [doi: 10.1089/g4h.2014.0046] [Medline: 26182069]
63. van Beugen S, Ferwerda M, Hoeve D, Rovers MM, Spillekom-van Koulil S, van Middendorp H, et al. Internet-based cognitive behavioral therapy for patients with chronic somatic conditions: a meta-analytic review. J Med Internet Res 2014 Mar 27;16(3):e88 [FREE Full text] [doi: 10.2196/jmir.2777] [Medline: 24675372]
64. Chiauzzi E, Pujol LA, Wood M, Bond K, Black R, Yiu E, et al. painACTION-back pain: a self-management website for people with chronic back pain. Pain Med 2010 Jul;11(7):1044-1058. [doi: 10.1111/j.1526-4637.2010.00879.x] [Medline: 20545873]

XSL•FO
RenderX

65. Vugts MA, Joosen MC, van Bergen AH, Vrijhoef HJ. Feasibility of Applied Gaming During Interdisciplinary Rehabilitation for Patients With Complex Chronic Pain and Fatigue Complaints: A Mixed-Methods Study. JMIR Serious Games 2016 Apr 01;4(1):e2 [FREE Full text] [doi: 10.2196/games.5088] [Medline: 27036094]
66. Vickers AJ, Cronin AM, Maschino AC, Lewith G, MacPherson H, Foster NE, Acupuncture Trialists' Collaboration. Acupuncture for chronic pain: individual patient data meta-analysis. Arch Intern Med 2012 Oct 22;172(19):1444-1453 [FREE Full text] [doi: 10.1001/archinternmed.2012.3654] [Medline: 22965186]
67. Hilton L, Hempel S, Ewing BA, Apaydin E, Xenakis L, Newberry S, et al. Mindfulness Meditation for Chronic Pain: Systematic Review and Meta-analysis. Ann Behav Med 2017 Apr;51(2):199-213 [FREE Full text] [doi: 10.1007/s12160-016-9844-2] [Medline: 27658913]
68. Fleming TM, de Beurs D, Khazaal Y, Gaggioli A, Riva G, Botella C, et al. Maximizing the Impact of e-Therapy and Serious Gaming: Time for a Paradigm Shift. Front Psychiatry 2016;7:65 [FREE Full text] [Medline: 27148094]
69. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
70. Graafland M, Dankbaar M, Mert A, Lagro J, De Wit-Zuurendonk L, Schuit S, et al. How to systematically assess serious games applied to health care. JMIR Serious Games 2014;2(2):e11 [FREE Full text] [doi: 10.2196/games.3825] [Medline: 25654163]
71. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]

Abbreviations

CIS: Checklist Individual Strength
CP: chronic pain
FSS: functional somatic syndromes
NPC: nonplaying character
PGIC: patient's global impression of change
RCI: reliable change index
RCT: randomized controlled trial
SCL-90: Symptoms Checklist
SES: socioeconomic status

Edited by G Eysenbach; submitted 27.12.17; peer-reviewed by H Thabrew, H Eide; comments to author 15.03.18; revised version received 30.04.18; accepted 18.06.18; published 15.08.18.

Please cite as:
Vugts MAP, Joosen MCW, Mert A, Zedlitz AME, Vrijhoef HJM
Effectiveness of Serious Gaming During the Multidisciplinary Rehabilitation of Patients With Complex Chronic Pain or Fatigue: Natural Quasi-Experiment
J Med Internet Res 2018;20(8):e250
URL: http://www.jmir.org/2018/8/e250/
doi:10.2196/jmir.9739
PMID:30111527

©Miel AP Vugts, Margot CW Joosen, Agali Mert, Aglaia ME Zedlitz, Hubertus JM Vrijhoef. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 15.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Informal Caregivers' Experiences and Perceptions of a Web-Based Peer Support Network: Mixed-Methods Study

Christine Vaughan[1], PhD; Thomas E Trail[2], PhD; Ammarah Mahmud[2], MPH; Stephanie Dellva[2], MHA; Terri Tanielian[2], MA; Esther Friedman[3], PhD

[1]RAND Corporation, San Francisco, CA, United States
[2]RAND Corporation, Arlington, VA, United States
[3]RAND Corporation, Santa Monica, CA, United States

Corresponding Author:
Christine Vaughan, PhD
RAND Corporation
1885 Mission Street
San Francisco, CA, 94103
United States
Phone: 1 310 393 0411 ext 6068
Email: cvaughan@rand.org

Abstract

Background: Web-based peer support interventions have shown promise in reducing social isolation and social support deficits among informal caregivers, but little research has examined how caregivers use and perceive these interventions.

Objective: In this study, we examined utilization and perceptions of a Web-based social support intervention for informal caregivers of wounded, ill, and injured United States military service members and veterans.

Methods: This was a mixed-methods study that used quantitative survey data and qualitative data from focus groups and interviews with informal caregivers enrolled in a Web-based peer support intervention to explore their use and perceptions of the intervention. The intervention was delivered via a website that featured interest groups organized around specific topics, webinars, webchats, and messaging functionality and was moderated by professionally trained peers. This study occurred in the context of a quasi-experimental outcome evaluation of the intervention, where intervention participants were compared with a group of military caregivers who were not enrolled in the intervention.

Results: Survey findings indicated that caregivers used the website infrequently, with 60.7% (128/211) visiting the website once a month or less, and passively, with a minority (32/144, 22.2%) of users (ie, those who had visited the website at least once during the past 3 months, N=144) posting comments or links to the network. Nonetheless, most users (121/144, 84.0%) endorsed moderate or greater satisfaction with the website on the survey, and focus group and interview participants reported benefiting sufficiently from passive use of the website (eg, reading posts). Quantitative and qualitative findings suggested that users viewed the website primarily as a source of informational support. Among 63.2% (91/144) of users who completed the survey, the most commonly reported network-related activity was obtaining information from the network's resource library, and focus group and interview participants viewed the network primarily as an informational resource. Focus group and interview participants expressed an unmet need for emotional support and the desire for a more *personal touch* in the forms of more active engagement with other caregivers in the network and the creation of local, in-person support groups for caregivers.

Conclusions: These findings suggest that Web-based peer support interventions may lend themselves better to the provision of informational (vs emotional) support and may need to be supplemented by in-person peer support groups to better meet caregivers' needs for emotional support.

(J Med Internet Res 2018;20(8):e257) doi:10.2196/jmir.9895

KEYWORDS

caregivers; social support; social isolation; biomedical technology; military family

XSL•FO
RenderX

Introduction

Background

Social isolation and social support deficits are strongly associated with adverse psychological and health outcomes. Defined as living alone, having few people in one's social network, or having infrequent contact with others [1,2], social isolation captures the objective social environment and has been shown to be at least as predictive of mortality as smoking, obesity, high blood pressure, and high cholesterol [3]. Social support is the resources perceived to be available through formal and informal groups or relationships [4], and lack of social support has been linked to greater risk of mortality, coronary heart disease [5], and depression [6]. Social support includes specific types of support, such as emotional support, that is, the provision of empathy, reassurance, and opportunities for emotional expression; informational support, that is, the provision of advice for dealing with problems; and instrumental support, that is, the provision of tangible forms of assistance, such as lending money or taking care of someone when they are sick [7].

Informal caregivers—those who provide unpaid care to family members, friends, or neighbors with disabling conditions—have an elevated risk of experiencing social isolation and social support deficits [8]. Caregivers face many challenges, including finding time for family and friends [9], and this may impede their ability to avail themselves of social support. Accordingly, interventions have been developed to decrease social isolation and increase social support among caregivers by strengthening existing social connections or creating new ones. Previous research on social support interventions for caregivers is characterized by mixed findings [10], with certain types of interventions demonstrating greater benefit than others. The authors of one systematic review found promising effects of remote interventions on caregivers' social outcomes (eg, satisfaction with support, companionship, and relationship quality with the care recipient) but concluded that more replications are required to have confidence in their benefits [10].

Web-based remote interventions have been recommended for further study, given their potential benefits relative to in-person interventions [10]. Web-based social support interventions may consist of Web-based meetings of group members at a regularly scheduled time, a Web-based network in which members can post comments and share information with each other at any time, *chat* functionality that allows members to send messages to each other in real-time, and webinars on featured topics of interest to group members. The ease of accessing these networks allows members to view information and interact with others at their own convenience without leaving home, thereby eliminating barriers to participation, such as travel time, distance, and the need to identify an alternate caregiver for the care recipient. This could be particularly beneficial to family caregivers. Moreover, given that previous research has documented caregivers' perceptions that caregiving is stigmatized [11] and that care recipients and their families (including caregivers) are devalued by others [12], Web-based interventions may be particularly attractive to caregivers because of the anonymity they afford.

The potential utility of Web-based interventions for caregivers is further suggested by additional recent research. For example, a systematic review of Web-based interventions for older adults found that interventions providing social support, professional support, and instructions in problem solving to caregivers yielded positive outcomes [13]. Nonetheless, the authors cautioned that the specific components of effective Web-based interventions cannot be clearly inferred from this review and called for additional research to illuminate the mechanisms of action. Similarly, little is known about the implementation of Web-based social support interventions, including how caregivers perceive and engage in these interventions, an understanding of which could help clarify the mechanisms of action. In two qualitative studies of the experiences and perceived benefits of Web-based interventions for caregivers of older adults [14] or people with dementia [15], caregivers perceived emotional benefits such as decreased social isolation and loneliness and informational benefits such as learning how to be a better caregiver. Caregivers also commented on their engagement with the networks, asserting that simply reading the material, rather than posting material, was sufficient to benefit from the intervention [15], and perceiving the interactive platform positively because it afforded a *protected environment* for communication with others who were experiencing similar challenges [14].

In addition, although thin, the existing evidence base highlights the unexploited potential of Web-based interventions to improve social support for caregivers. The specific ways in which caregivers engage in and benefit from Web-based social support interventions warrant further exploration to gain insight into how such interventions should be designed to meet the needs of caregivers. In addition, the existing evidence base consists largely of research conducted on caregivers of older adults, primarily those with dementia, with much less known about caregivers of family members with other types of conditions and care needs.

Objectives

In this study, we seek to fill these gaps by examining utilization and perceptions of a Web-based social support intervention for informal caregivers of wounded, ill, and injured United States (US) military service members and veterans. Military caregivers differ from caregivers of civilian care recipients (ie, without a history of military service) in several important ways, one of which is that they often provide care for individuals with mental health conditions such as posttraumatic stress disorder (PTSD) [9]. Military caregivers are also less likely to have a caregiving support network than civilian caregivers [9].

The intervention is built around a Web-based peer support network called the Military and Veteran Caregiver Network (MVCN), which was established in 2015 under the Tragedy Assistance Program for Survivors (TAPS) and is currently administered by the American Red Cross. On the basis of a peer support model [16], the primary goals of the network are to reduce social isolation, increase emotional support, and provide informational support in the form of centralized, high-quality

XSL•FO
RenderX

content and resources tailored to the unique needs of this population. Funded by the Bristol Myers Squibb Foundation, Elizabeth Dole Foundation, and others, the community was created by the staff of TAPS in conjunction with military caregivers who came to work for MVCN. In deciding on the network's features and content, its creators drew from existing research on peer support, the curricula used by TAPS and partner organizations that had developed similar types of social support groups for caregivers, and feedback from partner organizations and their members. The network allows caregivers to post and read comments, exchange information about relevant resources, and attend webchats and webinars about featured topics of interest to caregivers. The network also includes forum groups organized around specific topics, direct messaging functionality, and trained peer and professional moderators whose role is to make posts helpful and positive. Outside of MVCN's website, MVCN users can also access content through monthly question and answer calls, email digests, and MVCN's Facebook page. Content includes both caregiver-specific and noncaregiving topics and is organized by topics. Screenshots of the MVCN website are provided in Multimedia Appendix 1 to illustrate some of the intervention's components.

Methods

Study Design and Setting

We conducted a mixed-methods study of MVCN users, examining quantitative survey data to assess the frequency of their use and perceptions of the network and triangulating it with qualitative data from focus groups and interviews (FGIs) to obtain a richer, more detailed understanding of participants' use and perceptions of the network. The quantitative data were collected as part of a larger quasi-experimental, longitudinal study of military caregivers enrolled in MVCN and a comparison group of military caregivers who had not joined MVCN but were members of other military caregiver groups. Comparison group participants were recruited based on their membership in military caregiver organizations other than MVCN, such as Hidden Heroes, Operation Family Caregiver at the Rosalynn Carter Institute for Caregiving, the Caregiver Action Network, Blue Star Families, and the American Legion Auxiliary. In this study, we focused on the subset of survey participants who had joined MVCN, describing cross-sectional findings on their experiences and perceptions of MVCN from the last follow-up survey administered 6 months after they joined MVCN. Qualitative data were also collected from MVCN users over a similar time frame.

Quantitative Survey Data

Recruitment and Sampling

Eligibility criteria for inclusion in the survey data analysis were being at least 18 years old; providing unpaid care and assistance to a current or former member of the US military, National Guard, or Reserves who has an illness, injury, or condition for which they require outside support; and being an MVCN member. All US military and veteran caregivers are eligible to join MVCN. Those who wish to join must submit documentation to MVCN to verify their eligibility, and all caregivers who are approved receive an email from MVCN confirming their membership. All caregivers who joined MVCN during the study's enrollment period (September 2016 to February 2017) were invited to complete the online study eligibility screener in the confirmation email, and 62.0% (323/521) took the screener survey. Of 323 MVCN members who started the screener, 86.3% (279/323) met study eligibility criteria and were invited to enroll in the study. Informed consent was conducted on the Web before the baseline survey, with participants clicking a box to indicate their consent (or not) in lieu of written consent. Both the baseline and 6-month surveys were completed on the Web. Moreover, 6 months later, those who had completed the baseline survey were sent an email inviting them to complete the 6-month survey. Multiple reminder emails were sent to participants to maximize the likelihood of survey completion. The baseline and 6-month surveys were completed by 243 and 217 MVCN members, respectively, resulting in a retention rate of 89.3% (217/243). We compared survey completers with noncompleters on several characteristics, including their own and their care recipients' demographic characteristics, the types and extent of assistance provided to care recipients, and their care recipients' functional limitations, and we found that only the care recipient's age significantly differentiated survey completers from noncompleters. Participants with older care recipients had lower odds of completing the 6-month survey (odds ratio=0.97, P=.03). Participants were compensated for completing each survey (US $10 for the baseline survey and US $20 for the 6-month survey).

Data Collection and Analysis

The survey assessed MVCN users' participation in the network, including their frequency and duration of visiting the website and types of activities in which they engaged over the past 3 months; perceptions of the network, including potential barriers to using the network; and participation in and perceptions of other resources for caregivers. Several items assessing perceptions of MVCN were adapted from an existing scale designed to assess consumers' experiences on the Web [17]. All other items were created for this study.

Items assessing perceptions of MVCN and other resources for military caregivers were rated on a 5-point Likert scale with response options ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Responses were dichotomized to indicate agreement (*agree* or *strongly agree*) or lack of agreement (*neither agree nor disagree*, *disagree*, or *strongly disagree*) with the statement.

All analyses of quantitative survey data were univariate descriptive statistics of item responses that included data from all respondents who answered the item. Missing data on survey items were uniformly low (5% or less). Because the analyses presented here serve only a descriptive purpose, no tests of significance were conducted.

Qualitative Data From Focus Groups and Interviews

Recruitment and Sampling

To gain greater insight into how users experienced and perceived MVCN, we conducted FGIs with MVCN users. We had originally planned to conduct only focus groups but resorted to

XSL•FO
RenderX

individual interviews after having several *no-shows* for focus groups. Participants were recruited with assistance from an MVCN employee who advertised the study to all interest groups on MVCN's secure website, on MVCN's Facebook page, and in four weekly email digests sent to MVCN users. The advertisement included an invitation to complete a brief Web-based demographic questionnaire and eligibility screener, which was completed by 119 MVCN users. Similar to the eligibility criteria for survey participation, participation in the FGIs was limited to unpaid or informal caregivers who were members of MVCN, and the same eligibility screening questions used for the survey were used to identify unpaid or informal caregivers who were members of MVCN for the FGIs. Although survey participation was limited to caregivers who had joined MVCN very recently at the time of study enrollment (which is why the invitation to participate was sent right after caregivers registered for MVCN), participation in FGIs was open to all unpaid caregivers who were MVCN members, regardless of how much time had passed since they joined MVCN.

We planned to collect data from only 15 to 20 participants, which we estimated would be sufficient to achieve saturation. Therefore, we scheduled users to participate based on their availability until a final sample comprising 11 focus group participants (4 groups total, with 2 to 4 participants per group) and 4 interview participants was obtained and saturation was achieved.

Data Collection and Analysis

Qualitative data were collected using a semistructured protocol created for this study. In general, protocol topics aligned with survey topics, covering MVCN members' participation in and perceptions of MVCN and other resources for caregivers. One of the 3 researchers conducted FGIs between July 2016 and September 2017. FGIs were conducted over the phone to allow MVCN users in any part of the United States to participate. The informed consent process was conducted verbally before beginning the FGIs. Participants were compensated with a US $25 Amazon gift card. All the FGIs were audio-recorded and transcribed verbatim. All study procedures were approved by the Human Subjects Protection Committee at the institution where the research was conducted.

An inductive content analysis was conducted by 4 researchers to organize and condense the qualitative data and identify key themes and insights. All transcripts were double-coded (ie, coded independently by 2 members of the research team), and a codebook was created collaboratively. The codebook included predetermined codes based on the topics and subtopics covered in the semistructured protocol and more specific emergent codes derived from participants' comments. The codebook was created primarily from the first pass at coding the transcripts, in which the first coder marked each text fragment in the transcript with an appropriate topic, subtopic, and emergent code. For the second pass, another coder marked text fragments with one of the existing codes in the codebook (while blinded to the code assigned to the text fragment by the first coder) or, if no appropriate codes existed, generated a new code to capture the text fragment and added it to the codebook. After all the transcripts had been double-coded, the coders met to discuss and resolve discrepancies in the coding. In the few cases where the 2 coders of the same transcript could not reach agreement on the most appropriate code for a given text fragment, a third team member decided on the appropriate code.

Results

Participant Characteristics

MVCN members who completed the 6-month survey were mostly female (200/217, 92.2%); non-Hispanic white (164/217, 75.6%); under the age of 40 (121/217, 55.7%); married, living with their partner, or had a noncohabiting significant other (202/217, 93.1%); had at least 1 child under the age of 18 (130/217, 59.9%); had no more than an associate's degree or lower level of education (126/217, 58.1%); and had been a caregiver for at least 5 years (146/217, 67.9%). Most survey participants were married to or otherwise partnered with their care recipients (186/217, 85.7%) and resided with their care recipient (197/217, 90.8%). Their care recipients were mostly male service members or veterans (193/217, 88.9%) who were under the age of 40 (117/217, 53.9%) and had served in the military after September 11, 2001 (181/217, 85.3%). Nearly all care recipients had been diagnosed with at least one physical condition (eg, back pain, diabetes, paralysis, or spinal cord injury; 208/217, 95.9%), and a great majority of them had been diagnosed with a psychological condition (ie, PTSD, major depressive disorder, or a substance use disorder; 184/217, 84.8%). Nearly two-thirds of care recipients had been diagnosed with a neurological condition (ie, traumatic brain injury, Parkinson disease, or dementia; 141/217, 65.0%), and an equal number of care recipients had been diagnosed with at least 6 medical conditions (141/217, 65.0%).

All 15 FGI participants were women providing care for a military service member or veteran who had served after September 2001, and 5 participants reported that their care recipients had also served before September 2001. All but 3 FGI participants were married to their care recipients; of the 3 nonspouse caregivers, 2 were siblings and 1 was the mother of the care recipient. Most FGI participants (9/15, 60%) were under the age of 40.

Participation in the Network

Quantitative Findings

Frequency and Duration of Use

Overall, MVCN members (N=211) reported the infrequent use of the network over the past 3 months. Approximately one-third (67/211, 31.8%) of the members reported that they had not visited the website, nearly one-third (61/211, 28.9%) had visited once a month or less, 15.2% (32/211) had visited 2 or 3 times a month, and 24.2% (51/211) had visited once a week or more.

Of the MVCN members who had visited the website in the past 3 months (N=144), which we refer to throughout the description of survey results as *users*, 81.3% (117/144) reported that a typical visit was 30 min or less. Specifically, 25.0% (36/144) of users spent less than 10 min during a typical visit, 32.6% (47/144) spent between 10 and 20 min, 23.6% (34/144) spent

XSL•FO
RenderX

between 20 and 30 min, and 18.8% (27/144) spent more than 30 min.

Engagement in Specific Activities

When asked about engagement in various network-related activities over the past 3 months, users most commonly reported accessing information and resources from the website's resource library (91/144, 63.2%), followed by joining an interest group (eg, caregivers of care recipients with PTSD; 60/144, 41.7%), attending a webinar (42/144, 29.2%) or webchat (36/144, 25.0%), and posting comments, questions, or links to the network (32/144, 22.2%). Nearly one-third (45/144, 31.2%) of users reported interacting with other MVCN members outside of MVCN.

Qualitative Findings

Use of Informational Resources

When asked why and how they used MVCN, most FGI participants reported that they visited the website to find information, such as guidance on how to navigate caregiver resources and cope with the challenges of caregiving:

> *Having no clue what a caregiver is, you know, what is expected of me, resources I could reach out to, other caregivers in my situation and hearing from them about their challenges and how they've overcome them.*

Some reported using the MVCN website for a specific purpose. As one caregiver stated:

> *Usually if I'm going on there, it's a very specific item I'm looking for.*

Passive Engagement

Many participants described a pattern of use marked by passive and limited engagement, mostly reading others' posts rather than posting themselves. FGI participants perceived that they had benefited from simply reading posts that allowed them to obtain high-quality information and helped them to feel less alone. For example, as one participant commented:

> *I just really enjoy reading posts, realizing I'm not alone. I've gotten some really, really good information from other people...I've gained a lot just from reading information that others have shared.*

Another participant noted that:

> *The articles or the comments...enable me to either get through the day or get to the information that I need.*

Other Modes of Participation

FGI participants also reported engaging through other modes outside of the MVCN website. Most participants reported that they receive weekly email digests from MVCN and often review these in lieu of visiting the website. For some participants, the email digests served as a prompt to visit the website to explore available resources. In addition, many received MVCN updates through Facebook, which also functioned as an alternative to visiting the MVCN website. Many FGI participants had also attended MVCN's peer support calls or webinars, with some participants noting that they benefited from them.

Perceptions of the Network

Quantitative Findings

Perceived Benefits and Satisfaction

MVCN users were also asked to indicate their agreement with several statements about the potential benefits of using the website (Table 1). A slight majority of users reported that the website had improved their decision making, agreeing that this website "helps me make good caregiving decisions" (86/144, 59.7%) and "provides information that helps me make important decisions" (78/144, 54.2%). Slightly less than half of the users perceived that "this site helps me better manage my time and resources" (65/144, 45.1%). A little over half of the users endorsed the inspirational and self-improvement benefits of the network, agreeing that the website "makes me think of things in new, more positive ways" (84/144, 58.3%), "makes a difference in my life" (82/144, 56.9%), taught the user "how to improve myself" (76/144, 52.8%), and "inspires me in my own life" (73/144, 50.7%). Similarly, about half of the users agreed that "I am a better person for using this site" (71/144, 49.3%). Approximately half of the users endorsed positive perceptions of the community, agreeing that "I have learned a lot from the posts of other caregivers who visit this site" (79/144, 54.9%) and "this site does a good job of getting its visitors to contribute or provide feedback" (69/144, 47.9%).

When asked about their overall satisfaction with MVCN, the great majority of users endorsed at least moderate satisfaction. In addition, 45.1% (65/144) of users were *very or extremely satisfied*, 38.9% (56/144) were *moderately satisfied*, and 16.0% (23/144) were *not at all or slightly satisfied.*

Perceived Reasons for Limited Engagement

MVCN members were asked about several reasons why people may not visit or use resources on the website (Table 2). The most commonly endorsed barriers pertained to problems with usability or limited activity on the website. Specifically, 32.9% (69/210) of the members agreed that difficulty finding what one needs was a potential barrier, and 23.8% (50/210) of the members agreed that the website was difficult to use or did not have enough activity.

Members endorsed problems with the utility or accuracy of information and other users to a slightly lesser extent than limitations of the website. Approximately 20% of members endorsed concerns about the utility (43/210, 20.5%) or accuracy (36/210, 17.1%) of information given by others. Similarly, approximately 20% of members endorsed issues with other users, such as not having a lot in common with other users (45/210, 21.4%), gossiping about others (38/210, 18.1%), sniping or attacking of people who post on the website (37/210, 17.6%), and perceiving that other users are not welcoming or friendly (32/210, 15.2%).

XSL•FO
RenderX

Table 1. Users' perceptions of and satisfaction with Military and Veteran Caregiver Network (MVCN) over the past 3 months (N=144). Users are those who reported having visited the network at least once in the past 3 months.

Perceptions[a]	Participants who *agreed* or *strongly agreed*, n (%)[b]
Decision making or resources	
This site helps me make good caregiving decisions	86 (59.7)
This site provides information that helps me make important decisions	78 (54.2)
This site helps me better manage my time and resources	65 (45.1)
Inspiration or self-improvement	
This site makes me think of things in new, more positive ways	84 (58.3)
Using this site makes a difference in my life	82 (56.9)
I have learned how to improve myself from this site	76 (52.8)
This site inspires me in my own life	73 (50.7)
I am a better person for using this site	71 (49.3)
Community	
I have learned a lot from the posts of other caregivers who visit this site	79 (54.9)
This site does a good job of getting its visitors to contribute or provide feedback	69 (47.9)
Overall satisfaction with MVCN[c]	
Not at all or slightly satisfied	23 (16.0)
Moderately satisfied	56 (38.9)
Very or extremely satisfied	65 (45.1)

[a]Except for the item assessing overall satisfaction with MVCN, all items were adapted from an existing scale designed to assess consumers' experiences on the Web [17].

[b]The SE for all percentages was 0.04, except for the percentage of MVCN users who were *not at all or slightly satisfied* with MVCN, for which the SE was 0.03.

[c]Satisfaction with MVCN was rated on a scale that ranged from 1 (*not at all satisfied*) to 5 (*extremely satisfied*) and collapsed to form 3 categories: *not at all or slightly satisfied* (1 or 2), *moderately satisfied* (3), or *very or extremely satisfied* (4 or 5).

Table 2. Possible reasons why people may not visit Military and Veteran Caregiver Network (MVCN) or use resources on its website (N=210).

Possible reasons[a]	Participants who *agreed* or *strongly agreed*, n (%)[b]
Limitations of website	
It is difficult to find what you need on the website	69 (32.9)
There is not enough activity on the site (e.g., too few posts, not enough responses to posts or active discussion)	50 (23.8)
The website is difficult to use (slow to load, unorganized)	49 (23.3)
Utility or accuracy of information	
The information given by other users is not useful	43 (20.5)
The information given by other users is not accurate	36 (17.1)
Problems with other users	
I don't have a lot in common with other users	45 (21.4)
There is a lot of gossip posted by other users	38 (18.1)
There is a lot of sniping/attacking of people who post to the website	37 (17.6)
The other users are not welcoming/friendly	32 (15.2)

[a]This series of items was rated by all MVCN members, regardless of whether they had visited the MVCN website in the last 3 months.

[b]The SE for all percentages was 0.03, except for the percentage for the item *The other users are not welcoming/friendly*, for which the SE was 0.02.

XSL•FO
RenderX

Qualitative Findings

Perceived Benefits

In general, FGI participants reported positive perceptions of MVCN and noted few, if any, undesirable features. Some participants commented that, from the outset, MVCN was perceived to be a trustworthy, reliable resource because it has been vetted by other military caregiver support organizations. In addition, FGI participants perceived that, overall, the information and resources shared on MVCN were high-quality, commenting that the content seemed objective and informed by research. One participant highlighted the comprehensiveness of the information provided and expressed confidence in being able to find the necessary resources and information when needed:

> *It seems like if there's anything that I need, anything that I need to know, I could reach out and someone would get the information for me or guide me to the place to get it. So, I do feel like it's very comprehensive in that way.*

When asked about the supportiveness of the MVCN community, FGI participants reported that MVCN had a positive, *drama-free* environment, in contrast to some other social support groups, and attributed this environment to MVCN's professionally trained peer mentors:

> *Whereas some of the groups I've tried in the past...they don't have a social worker facilitator or something like that...I think that [at] MVCN the people are like peer mentors. Some of them have some type of positivity training to help keep that more positive, because otherwise it can just spiral and everyone's just kind of like, "My life is worse than yours," which is not helpful.*

Although some FGI participants considered the peer support from MVCN members helpful, some noted that support from other members was inconsistent or limited, again emphasizing that the benefits derived from MVCN were primarily of an informational, rather than emotional, nature. For example, 1 participant commented that she received feedback "only during the workshop." Another participant elaborated further:

> *One [caregiver platform] might be for...venting, you know, sharing stories kind of environment. And then MVCN for me is more like a resource center that is managed by professionals...I need to tap into some-thing that I know I can't get somewhere else or no one else knows, and I don't want the chatter around it; I know that's the place I'm going to go, and I can count on whatever is going to come out of it is going to be probably what I need.*

When asked about other benefits of MVCN, many FGI participants highlighted its privacy, noting that limiting access to verified caregivers contributes to a safe environment. As 1 participant said:

> *I appreciate the fact that MVCN is private and it's held in an online environment that isn't Facebook. And I feel like I have a little bit more control over how far my words go and sort of that it's a safe place.*

Some participants also explained that the benefits derived from MVCN are in direct proportion to the user's level of engagement, asserting that "you get out of it what you put into it."

Perceived Reasons for Limited Engagement and Suggested Solutions

FGI participants provided possible explanations for the limited engagement of some community members. A small number of FGI participants who reported limited engagement with MVCN attributed this to other members' lack of active participation in MVCN. For example, 1 participant noted that she is not motivated to visit MVCN because other users do not participate or respond to posts. Some FGI participants noted that, because of other commitments and responsibilities, caregivers "have very little extra time" to participate in activities such as calls and webinars. In addition, some FGI participants attributed their low participation in MVCN to the impersonal nature of Web-based groups, citing a preference for in-person support:

> *I'm more of a physical...group type of person, and that's scary. But at the same time, that gets me out of isolation too. So, having that physical contact, which I know is not easy, and depending where we live...it's a challenge. But having those physical groups in different areas, I know for me, would be helpful.*

Participants suggested that engagement with MVCN could potentially be increased by making MVCN more accessible and user-friendly. Noting that they were less motivated to log-in to MVCN if they had to go through a separate website, participants recommended creating a mobile app for MVCN. Similarly, they suggested making MVCN more user-friendly by allowing users to see their view history and what has changed since their last log-in. Participants also expressed the desire for a more *personal touch* both within and beyond MVCN. Within MVCN, this included recommending tailored resources for users and actively encouraging participation and interaction among users. As 1 participant suggested:

> *So, kind of like one step more to help the community interact, and then maybe every so often looking at what you know about the different members and saying, "Hey, I think that this is useful for you," or, "Have you connected with this person? I think that you would have a similarity and maybe be peer mentors and things like that." Or just tagging you on topics and go like, "Oh, this might help you in what you're looking for. Or do you have any feedback regarding this?"*

Beyond the Web-based community, participants suggested facilitating in-person meetings for caregivers who live in the same geographic area.

XSL•FO
RenderX

Participation in and Perceptions of Other Peer Support Communities for Caregivers

Quantitative Findings

Participation in other caregiver support groups was common (Table 3). The great majority of MVCN members were in at least one other Web-based-only caregiver support group (171/211, 81.0%), and more than half participated in at least two groups (116/211, 55.0%). Nearly two-thirds of members were in a military or veteran caregiver group on Facebook (130/211, 61.6%). Relatedly, a slight majority of the members reported monitoring several different military caregiver websites to get the information, resources, and support they need (119/211, 56.4%; SE=0.03; data not shown in table). Although less common than participation in Web-based-only groups, participation in in-person caregiver support groups was nonetheless fairly common, with half of MVCN members belonging to at least one in-person group (106/211, 50.2%). When asked how often they participated in other caregiver support groups, both on the Web and in-person, nearly one-third of the members said once a week or more (64/211, 30.3%), one-third said one to three times a month (69/211, 32.7%), and over one-third said every few months or less (78/211, 37.0%).

The survey also assessed MVCN members' perceptions of resources for military caregivers in general. Nearly three-quarters (154/211, 73.0%) of members agreed that more in-person peer support groups for military caregivers are needed, whereas only 45.0% (95/211) of members agreed that more Web-based peer support groups for military caregivers were needed. Only 21.8% (46/211) of members believed there were too many Web-based groups for military caregivers. When asked about perceived needs related to informational resources, over two-thirds (147/211, 69.7%) of members reported a need for Web-based resources that provide specific types of help for military caregivers (eg, help with alcohol abuse or depression), and over two-thirds (144/211, 68.2%) of members indicated a need for resources providing information and support for military caregivers to be located in one central place on the Web.

Qualitative Findings

All FGI participants reported involvement with other caregiver support groups, most of which were military-specific. One participant observed that, relative to other Web-based caregiver support groups, MVCN "seems to be more positive and objective," focusing on "different research coming out, different types of therapies and modalities." Other participants echoed these viewpoints.

Table 3. Military and Veteran Caregiver Network (MVCN) members' participation in caregiver support groups other than MVCN (N=211).

Participation in other caregiver support groups	Participants, n (%)
Member of a military or veteran caregiver group on Facebook	
Yes	130 (61.6)[a]
No	81 (38.4)[a]
Number of Web-based-only caregiver support groups other than MVCN	
0	30 (14.2)[b]
1	55 (26.1)[a]
2	51 (28.9)[a]
3 or more	65 (30.8)[a]
Number of in-person caregiver support groups	
0	105 (49.8)[a]
1	57 (27.0)[a]
2	39 (18.5)[a]
3 or more	10 (4.7)[c]
Frequency of participation in Web-based and in-person caregiver support groups other than MVCN	
Once a week or more	64 (30.3)[a]
One to three times a month	69 (32.7)[a]
Every few months or less	78 (37.0)[a]

[a]SE=0.03.

[b]SE=0.02.

[c]SE=0.01.

XSL•FO
RenderX

Participants highlighted the variability across Web-based support groups in the types and amount of informational support available and the supportiveness of the community environment. As 1 participant explained, each group fills a different niche:

> *I think that each organization has their niche...or I determine that they just have some sort of a niche. And so, I use each organization and their information for that purpose. For instance, [one organization] runs retreats. And they have other slots(?), but that's kind of what I've done with them so far. And some of the private, closed, Facebook groups, where the members are vetted, if I need to vent about something, which I pretty much don't do, I would do it there...*

Some FGI participants reported that other Web-based communities often have a lot of gossip or an otherwise negative environment:

> *The drawback, you always end up with one or two that are negative Nellys, the negative ones that don't have anything positive going on in their life...You're going to have your ones that just won't ever be happy. Because they're not happy, they don't want anybody else to be happy.*

Participants perceived an overall deficit in in-person support. Many participants reported little to no local community support for caregivers, requiring them to drive long distances to participate in in-person groups, create informal local support groups, or utilize resources over the phone or on the Web. As 1 participant commented:

> *We're heavily disjointed, and California is a really, really large state. So there's a lot of phone interaction, or like we're doing now, Skype. And a lot of online...my intention is to be able to pull together...just a group of individuals, but I think there's something to be said about people who can meet together and just-sometimes it's just holding someone's hand through a rough day, sit there through a rough story.*

Discussion

Principal Findings

Collectively, the findings from this study suggest that, although many caregivers in our study reported infrequent and passive engagement with the Web-based network examined, it is generally viewed as a source of high-quality informational support. In addition, caregivers perceived that in-person contact is necessary to meet their emotional support needs and noted that local in-person support groups are rare. Below we discuss these findings and their implications in greater detail.

Overall, MVCN members reported infrequent use of the network, with approximately one-third of survey participants not having visited it in the past 3 months. Similar levels of use were reported in another study of a Web-based community for a smoking cessation program in which approximately one-third of participants never visited the community [18]. Moreover, only 24.2% (51/211) of MVCN members reported using the website once a week or more. In contrast, 43% of caregivers in the comparison group of our study reported having used their most frequently visited website once a week or more (TET, PhD, unpublished data, January 2018).

Although FGI participants acknowledged the importance of engaging actively with the network to benefit maximally, a minority (32/144, 22.2%) of survey participants reported its active use, such as posting links or comments to the network. Furthermore, FGI participants reported primarily passive use of the network, such as reading email digests or going to the website to find information about topics or resources of interest. Consistent with previous research [15], FGI participants felt that they benefited sufficiently from passive use of the website.

There are several possible explanations for caregivers' infrequent use of and typically passive engagement with the network. Some FGI participants reported that they did not have time to use the website as often as they would like. In addition, most survey participants belonged to multiple Web-based support groups for caregivers and many participated in in-person support groups. Thus, the small amount of discretionary time available to caregivers may be split across several different groups.

Quantitative and qualitative data indicated that caregivers used and viewed the network primarily as a source of informational support. Among survey participants, accessing resources and information from the network's library was the most commonly endorsed type of MVCN-related activity (91/144, 63.2%). Similarly, many FGI participants reported using the website in a much-targeted way to obtain needed information and resources. This may help to explain why many members used the website infrequently. If viewed mainly as a resource directory, rather than as a social network, the network would be visited only on an *as-needed* basis. Thus, infrequent use may primarily reflect how caregivers use the website, rather than a lack of interest in or appreciation of the website. Indeed, the informational support provided by the network was positively regarded by most caregivers, with 54.2% (78/144) of survey participants agreeing that the website provides information that helps them make important decisions. Moreover, FGI participants commended the comprehensiveness, reliability, and quality of the information provided by the network, noting the professional curation of content and expressing confidence that they could find the needed resources.

Although caregivers generally perceived more informational than emotional benefits of the network, they nonetheless valued the positive environment and privacy of the network. In FGIs, multiple caregivers noted that MVCN differed from some other websites in which community members would begin complaining and set off a downward spiral. Among survey participants, 58.3% (84/144) agreed that the website makes them "think of things in new, more positive ways." Conversely, a minority of survey participants endorsed problems with other users pertaining to gossip, sniping, or being unfriendly. FGI participants attributed the network's positive environment to its professionally trained peer mentors. In addition, FGI participants appreciated the vetting of members and privacy afforded by the network, a finding that dovetailed with findings from a previous study of a Web-based social support intervention for caregivers [14]. Thus, it is important to

XSL•FO
RenderX

incorporate these characteristics for other Web-based caregiver support communities to gain the trust of prospective members and encourage them to participate in the community.

Although caregivers appreciated MVCN's positive environment, they also found it inadequate at meeting their need for emotional support and expressed a desire for a more *personal touch*, particularly in the form of in-person peer support. When asked about the perceived need for in-person and Web-based peer support in the survey, nearly three-quarters (154/211, 73.0%) of MVCN members agreed that more in-person peer support groups for military caregivers are needed, whereas only 45.0% (95/211) of MVCN members agreed that more Web-based peer support groups for military caregivers are needed. FGI participants explained that in-person contact fulfills a need for emotional support that cannot be met by the Web-based contact alone. Similar sentiments about Web-based versus in-person support have been obtained in previous qualitative research conducted with members of a Web-based support community for Parkinson disease [19]. Moreover, experimental research suggests that, under stressful conditions, emotional support provided in-person more strongly bolsters positive affect than the support provided via text messaging [20]. These findings are not surprising in light of evidence that certain types of nonverbal cues (eg, physical warmth from touch and facial expressions such as smiling genuinely) that are present only in in-person interactions may promote the formation of trust in relationships [21,22]. Furthermore, in-person support groups might enable and encourage caregivers to provide help to one another in more tangible ways, thus increasing their levels of instrumental support. FGI participants also indicated very limited availability of local in-person support groups. Thus, making local in-person peer support groups available in geographic areas in which there is a critical mass of caregivers who are interested in attending in-person support groups may help improve emotional support.

Although FGI participants identified deficits in the network's ability to meet their emotional support needs, they did not report perceiving any adverse effects of the use of the network on their mood or loneliness. This is interesting in light of recent research indicating that the use of social media websites (eg, Facebook and Twitter) increases the risk of experiencing loneliness [23] and depressive symptoms [24]. Of course, the Web-based support community for caregivers likely differs from a typical social media website in many ways that might make use of the support community more likely to benefit their well-being and reduce social isolation. Scholars have recently argued that the benefits of social network groups depend on the network's ability to facilitate social connections between members [25,26]. The caregiver community studied here comprises individuals who are united by shared problems, and most of the website's content specifically aims to address these common problems and is professionally curated and moderated for that purpose. In contrast, a social media website such as Facebook lacks a well-defined purpose (eg, people can post a much broader range of content) and professional curation and moderation of content and comments made by other network members. In the absence of the constraints that characterize MVCN, a social media website such as Facebook might simply perpetuate users' existing insecurities and thus fuel feelings of loneliness and depressive symptoms rather than feelings of connection with fellow users [26].

Caregivers also identified areas of improvement for the network. Although survey findings suggested that nearly half of MVCN users perceived that the website effectively encouraged members to contribute to or provide feedback on the website, nearly one-quarter of MVCN members (including those who had not visited the website in the past 3 months) reported that the lack of activity on the website (eg, not enough posts or responses) might prevent some caregivers from visiting it. Moreover, some FGI participants also noted that the lack of active engagement among other members limited their own engagement. A lack of responsiveness has also been identified as a problem in other research conducted on Web-based support communities for individuals with Parkinson disease, indicating that it is not unique to MVCN [19]. In a related vein, some FGI participants suggested that MVCN could do more to encourage more active engagement among members by, for example, having peer moderators invite users to comment and nudge them to interact with other members who are similar to them in various ways.

Similar strategies have been tested in other social media websites and found to be effective at increasing user engagement, lending credence to this suggestion. Specifically, one study on Facebook newcomers found that those who were initially disinclined to contribute actively and who were subsequently *singled out* by other users (ie, tagged in photos posted by other users) exhibited greater long-term sharing [27]. Similarly, an experiment conducted at another social media website used recommender systems at sign-up to make tailored recommendations to new users regarding relevant content and other users with whom to connect, and found that the new users who received these tailored recommendations significantly increased their viewing of and contributions to the website [28]. Moreover, the same study found that recommending more active users as connections for new users was associated with greater engagement of new users [28].

Strengths and Limitations

This study's primary strengths include the use of both quantitative and qualitative data to explore an important but relatively understudied topic, that of caregivers' use and perceptions of a Web-based peer support network; the extensive coverage of members of the network with the sample of survey participants; and the in-depth examination of why and how caregivers use Web-based peer support. This study also had some limitations. One limitation is the exclusive reliance on participants' self-report, which may have been biased by social desirability concerns or affected by difficulty recalling the requested information (eg, the frequency of visits to the MVCN website). Similarly, participants' beliefs about how they benefited from MVCN and suggested changes that would reap additional benefits (eg, providing in-person peer support groups) may not be accurate; that is, participants may not know whether or how they have actually benefited from MVCN or how they would respond to future changes made to MVCN for their benefit. In addition, the generalizability of the findings of this

XSL•FO
RenderX

study to a broader caregiving population (eg, dementia caregivers) and to other peer support networks is unclear.

Conclusions

The Web-based peer support network examined in this study was valued by its members for the provision of trustworthy, readily accessible information on a wide variety of topics and maintenance of a private, positive environment by professionally trained peer mentors. In general, members engaged with the network infrequently and passively, which they attributed to other members' limited engagement and their own limited time to visit the network. Members expressed a desire for the network to provide a more *personal touch* by actively encouraging interactions among users and facilitating local in-person peer support groups for caregivers in areas with critical masses of caregivers who are interested in and able to attend such groups.

Acknowledgments

The authors wish to acknowledge the staff at the MVCN for their support and assistance in the recruitment of study participants and the caregivers who participated in the study. This study was supported by a research grant from the Bristol-Myers Squibb Foundation (BMSF; BMSF_10.01.15_) awarded to Thomas E Trail and Terri Tanielian.

Authors' Contributions

CV took primary responsibility for designing this study, analyzing the quantitative and qualitative data, and writing the paper. TET took primary responsibility for designing the parent study for which the data presented here were collected and participated in the design and development of this study, analysis of the qualitative data, and writing of the paper. AM and SD participated in the design and development of this study, analysis of the qualitative data, and writing of the paper. TT and EF participated in the design and development of this study and writing of the paper. Neither MVCN nor BMSF played a role in the design of this study; the collection, analysis, and interpretation of data; the writing of this paper; or the decision to submit the paper for publication. The views expressed are those of the authors and not necessarily those of BMSF or MVCN.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshots of the Military and Veteran Caregiver Network website.

[PDF File (Adobe PDF File), 837KB - jmir_v20i8e257_app1.pdf]

References

1. Cacioppo JT, Cacioppo S. Social relationships and health: the toxic effects of perceived social isolation. Soc Personal Psychol Compass 2014 Feb 1;8(2):58-72 [FREE Full text] [doi: 10.1111/spc3.12087] [Medline: 24839458]
2. Holt-Lunstad J, Smith TB, Baker M, Harris T, Stephenson D. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. Perspect Psychol Sci 2015 Mar;10(2):227-237. [doi: 10.1177/1745691614568352] [Medline: 25910392]
3. Pantell M, Rehkopf D, Jutte D, Syme SL, Balmes J, Adler N. Social isolation: a predictor of mortality comparable to traditional clinical risk factors. Am J Public Health 2013 Nov;103(11):2056-2062. [doi: 10.2105/AJPH.2013.301261] [Medline: 24028260]
4. Cohen S, Underwood LG, Gottlieb BH. Social Support Measurement and Intervention: A Guide for Health and Social Scientists. Oxford, New York: Oxford University Press; 2000.
5. Barth J, Schneider S, von Känel R. Lack of social support in the etiology and the prognosis of coronary heart disease: a systematic review and meta-analysis. Psychosom Med 2010 Apr;72(3):229-238. [doi: 10.1097/PSY.0b013e3181d01611] [Medline: 20223926]
6. Lakey B, Cronin A. Low social support and major depression: research, theory and methodological issues. In: Risk factors for depression. San Diego, CA: Academic Press; 2008:385-408.
7. Barrera M. Distinctions between social support concepts, measures, and models. Am J Community Psychol 1986;14(4):413-445. [doi: 10.1007/BF00922627]
8. Brodaty H, Donkin M. Family caregivers of people with dementia. Dialogues Clin Neurosci 2009;11(2):217-228 [FREE Full text] [Medline: 19585957]
9. Ramchand R, Tanielian T, Fisher MP, Vaughan CA, Trail TE, Batka C, et al. Rand. 2014. Hidden Heroes: America's Military Caregivers URL: https://www.rand.org/content/dam/rand/pubs/research_reports/RR400/RR499/RAND_RR499.pdf [accessed 2018-08-13] [WebCite Cache ID 71dZu1ymq]

XSL•FO
RenderX

10. Dam AE, de Vugt ME, Klinkenberg IP, Verhey FR, van Boxtel MP. A systematic review of social support interventions for caregivers of people with dementia: are they doing what they promise? Maturitas 2016 Mar;85:117-130. [doi: 10.1016/j.maturitas.2015.12.008] [Medline: 26857890]
11. Phelan SM, Griffin JM, Hellerstedt WL, Sayer NA, Jensen AC, Burgess DJ, et al. Perceived stigma, strain, and mental health among caregivers of veterans with traumatic brain injury. Disabil Health J 2011 Jul;4(3):177-184. [doi: 10.1016/j.dhjo.2011.03.003] [Medline: 21723524]
12. Struening EL, Perlick DA, Link BG, Hellman F, Herman D, Sirey JA. Stigma as a barrier to recovery: the extent to which caregivers believe most people devalue consumers and their families. Psychiatr Serv 2001 Dec;52(12):1633-1638. [doi: 10.1176/appi.ps.52.12.1633] [Medline: 11726755]
13. Guay C, Auger C, Demers L, Mortenson WB, Miller WC, Gélinas-Bronsard D, et al. Components and outcomes of internet-based interventions for caregivers of older adults: systematic review. J Med Internet Res 2017 Sep 19;19(9):e313 [FREE Full text] [doi: 10.2196/jmir.7896] [Medline: 28928109]
14. Barbabella F, Poli A, Andréasson F, Salzmann B, Papa R, Hanson E, et al. A web-based psychosocial intervention for family caregivers of older people: results from a mixed-methods study in three European countries. JMIR Res Protoc 2016 Oct 6;5(4):e196 [FREE Full text] [doi: 10.2196/resprot.5847] [Medline: 27713113]
15. McKechnie V, Barker C, Stott J. The effectiveness of an internet support forum for carers of people with dementia: a pre-post cohort study. J Med Internet Res 2014 Feb;16(2):e68 [FREE Full text] [doi: 10.2196/jmir.3166] [Medline: 24583789]
16. Mentalhealthamerica. 2011. Best Practices Identified for Peer Support Programs URL: https://www.mentalhealthamerica.net/sites/default/files/Best_Practices_Identified_for_Peer_Support_Programs_Jan_2011.pdf [accessed 2018-08-13] [WebCite Cache ID 71da98gvc]
17. Calder BJ, Malthouse EC, Schaedel U. An experimental study of the relationship between online engagement and advertising effectiveness. J Interact Mark 2009 Nov;23(4):321-331. [doi: 10.1016/j.intmar.2009.07.002]
18. Papandonatos GD, Erar B, Stanton CA, Graham AL. Online community use predicts abstinence in combined internet/phone intervention for smoking cessation. J Consult Clin Psychol 2016 Jul;84(7):633-644 [FREE Full text] [doi: 10.1037/ccp0000099] [Medline: 27100127]
19. Attard A, Coulson NS. A thematic analysis of patient communication in Parkinson's disease online support group discussion forums. Comput Human Behav 2012 Mar;28(2):500-506. [doi: 10.1016/j.chb.2011.10.022]
20. Holtzman S, DeClerck D, Turcotte K, Lisi D, Woodworth M. Emotional support during times of stress: can text messaging compete with in-person interactions? Comput Human Behav 2017;71:130-139. [doi: 10.1016/j.chb.2017.01.043]
21. Williams LE, Bargh JA. Experiencing physical warmth promotes interpersonal warmth. Science 2008 Oct 24;322(5901):606-607 [FREE Full text] [doi: 10.1126/science.1162548] [Medline: 18948544]
22. Krumhuber E, Manstead AS, Cosker D, Marshall D, Rosin PL, Kappas A. Facial dynamics as indicators of trustworthiness and cooperative behavior. Emotion 2007 Nov;7(4):730-735. [doi: 10.1037/1528-3542.7.4.730] [Medline: 18039040]
23. Primack BA, Shensa A, Sidani JE, Whaite EO, Lin LY, Rosen D, et al. Social media use and perceived social isolation among young adults in the U.S. Am J Prev Med 2017 Jul;53(1):1-8. [doi: 10.1016/j.amepre.2017.01.010] [Medline: 28279545]
24. Steers MN, Wickham RE, Acitelli LK. Seeing everyone else's highlight reels: how Facebook usage is linked to depressive symptoms. J Soc Clin Psychol 2014;33(8):701-731. [doi: 10.1521/jscp.2014.33.8.701]
25. Clark JL, Algoe SB, Green MC. Social network sites and well-being: the role of social connection. Curr Dir Psychol Sci 2017 Dec 19;27(1):32-37. [doi: 10.1177/0963721417730833]
26. Nowland R, Necka EA, Cacioppo JT. Loneliness and social internet use: pathways to reconnection in a digital world? Perspect Psychol Sci 2018 Jan;13(1):70-87. [doi: 10.1177/1745691617713052] [Medline: 28937910]
27. Burke M, Marlow C, Lento T. Feed me: motivating newcomer contribution in social network sites. 2009 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 4-9, 2009; Boston, MA. [doi: 10.1145/1518701.1518847]
28. Freyne J, Jacovi M, Guy I, Geyer W. Increasing engagement through early recommender intervention. 2009 Presented at: Proceedings of the third ACM conference on Recommender systems; October 23-25, 2009; New York, NY. [doi: 10.1145/1639714.1639730]

Abbreviations

FGI: focus groups and interviews
MVCN: Military and Veteran Caregiver Network
PTSD: posttraumatic stress disorder
TAPS: Tragedy Assistance Program for Survivors
US: United States

XSL•FO
RenderX

Edited by G Eysenbach; submitted 26.01.18; peer-reviewed by G Baumblatt, L Boots, J Griffin, M Reblin; comments to author 30.03.18; revised version received 25.05.18; accepted 29.05.18; published 28.08.18.

Please cite as:
Vaughan C, Trail TE, Mahmud A, Dellva S, Tanielian T, Friedman E
Informal Caregivers' Experiences and Perceptions of a Web-Based Peer Support Network: Mixed-Methods Study
J Med Internet Res 2018;20(8):e257
URL: http://www.jmir.org/2018/8/e257/
doi:10.2196/jmir.9895
PMID:30154074

©Christine Vaughan, Thomas E Trail, Ammarah Mahmud, Stephanie Dellva, Terri Tanielian, Esther Friedman. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 28.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Relationship Between Internet Health Information and Patient Compliance Based on Trust: Empirical Study

Xinyi Lu[1*], B Eng; Runtong Zhang[1*], PhD; Wen Wu[1*], PhD; Xiaopu Shang[1*], PhD; Manlu Liu[2], PhD

[1]School of Economics and Management, Beijing Jiaotong University, Beijing, China

[2]Saunders College of Business, Rochester Institute of Technology, Rochester, NY, United States

[*]these authors contributed equally

Corresponding Author:
Runtong Zhang, PhD
School of Economics and Management
Beijing Jiaotong University
3 Shangyuancun
Haidian District
Beijing, 100044
China
Phone: 86 10 5168 3854
Email: rtzhang@bjtu.edu.cn

Abstract

Background: The internet has become a major mean for acquiring health information; however, Web-based health information is of mixed quality and may markedly affect patients' health-related behavior and decisions. According to the social information processing theory, patients' trust in their physicians may potentially change due to patients' health-information-seeking behavior. Therefore, it is important to identify the relationship between internet health information and patient compliance from the perspective of trust.

Objective: The objective of our study was to investigate the effects of the quality and source of internet health information on patient compliance using an empirical study based on the social information processing theory and social exchange theory.

Methods: A Web-based survey involving 336 valid participants was conducted in China. The study included independent variables (internet health information quality and source of information), 2 mediators (cognition-based trust [CBT] and affect-based trust [ABT]), 1 dependent variable (patient compliance), and 3 control variables (gender, age, and job). All variables were measured using multiple-item scales from previously validated instruments, and confirmative factor analysis as well as structural equation modeling was used to test hypotheses.

Results: The questionnaire response rate was 77.16% (375/486), validity rate was 89.6% (336/375), and reliability and validity were acceptable. We found that the quality and source of internet health information affect patient compliance through the mediation of CBT and ABT. In addition, internet health information quality has a stronger influence on patient compliance than the source of information. However, CBT does not have any direct effect on patient compliance, but it directly affects ABT and then indirectly impacts patient compliance. Therefore, the effect of ABT seems stronger than that of CBT. We found an unexpected, nonsignificant relationship between the source of internet health information and ABT.

Conclusions: From patients' perspective, internet health information quality plays a stronger role than its source in impacting their trust in physicians and the consequent compliance with physicians. Therefore, patient compliance can be improved by strengthening the management of internet health information quality. The study findings also suggest that physicians should focus on obtaining health information from health websites, thereby expanding their understanding of patients' Web-based health-information-seeking preferences, and enriching their knowledge structure to show their specialization and reliability in the communication with patients. In addition, the mutual demonstration of care and respect in the communication between physicians and patients is important in promoting patients' ABT in their physicians.

(J Med Internet Res 2018;20(8):e253) doi:10.2196/jmir.9364

XSL•FO
RenderX

KEYWORDS

affect-based trust; cognition-based trust; internet health information; patient compliance; patient-physician relationship; social information processing theory; social exchange theory; structural equation modeling

Introduction

Background

The patient-physician relationship has become the second most important relationship, following the family relationship [1]. Communication plays an important role in the patient-physician relationship, and patient compliance is important in patient-physician interactions and receives considerable attention from scholars. Specifically, medical diagnoses and treatments can be effective if patients follow their physicians' directions [2]. Patients' attitudes and self-management are proposed to be critical in preventing diseases and promoting communication with physicians [3]. Therefore, the treatment effects of highly compliant patients are better than that of patients with low compliance [4].

Traditionally, patients have been primarily obtaining health information from physicians [5]. However, today, patients have begun to take advantage of the wide range of health information sources available beyond their physicians, including family, friends, and traditional mass media [6]. However, information from these sources cannot meet patients' increasing demands [7]; moreover, patients are sensitive [8] regarding whether they would like to seek information on their own in order to monitor, for example, their physicians' decision making. Also, patients with minor symptoms may prefer to diagnose themselves [9]. The development of the internet has provided patients with a considerable amount of health information [10], and thus, the internet has almost become a major source for patients to seek health information [11]. Nonetheless, health information available on the internet is of mixed quality, and some information is oversimplified, incomplete, inaccurate, or even misleading [12]. The current health information environment provides patients with an access to the information of different quality, and the information can directly influence patients' trust in physicians and their decisions to follow physicians' advice [13]. For example, if the health information from the internet is inconsistent with that from physicians, patients may doubt the advice of their physicians.

The relationship between internet health information and patient compliance has been studied on the basis of a perceived information asymmetry for patients [1]. This study, however, actually uses the perspective of psychology to study patient compliance. Reportedly, the interrelationship between patients' behavior and their health is complex, and psychological factors, such as motivation, play an important role in this relationship [14]. Recently, researchers have begun to focus on behavioral psychology as well as its application in healthy choices and patient-physician interactions [15]. However, related previous studies that have discussed patient compliance from the angle of psychology are limited. Patient compliance, in fact, is a dynamic parameter that sometimes changes unintentionally because of cognitive deficiencies such as poor attention. In addition, noncompliance stems from other factors such as psychosocial stress [16]. Therefore, in this study, based on the social processing theory and social exchange theory, we intended to explore how internet health information impacts patient compliance through mediation of trust.

Internet Health Information

The internet has become an important source of health information [17]. An increasing number of institutions, including governments, medical institutions, and business corporations, have established health information portals to provide public health information and to meet growing demands for such information [18]. However, problems like confusion and uncertainty about information quality remain serious. For example, mismatches between the Web-based health information obtained by patients and the actual demands of patients may arise [19]. Hence, it is important to conduct studies focusing on the internet health information. Previous studies have proposed that many people worry about the quality of internet health information, which is an important problem [20]. A marked and indirect relationship between the quality of internet health information and patient compliance has been identified; however, information asymmetry is considered to be a nonsignificant mediator [1]. Therefore, in this study, we used trust as another mediator between internet health information quality and patient compliance. Internet health information quality is used [1] to describe the information fitness for use and information reliability [3], comprising 4 dimensions: relevance, understandability, adequacy, and usefulness [1]. In terms of quantity, patients commonly acquire health information through search engines, such as Google, Bing, and Yahoo, and many health websites are available [12]; thus, a majority of patients have access to a great amount of internet health information regardless of their health literacy. Moreover, they tend to trust and select the first few results provided by these search engines [21]. On the other hand, quality includes adequacy as a dimension. In that case, we did not consider the quantity of Web-based health information in this study.

The source of internet health information is a critical attribute [22], and its evaluation in the health-information-seeking process is important [23]. Numerous health websites exist that have been built by several institutions, and one factor influencing patients' selection of websites is their trust [6]. Therefore, how the source of internet health information impacts patient compliance should also be considered. The source of internet health information can be described using several related attributes (eg, reliability, accessibility, trustworthiness, and authority) [21,24]. In this study, we have described the source of internet health information with reliability, authority, and accessibility. In summary, we have discussed internet health information from the perspectives of its quality and source.

Patient Compliance

Patient compliance is an important term that represents how patients follow the medical diagnoses and treatment regimens recommended by their physicians [1]; it plays a vital role in the

XSL•FO
RenderX

patient-physician relationship. Patients' attitudes and personal involvement, including self-management and self-monitoring, are all propitious to the improvement of patient compliance [8]. Patient compliance is mainly manifested in 2 aspects as follows: (1) maintaining a healthy lifestyle by following physicians' advice and (2) medicine adherence. Khera et al [25], Stonerock and Blumenthal [26], and Johal et al [27] proposed that maintaining a healthy lifestyle is beneficial for preventing cardiovascular disease and reducing its incidence. In addition, living a healthy life has a substantial effect on decreasing the risk of cancer [28,29]. Nevertheless, when physicians suggest that patients make major changes in lifestyle, the patients are unlikely to comply [26]. In terms of medicine, for example, Varleta et al [30] concluded that a lack of adherence to taking medicines according to the prescription may lead to poor blood pressure control.

Noncompliance in health care may have three consequences. (1) For patients: the probability of morbidity and mortality is likely to increase [31]. (2) For economy: medical productivity and resources may be wasted because patients ignore the medical diagnoses and treatment regimens recommended by their physicians [31,32]. (3) For society: the use of genuinely beneficial drugs may be terminated due to noncompliance in clinical practice [31]. The treatment of chronic diseases largely relies on self-management and self-monitoring by patients [4]; thus, improving patient compliance may lead to better health-related outcomes than discovering any new therapy, given the increasing proportion of patients with chronic diseases [32]. However, the proportion of high compliant patients is relatively low, with a study revealing the rate of patients' noncompliance with medicine to be as high as 50% [32]. Therefore, in this study, we aimed to discuss the improvement of patient compliance.

Trust

Patients' trust in their physicians is the core of the patient-physician relationship [33]. Trust has been discussed and defined in many previous studies, and this study focuses on the interpersonal trust between patients and physicians. Interpersonal trust is a pervasive phenomenon defined as "the extent to which a person is confident in, and willing to act on the basis of, the words, actions, and decisions of another" [34]. Furthermore, interpersonal trust is conceptualized into two different dimensions [34]: (1) cognition-based trust (CBT), which is grounded in the available knowledge, competence, and responsibility of individuals [35], and (2) affect-based trust (ABT), which is grounded in mutual respect, genuine care, and concern for the needs of others [36]. The patient-physician relationship is an important interpersonal relationship. The trust that patients have in physicians has been studied frequently; however, previous literature has rarely studied CBT and ABT in the patient-physician relationship. Nonetheless, these two types of trust have been considered to be important factors in behavioral studies [37], and they explain behavior from different points. Specifically, CBT is associated with individuals' perceived competence and ABT with emotional connections [38]. Furthermore, cognition and affect are strongly linked [39], and CBT is built more easily than ABT [37]. Therefore, to effectively study the patient-physician relationship, trust is considered on both the cognitive and affective levels [40].

In CBT, when patients perceive their physicians as reliable, competent, and likely to offer useful help, they may be willing to trust these physicians [41]. While exchanging health information with their physicians, if the patients find that their physicians have shared or are sharing health information that is consistent with what they obtain from the internet, they may presume that their physicians are professional; thus, they will cognitively trust these physicians [41]. Consequently, they may follow the medical diagnoses and treatment regimens that these physicians propose. Regarding ABT, when an emotional connection is established, patients may feel safer while communicating with their physicians. Thus, they may be willing to show their vulnerability and are likely to express their personal attitudes related to health [41]. Therefore, disagreements, conflicts, and biased speculations can be reduced because of the full processing of information between patients and physicians [41]. In addition, CBT can impact ABT [34].

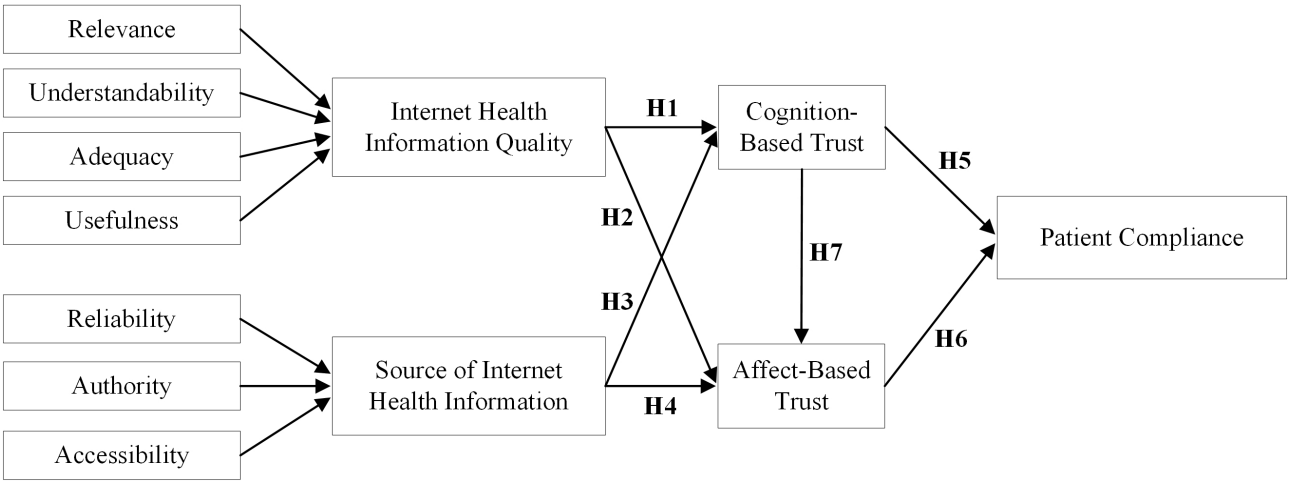
Model and Hypotheses

Internet health information is advantageous to health-related decision making and patient-physician communication [42]. In this study, we investigated patient compliance by focusing on internet health information. Figure 1 shows the research model describing how internet health information impacts patient compliance through the mediation of CBT and ABT.

According to the social information processing theory [13], social contexts provide environmental cues, such as social information, which influence people's behavioral options [43]. Especially when individuals do not have sufficient suitable information related to their targets, they are more likely to seek information from other sources, which may shape their attitudes, beliefs, and opinions [44]. In this study, we applied the social information processing theory to the patient-physician relationship, in the context of internet health information. Before or after visiting physicians, patients tend to seek health information from other sources in addition to physicians, such as books, news, mass media, and, especially, the internet [1]. Patients compare this obtained health information with that obtained from the patient-physician communication, which further helps establish their attitudes toward their physicians and may, in turn, influence their treatment behavior [45], such as patient compliance. In addition, the patient-physician relationship is a type of social exchange relationship [46], in which resources (love, status, information, money, goods, and services) are exchanged between physicians and patients [47,48]. Patients want to acquire correct information and suitable treatment, and physicians want to achieve the satisfaction of their patients, which eventually enhances doctors' reputations. In that case, patients' attitudes based on social information help them decide whether or not to exchange resources with their physicians. For example, if patients hold the view that their physician is unprofessional, they may switch to another physician and even provide negative comments about the first physician when other patients ask their advice.

XSL•FO
RenderX

Figure 1. Research model.



To date, the significant relationship between internet health information quality and patients' trust in physicians has not been directly supported by evidence. Reportedly, trust is a dynamic parameter [49], and the trust that patients have in their physicians may change because of certain influencing factors [50]. When patients obtain high-quality health information from the internet, they are likely to establish correct health-related views and beliefs, which, in turn, may cause them to realize that their physicians have truly shared useful health-related knowledge with them [1]. In such a case, patients are willing to trust their physicians. Hence, we proposed the following hypotheses:

> H1: Internet health information quality has a positive impact on patients' CBT in their physicians.
>
> H2: Internet health information quality has a positive impact on patients' ABT in their physicians.

Undoubtedly, the trust that patients place in internet health information changes with the information source, and the use of such information is influenced by their trust [6]. A good information source, such as an official health website, can provide high-quality information and help patients correctly understand their health conditions, as well as their physicians' medical diagnoses and treatment regimens. Therefore, patients may perceive that they have received or are receiving suitable therapies from their physicians, which promotes patients' trust in their physicians. Therefore, we proposed the following hypotheses:

> H3: The source of internet health information has a positive impact on patients' CBT in their physicians.
>
> H4: The source of internet health information has a positive impact on patients' ABT in their physicians.

As an interpersonal relationship, interpersonal trust plays an important role in the physician-patient relationship [50,51] and is also one of the critical principles of effective social exchange [46]. A previous study has proposed that CBT and ABT have direct effects on behaviors [52]. Regarding CBT, when the Web-based health information is consistent with that obtained from the physicians, the patients may cognitively trust their physicians. Then, the reliability and competence of physicians could be established from the patients' points of view [34], by removing the uncertainty of the patient-physician relationship [53]. Consequently, patients are increasingly willing to trust the information provided by their physicians [41], and they feel the obligation to comply with physicians' medical diagnoses and treatment regimens [54]. Therefore, we proposed the following hypothesis:

> H5: Patients' CBT in their physicians has a positive impact on the compliance.

In the patient-physician relationship, ABT is conducive to establishing an emotional connection based on mutual respect, care, and concern [36] and promoting a sense of safety in patients about expressing themselves when they interact with their physicians. Based on the social information processing theory, patients' ABT in their physicians comes from the realization that their physicians are reliable and dependable, as evidenced by the health information that the patients have obtained through the internet. Consequently, patients tend to feel that they are taken seriously and are, thus, willing to communicate with their physicians and obey their physicians' recommendations, as the reciprocation of physicians' sincere treatments. Thus, disagreements, conflicts, and biased speculations may be reduced [41], and patients' trust may lead to a high patient compliance. Consequently, we suggested the following hypothesis:

> H6: Patients' ABT in their physicians has a positive impact on the compliance.

CBT impacts ABT [34] and promotes patients to feel at ease in response to a reliable and professional atmosphere. In turn, patients come to regard their physicians as being reliable and gradually become emotionally dependent on them; consequently, ABT is established. Hence, this situation led us to derive the following hypothesis:

> H7: Patients' CBT in their physicians has a positive impact on their ABT in physicians.

Methods

Instrument Development

We used a multiple-item measurement scale to measure the constructs and previously validated instruments, which have

XSL•FO
RenderX

been used in published works, for instrument development to ensure reliability and validity. A 7-point Likert-type response format that ranged from "strongly disagree" to "strongly agree" was used to measure items. The 5 variables in the research model (see Figure 1) were covered by the survey instrument (Multimedia Appendix 1). Patient compliance, which was discussed from a different perspective by Laugesen et al [1], was measured using a 5-item scale from the same reference. CBT and ABT were proposed by Mcallister [34] and measured by the same author using 6-item and 5-item scales, respectively. To address the subject of this study, these two scales were adjusted for measuring the CBT and ABT of patients in physicians. Internet health information quality consisted of the factors of relevance, understandability, adequacy, and usefulness [1] and was measured using a 16-item scale from Laugesen et al [1]. Similarly, the source of the internet health information, comprising the dimensions of reliability, authority, and accessibility, was measured using a 28-item scale divided into 3 parts. Specifically, reliability was measured using a scale from Singh et al [55] and authority and accessibility measured using an 18-item and a 5-item scale from Provost et al [24], respectively.

Analysis Tool Selection

Research methods from previous studies [1,56] have used structural equation modeling (SEM) to analyze relationships between variables and to test hypotheses. In contrast with simple correlation-based models, we included mediators and complex relationships between variables in this research model. SEM has been widely accepted in several studies [57], and it accommodates intricate causal networks [1,58], such as testing hypotheses covering all variables and analyzing causal relationships in research models [59]. We used SEM for the following two reasons: (1) the measurement error can be incorporated, and the detected effects can be provided power through SEM and (2) the research model can be improved through a combination with confirmative factor analysis [60]. We used IBM's SPSS 22.0 and Amos 22.0 (Armonk, New York, United States), which can achieve efficient and unbiased analysis and evaluate the latent variable interactions.

Data Collection and Respondent Profile

The scales had to be translated into Chinese because the questionnaires would be distributed among Chinese respondents in China. First, as was done with the translation process in previous works [61,62], we recruited native Chinese speakers who had a master's degree and above and were fluent in speaking English as well as skilled in scientific research translation to translate our scales into Chinese. As cross-cultural adaptation had to be considered [63], certain expressions needed to be modified. Second, 10 individuals from different professions and different ages, genders, and educational levels were invited to read the translated scales and provide recommendations for our modifying scales, consequently ensuring the comprehensibility, appropriateness, and readability in the context of Chinese culture. Finally, the scales underwent a reverse translation process performed by an English-speaking professional to check for the conceptual discrepancies and to ensure consistency with the original English version.

In the weeks preceding the formal investigation, a pretest was conducted with 112 subjects to ensure the clarity, conciseness, and readability of the scales and to determine the approximate time required to complete the questionnaire. Our subjects were Chinese individuals who had received medical therapies within the previous month and had sought Web-based health information. The formal investigation was anonymously conducted through a Web-based questionnaire survey addressed to participants in June 2017. The respondents were assured that their privacy was protected, and their informed consent was secured. Moreover, to control the duplication of responses, we only accepted the first response from the same IP address and deleted other responses within 1 hour. Of course, we did not tell participants this rule.

With the help of a medical association in China, we sent questionnaires to 486 participants and received a total of 375 responses, 336 of which were valid and covered 28 provinces of China (except Macao, Qinghai, Ningxia, Tibet, Xinjiang, and Hainan). Therefore, the response rate was 77.16% (375/486) and the validity rate was 89.6% (336/375). Table 1 presents the demographics of the research sample, in which 56.5% (190/336) participants were 20-40 years old, 53.6% (180/336) were females, and 62.2% (209/336) had, at least, a college education or a bachelor's degree. Thus, more than half of this sample was young, female, and highly educated. A previous study has also reported that internet health information users are likely to be young, female, and educated [64]. The objective of our study was to identify the relationship between internet health information and patient compliance, and the investigative channel used was the internet. Therefore, the sample met our requirements.

XSL•FO
RenderX

Table 1. Sample demographics (N=336).

Demographic characteristics	Participants, n (%)
Age (years)	
<20	22 (6.6%)
20-29	83 (24.7%)
30-39	107 (31.8%)
40-49	59 (17.6%)
50-59	47 (14.0%)
≥60	18 (5.4%)
Gender	
Male	156 (46.4%)
Female	180 (53.6%)
Resident status	
Urban	184 (54.8%)
Rural	152 (45.2%)
Education	
Junior middle school	31 (9.2%)
High school	96 (28.6%)
Junior college	68 (20.2%)
Bachelor's degree	127 (37.8%)
Master's degree	9 (2.7%)
Doctor's degree	5 (1.5%)
Job	
Private business owners	28 (8.3%)
Factory workers	31 (9.2%)
Professional and technical workers	77 (22.9%)
Commercial service workers	63 (18.8%)
Students	38 (11.3%)
Liberal professionals	27 (8.0%)
Employees in government offices and public institutions	40 (11.9%)
Retirees	22 (6.5%)
Farmers	10 (3.0%)

Results

Data Analysis

We analyzed data on the basis of methods from previous studies [1,56]. The reliability and validity of the measures were analyzed using the SPSS 22.0 software. Cronbach alpha, which was used to assess the reliability, needed to be at least .700 [59]. Table 2 presents the Cronbach alpha of each construct, and these results show that the scale in this study had good reliability. The Kaiser-Meyer-Olkin value (weak, .500; medium, .600; good, .700; very good, .800; and perfect, .900) [65-68] was equal to .907 (P<.001, significant) above the cutoff value of .900; thus, the construct validity was fully acceptable.

In accordance with a study by Wu et al [69], we evaluated the discriminant validities of the constructs and ensured whether internet health information quality, the source of internet health information, CBT and ABT, and patient compliance were distinct from each other and from the indicators loaded onto their intended latent variables, by means of nested confirmatory factor analytic models. We established and compared 6 nested models based on the research model (see Figure 1): (1) a 5-factor model treating each of the variables as separate factors; (2) a 4-factor model treating internet health information quality and the source of internet health information as the first factor, CBT as the second factor, ABT as the third factor, and patient compliance as the fourth factor; (3) a 4-factor model treating internet health information quality as the first factor, treating the source of internet health information as the second factor,

XSL•FO
RenderX

CBT and ABT as the third factor, and patient compliance as the fourth factor; (4) a 3-factor model treating internet health information quality and source of internet health information as the first factor, CBT and ABT as the second factor, and patient compliance as the third factor; (5) a 2-factor model treating internet health information quality, the source of internet health information, CBT, and ABT as the first factor and patient compliance as the second factor; and (6) a 1-factor model treating all 5 factors as one factor. As shown in Table 3, there was a good fit between the data and the 5-factor model (model 1; χ^2_{1390}=1793.1, χ^2/df=1.29<3; comparative fit index=.96>.90; Tucker-Lewis index=.95>.90; incremental fit index=.96>.90; root mean square error of approximation=.029<.050) [70-74]. Compared with model 1, the other 5 nested models (Models 2-6) were worse fits to the data, according to all fit indices. Therefore, we concluded that internet health information quality, the source of internet health information, CBT and ABT, and patient compliance were 5 different factors.

Hypothesis Testing

The demographical statistics were used to identify any significant relationship between the demographic factors and variables of the research model [1]. Our analytical results indicated the following: (1) Gender: the relationship between gender and CBT and that between gender and patient compliance was significant. Specifically, females were more likely than males to cognitively trust their physicians and behaved with higher compliance. (2) Age: according to the results of the analysis of covariance, age exhibited a marked effect on the relationship between internet health information quality and patient compliance and between ABT and patient compliance. (3) Job: private business owners held more positive attitudes about the sources of internet health information (health websites) than commercial service workers, students, and liberal professionals. Therefore, we added gender, age, and job as control variables into the research model.

Table 2. Cronbach alpha of the constructs.

Constructs	Cronbach alpha
Internet health information quality	.933
Source of internet health information	.910
Cognition-based trust (CBT)	.865
Affect-based trust (ABT)	.756
Patient compliance	.870
Total[a]	.950

[a]For the total value, all five constructs were regarded as one and were used to calculate the total Cronbach alpha.

Table 3. Comparison of measurement models in confirmatory factor analysis.

Distinctiveness test for all variables (model factors)	Fit indices					
	χ^{2a} (df^b)	χ^2/df	RMSEA[c]	CFI[d]	IFI[e]	TLI[f]
Model 1 (5 factors): internet health information quality, source of internet health information, CBT[g], ABT[h], patient compliance	1793.1 (1390)	1.29	.029	.96	.96	.95
Model 2 (4 factors): internet health information quality and source of internet health information combined into 1 factor	2322.5 (1394)	1.67	.045	.91	.92	.89
Model 3 (4 factors): CBT and ABT combined into 1 factor	1878.8 (1394)	1.35	.032	.96	.96	.94
Model 4 (3 factors): internet health information quality and source of internet health information combined into 1 factor and CBT and ABT combined into 1 factor	2399.4 (1397)	1.72	.046	.91	.91	.88
Model 5 (2 factors): internet health information quality, source of internet health information, CBT and ABT combined into 1 factor	3046.8 (1399)	2.18	.059	.85	.85	.81
Model 6 (1 factor): internet health information quality, source of internet health information, CBT and ABT, and patient compliance combined into 1 factor	3303.1 (1400)	2.36	.064	.83	.83	.78

[a]χ^2: Pearson chi-square.

[b]*df*: degrees of freedom.

[c]RMSEA: root mean square error of approximation.

[d]CFI: comparative fit index.

[e]IFI: incremental fit index.

[f]TLI: Tucker-Lewis index.

[g]CBT: cognition-based trust.

[h]ABT: affect-based trust.

XSL•FO
RenderX

First, we used a hierarchical multiple linear regression method to test our hypotheses and to evaluate the effects of the control variables. Table 4 presents the path coefficient and significance of each relationship and shows that the relationships proposed by H1, H2, H3, H5, H6, and H7 were significant; however, the relationship hypothesized by H4 was nonsignificant. In addition, Cohen f^2 [75] was used to assess the effects of the control variables, with results divided into several categories (ie, insignificant: <.020; small: ≥.020 and <.150; medium: ≥.150 and <.300; and large: ≥.350). As shown in Table 5, the effect sizes of gender, age, and job were all small. Table 6 shows the effect sizes of variables.

We found that only the effect size of CBT on ABT was large, whereas the effect size of CBT on patient compliance was small. In addition, the effect sizes of the quality as well as source of internet health information and ABT were all small.

Table 4. Results of hierarchical multiple linear regression.

Hypothesis	Path coefficient	*P* value
H1: Internet health information quality → CBT[a]	.317	<.001
H2: Internet health information quality → ABT[b]	.213	<.001
H3: Source of internet health information → CBT	.224	<.001
H4: Source of internet health information → ABT	.076	.13
H5: CBT → patient compliance	.326	<.001
H6: ABT → patient compliance	.378	<.001
H7: CBT → ABT	.535	<.001

[a]CBT: cognition-based trust.

[b]ABT: affect-based trust.

Table 5. Multivariate coefficient of determination (R^2) results, where ΔR^2 is $R^2_{\text{with control variables}} - R^2_{\text{without control variables}}$.

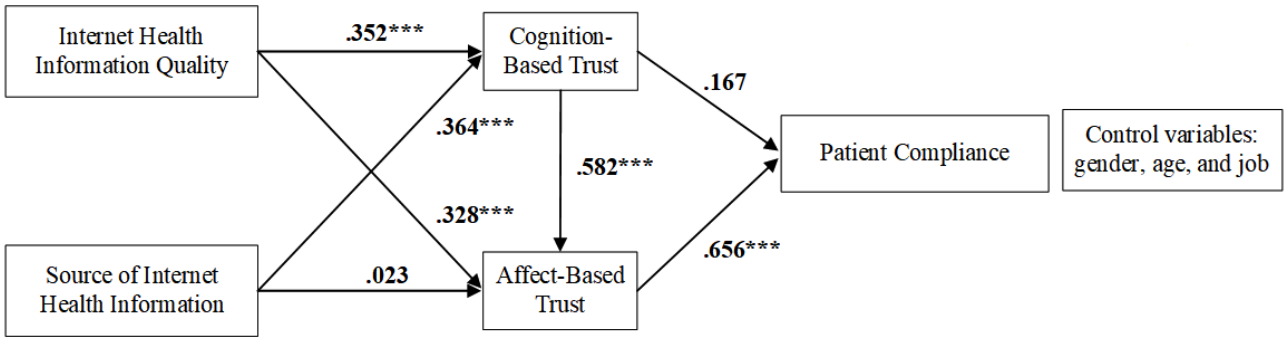
Variables	R^2		Control variable effects		
	With control variables	Without control variables	ΔR^2	Cohen f^2	Effects
Cognition-based trust (CBT)	0.252	0.229	0.023	0.031	Small
Affect-based trust (ABT)	0.487	0.470	0.017	0.033	Small
Patient compliance	0.461	0.443	0.018	0.033	Small

Table 6. Hierarchical multiple linear regression effect size analysis, where R^2 is multivariate coefficient of determination and ΔR^2 is $R^2_{\text{with control variables}} - R^2_{\text{without control variables}}$.

Variables	R^2		ΔR^2	Cohen f^2	Effect size
	In	Out			
Patient compliance					
Cognition-based trust (CBT)	.461	.406	.055	.102	Small
Affect-based trust (ABT)	.461	.388	.073	.135	Small
Cognition-based trust (CBT)					
Internet health information quality	.252	.185	.067	.090	Small
Source of internet health information	.252	.219	.033	.044	Small
Affect-based trust (ABT)					
Internet health information quality	.487	.459	.028	.055	Small
Source of internet health information	.487	.483	.004	.008	Small
Cognition-based trust (CBT)	.487	.272	.215	.419	Large

XSL•FO
RenderX

Figure 2. Research model with path coefficients. ***P<.001, *P<.05.



We used SPSS 22.0 and Amos 22.0 to test our hypotheses and found that age had a positive effect on patient compliance and that older patients were more likely to follow their physicians' suggestions. Females were more willing to cognitively trust their physicians and comply with their physicians, whereas males were more likely to have ABT in their physicians. The job type only significantly impacted ABT. Specifically, liberal professions held the highest ABT in their physicians, whereas private business owners were least likely to affectively trust their physicians.

Figure 2 indicates the SEM results, and the magnitude and significance of the path coefficients are shown in Table 7. Five hypotheses were supported (H1, H2, H3, H6, and H7), but H4 and H5 were not supported, and we have provided the possible reasons for these nonsignificant relationships in the next section.

In addition, we used the bootstrapping method (n=5000, 95% CI) to further analyze the mediating effects in the research model.

As Table 8 reveals, we first concluded that the significant mediating effects of CBT and ABT between the source of internet health information and patient compliance were nonsignificant. In addition, Sobel test was used to evaluate the mediating role played by CBT between internet health information quality and patient compliance. To specify, we found that the value of Z was .489, significantly lower than .900, indicating that the mediation of CBT was nonsignificant. Furthermore, ABT had a significant mediating effect on the relationship between internet health information quality and patient compliance, with all the effects (direct, indirect, and total) being significant.

Table 7. Hypothesis testing.

Hypothesis	Path coefficient	P value
Internet health information quality has a positive impact on patients' CBT[a] in their physicians.	.352	<.001
Internet health information quality has a positive impact on patients' ABT[b] in their physicians.	.328	<.001
The source of the internet health information has a positive impact on patients' CBT in their physicians.	.364	<.001
The source of the internet health information has a positive impact on patients' ABT in their physician.	.023	.72
Patients' CBT in their physicians has a positive impact on the compliance.	.167	.09
Patients' ABT in their physicians has a positive impact on the compliance.	.656	<.001
Patients' CBT in their physicians has a positive impact on their ABT in physicians.	.582	<.001

[a]CBT: cognition-based trust.

[b]ABT: affect-based trust.

XSL•FO
RenderX

Table 8. Path coefficients by bootstrapping method. Amos 22.0 used to calculate direct, indirect, and total effects.

Effect	Path coefficient (SD)	*P* value
Direct effects		
internet health information quality → CBT[a]	.309 (.076)	.001
source of internet health information → CBT	.310 (.069)	<.001
CBT → patient compliance	.142 (.288)	.46
internet health information quality → ABT[b]	.308 (.073)	<.001
source of internet health information → ABT	.062 (.074)	.39
ABT → patient compliance	.737 (.330)	.001
Indirect effects		
internet health information quality → patient compliance	.412 (.111)	<.001
source of internet health information → patient compliance	.232 (.081)	.008
Total effects		
internet health information quality → patient compliance	.393 (.067)	<.001
source of internet health information → patient compliance	.083 (.076)	.28

[a]CBT: cognition-based trust.

[b]ABT: affect-based trust.

Discussion

Principal Results

This is the first study that discusses patient compliance from the perspective of both CBT and ABT. This study makes several theoretical contributions and practical implications for future study of patient compliance and ways to improve the patient-physician relationship. First, we constructed a research model to identify the relationship between internet health information and patient compliance, mediated by trust. We clarified the mechanisms through which internet health information (quality and source) impacts patient compliance. To specify, we used CBT and ABT that patients have in their physicians as the mediators. Internet health information quality directly impacts patients' CBT and ABT in their physicians and indirectly impacts patient compliance. Laugesen et al [1] identified the indirect impact that high-quality internet health information could increase patient compliance. Therefore, patient compliance can be improved by strengthening the management of certain aspects of internet health information quality, such as the information topics, categories, meaning, and usability. In addition, the path coefficient from internet health information quality to CBT (.352) was higher than that from internet health information quality to ABT (0.328), and the source of internet health information only directly affected CBT but did not have any direct effect on ABT. Hence, we found that internet health information (quality and source) had more significant impacts on CBT than on ABT, implying that individuals always rationally deal with internet health information on the basis of their own cognition. Thus, CBT in the patient-physician relationship can be improved by improving the quality and source of internet health information. For example, physicians should focus on the health information obtained from health websites to understand the health-information-seeking preferences of their patients. Moreover, physicians should communicate with their patients using health websites, which can enable them to share health information with their patients through Web and establish CBT.

Second, ABT has positive effects on patient compliance, but we did not identify any significant direct effect of CBT on patient compliance. In accord with Mcallister's [34] proposal, we too found that CBT directly impacts ABT, thus, indirectly impacting patient compliance. In other words, ABT has a mediating effect on the relationship between CBT and patient compliance. Lee et al [76] reported that trust in physicians was related to patient compliance, such that when patients highly trusted their physicians, they appeared to be more likely to report their health status to these physicians; this finding indicates that patient compliance can be improved by enhancing patients' CBT and ABT in their physicians. On the one hand, physicians are advised to enrich their knowledge structures to show their specialization and reliability in the interaction with their patients to inspire their CBT. On the other hand, communication is important in the patient-physician relationship, and physicians should focus on mutual care and respect for patients in communication to promote patients to affectively trust their physicians. In addition, we suggest that physicians obtain internet health information to enrich themselves and they should actively participate in discussions with their patients on health websites to establish a good atmosphere for communication.

Third, and surprisingly, a nonsignificant relationship was found between the source of internet health information and ABT in physicians. We considered that there might be suppression effects of other factors in the model. Thus, we removed the relationships between (1) internet health information quality and ABT, (2) the source of internet health information and CBT, and (3) CBT and ABT and retained only the relationship between the source of internet health information and ABT. Then, we added the three abovementioned relationships to the

XSL•FO
RenderX

model one by one and investigated how each of them affected the strength of the path from the source of internet health information to ABT. Eventually, we found that the quality of internet health information played a critical role in impacting the relationship between the source of internet health information and ABT. Furthermore, this relationship became nonsignificant when the relationship between internet health information quality and ABT was added to the model. According to the results of bootstrapping, both the quality and the source of internet health information have indirect effects on patient compliance, but only the quality totally affects patient compliance. In addition, the path coefficient of the indirect relationship between internet health information quality and patient compliance is larger than that between internet health information source and compliance. In fact, ABT is based on an emotional connection and subjective judgment. Therefore, the quality of internet health information may be stronger than the source, from the patients' point of view, because the demand for quality may be greater than that for the source whose impact on ABT was ignored in this analysis.

Limitations

The limitations of the study must be considered. First, we focused on only 2 dimensions of internet health information: quality and source. Other dimensions may also be worthy of investigation. Second, in this study, we examined the relationship between internet health information and patient compliance in the context of China. However, China is a special country in terms of health care due to its large population and unbalanced health care development in different areas. In addition, China is currently actively promoting Web-based health care. There may be several common and different aspects between China and other places, which can be addressed in future research through additional surveys. Third, a cross-sectional survey was used to collect data from respondents. Thus, changes in patient compliance that were associated with the attitudes of patients toward internet health information may not have been captured. Fourth, this study proposed universal and guiding suggestions on the basis of discussions about internet health information quality and source. Future studies may aim to evaluate the quality and source of internet health information, focusing on and collecting user data from specific health websites. Fifth, all concepts and relationships were measured only once. This study was conducted from a static perspective and, therefore, failed to consider dynamic changes in patient attitudes. Last, although our sample met the characteristics of typical internet health information seekers, we did not consider the feature of Chinese census data, and the number of respondents was relatively small.

Conclusions

This study indicates that both the quality and source of internet health information markedly impact patient compliance through the mediations of CBT and ABT. In our research model, the quality of internet health information showed a stronger effect on patient compliance than the source of that information and the information quality also showed a positive impact on CBT and ABT. Consequently, health information quality indirectly affects patient compliance. In terms of trust, ABT appears to have a stronger effect than CBT on patient compliance. These findings suggest the follwing: (1) patient compliance can be improved by strengthening the management of certain aspects of internet health information quality, such as its topics, categories, meaning, and usability; (2) physicians could focus on obtaining information from health websites to understand patients' health-information-seeking preferences and to enrich their own knowledge to enhance their specialization and reliability; (3) physicians can communicate with patients on health websites and share information as well as establish CBT with them through Web; and (4) a mutual demonstration of care and respect in the communication between physicians and patients is important during treatment and is beneficial in promoting patients' affective trust in their physicians.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (key program) [grant number 71532002] and the Fundamental Research Funds for the Central Universities [grant number 2017YJS068]. The authors also appreciate the support of the Beijing Logistics Informatics Research Base.

Authors' Contributions

XL, RZ, WW, and XS conceived and designed the study, developed the research model, designed the questionnaire, conducted data collection and analysis, and drafted as well as modified the manuscript; ML modified the manuscript and improved the languages. All authors approved the final version of the manuscript for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Measurement instruments.

[PDF File (Adobe PDF File), 47KB - jmir_v20i8e253_app1.pdf]

References

XSL•FO
RenderX

1. Laugesen J, Hassanein K, Yuan Y. The Impact of Internet Health Information on Patient Compliance: A Research Model and an Empirical Study. J Med Internet Res 2015;17(6):e143 [FREE Full text] [doi: 10.2196/jmir.4333] [Medline: 26068214]
2. Tustin N. The role of patient satisfaction in online health information seeking. J Health Commun 2010 Jan;15(1):3-17. [doi: 10.1080/10810730903465491] [Medline: 20390974]
3. Deng Z, Liu S, Hinz O. The health information seeking and usage behavior intention of Chinese consumers through mobile phones. Info Technology & People 2015 Jun;28(2):405-423. [doi: 10.1108/ITP-03-2014-0053]
4. Horwitz RI, Horwitz SM. Adherence to treatment and health outcomes. Arch Intern Med 1993 Aug 23;153(16):1863-1868. [Medline: 8250647]
5. Kassirer JP. Patients, physicians, and the Internet. Health Aff (Millwood) 2000;19(6):115-123 [FREE Full text] [Medline: 11192394]
6. Ruppel EK. Scanning Health Information Sources: Applying and Extending the Comprehensive Model of Information Seeking. J Health Commun 2016;21(2):208-216. [doi: 10.1080/10810730.2015.1058438] [Medline: 26716985]
7. Hu Y, Shyam Sundar S. Effects of Online Health Sources on Credibility and Behavioral Intentions. Communication Research 2009 Nov 25;37(1):105-132. [doi: 10.1177/0093650209351512]
8. Cohn A, Richters J. 'My Vagina Makes Funny Noises': Analyzing Online Forums to Assess the Real Sexual Health Concerns of Young People. International Journal of Sexual Health 2013 Apr;25(2):93-103. [doi: 10.1080/19317611.2012.719852]
9. Mendes Á, Abreu L, Vilar-Correia MR, Borlido-Santos J. "That Should Be Left to Doctors, That's What They are There For!"-Exploring the Reflexivity and Trust of Young Adults When Seeking Health Information. Health Commun 2017 Sep;32(9):1076-1081. [doi: 10.1080/10410236.2016.1199081] [Medline: 27485762]
10. Clarke MA, Moore JL, Steege LM, Koopman RJ, Belden JL, Canfield SM, et al. Health information needs, sources, and barriers of primary care patients to achieve patient-centered care: A literature review. Health Informatics J 2016 Dec;22(4):992-1016. [doi: 10.1177/1460458215602939] [Medline: 26377952]
11. Yang Q, Chen Y, Wendorf MJ. Social Support, Trust in Health Information, and Health Information-Seeking Behaviors (HISBs): A Study Using the 2012 Annenberg National Health Communication Survey (ANHCS). Health Commun 2017 Sep;32(9):1142-1150. [doi: 10.1080/10410236.2016.1214220] [Medline: 27589249]
12. Seçkin G, Yeatts D, Hughes S, Hudson C, Bell V. Being an Informed Consumer of Health Information and Assessment of Electronic Health Literacy in a National Sample of Internet Users: Validity and Reliability of the e-HLS Instrument. J Med Internet Res 2016 Jul 11;18(7):e161 [FREE Full text] [doi: 10.2196/jmir.5496] [Medline: 27400726]
13. Salancik GR, Pfeffer J. A social information processing approach to job attitudes and task design. Adm Sci Q 1978 Jun;23(2):224-253. [Medline: 10307892]
14. Doi SAR, Amigo MF. Nurses' intentions to wear gloves during venipuncture procedures: a behavioral psychology perspective. Infect Control Hosp Epidemiol 2007 Jun;28(6):747-750. [doi: 10.1086/517978] [Medline: 17520556]
15. Swindell JS, McGuire AL, Halpern SD. Beneficent persuasion: techniques and ethical guidelines to improve patients' decisions. Ann Fam Med 2010;8(3):260-264 [FREE Full text] [doi: 10.1370/afm.1118] [Medline: 20458111]
16. Umaki TM, Umaki MR, Cobb CM. The psychology of patient compliance: a focused review of the literature. J Periodontol 2012 Apr;83(4):395-400. [doi: 10.1902/jop.2011.110344] [Medline: 21819248]
17. Ritterband LM, Borowitz S, Cox DJ, Kovatchev B, Walker LS, Lucas V, et al. Using the internet to provide information prescriptions. Pediatrics 2005 Nov;116(5):e643-e647. [doi: 10.1542/peds.2005-0404] [Medline: 16263978]
18. Li F, Li M, Guan P, Ma S, Cui L. Mapping publication trends and identifying hot spots of research on Internet health information seeking behavior: a quantitative and co-word biclustering analysis. J Med Internet Res 2015 Mar 25;17(3):e81 [FREE Full text] [doi: 10.2196/jmir.3326] [Medline: 25830358]
19. Pang PC, Chang S, Verspoor K, Pearce J. Designing Health Websites Based on Users' Web-Based Information-Seeking Behaviors: A Mixed-Method Observational Study. J Med Internet Res 2016 Jun 06;18(6):e145 [FREE Full text] [doi: 10.2196/jmir.5661] [Medline: 27267955]
20. Maloney EK, D'Agostino TA, Heerdt A, Dickler M, Li Y, Ostroff JS, et al. Sources and types of online information that breast cancer patients read and discuss with their doctors. Palliat Support Care 2015 Apr;13(2):107-114 [FREE Full text] [doi: 10.1017/S1478951513000862] [Medline: 24182945]
21. Rowley J, Johnson F, Sbaffi L. Gender as an influencer of online health information-seeking and evaluation behavior. J Assn Inf Sci Tec 2015 Aug 25:n/a-n/a. [doi: 10.1002/asi.23597]
22. Kostagiolas P, Korfiatis N, Kourouthanasis P, Alexias G. Work-related factors influencing doctors search behaviors and trust toward medical information resources. International Journal of Information Management 2014 Apr;34(2):80-88. [doi: 10.1016/j.ijinfomgt.2013.11.009]
23. Sbaffi L, Rowley J. Trust and Credibility in Web-Based Health Information: A Review and Agenda for Future Research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
24. Provost M, Koompalum D, Dong D, Martin BC. The initial development of the WebMedQual scale: domain assessment of the construct of quality of health web sites. Int J Med Inform 2006 Jan;75(1):42-57. [doi: 10.1016/j.ijmedinf.2005.07.034] [Medline: 16169770]

XSL•FO
RenderX

25. Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, et al. Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. N Engl J Med 2016 Dec 15;375(24):2349-2358 [FREE Full text] [doi: 10.1056/NEJMoa1605086] [Medline: 27959714]
26. Stonerock GL, Blumenthal JA. Role of Counseling to Promote Adherence in Healthy Lifestyle Medicine: Strategies to Improve Exercise Adherence and Enhance Physical Activity. Prog Cardiovasc Dis 2017;59(5):455-462 [FREE Full text] [doi: 10.1016/j.pcad.2016.09.003] [Medline: 27640186]
27. Johal S, Jamsen KM, Bell JS, Mc NKP, Magliano DJ, Liew D, et al. Do statin users adhere to a healthy diet and lifestyle? The Australian Diabetes, Obesity and Lifestyle Study. Eur J Prev Cardiol 2017 Dec;24(6):621-627. [doi: 10.1177/2047487316684054] [Medline: 28326830]
28. Grosso G, Bella F, Godos J, Sciacca S, Del RD, Ray S, et al. Possible role of diet in cancer: systematic review and multiple meta-analyses of dietary patterns, lifestyle factors, and cancer risk. Nutr Rev 2017 Dec 01;75(6):405-419. [doi: 10.1093/nutrit/nux012] [Medline: 28969358]
29. Stacey FG, Lubans DR, Chapman K, Bisquera A, James EL. Maintenance of Lifestyle Changes at 12-month Follow-up in a Nutrition and Physical Activity Trial for Cancer Survivors. Am J Health Behav 2017 Nov 01;41(6):784-795. [doi: 10.5993/AJHB.41.6.12] [Medline: 29025506]
30. Varleta P, Acevedo M, Akel C, Salinas C, Navarrete C, García A, et al. Mobile phone text messaging improves antihypertensive drug adherence in the community. J Clin Hypertens (Greenwich) 2017 Dec;19(12):1276-1284 [FREE Full text] [doi: 10.1111/jch.13098] [Medline: 28941056]
31. Hensher DA, Rose JM, Greene WH. Applied Choice Analysis. Cambridge: Cambridge University Press; 2015.
32. Brown MT, Bussell J, Dutta S, Davis K, Strong S, Mathew S. Medication Adherence: Truth and Consequences. Am J Med Sci 2016 Apr;351(4):387-399. [doi: 10.1016/j.amjms.2016.01.010] [Medline: 27079345]
33. Zhao D, Rao K, Zhang Z. Patient Trust in Physicians: Empirical Evidence from Shanghai, China. Chin Med J (Engl) 2016 Apr 05;129(7):814-818 [FREE Full text] [doi: 10.4103/0366-6999.178971] [Medline: 26996477]
34. McAllister DJ. AFFECT- AND COGNITION-BASED TRUST AS FOUNDATIONS FOR INTERPERSONAL COOPERATION IN ORGANIZATIONS. Academy of Management Journal 1995 Feb 01;38(1):24-59. [doi: 10.2307/256727]
35. Casimir G, Lee K, Loon M. Knowledge sharing: influences of trust, commitment and cost. J of Knowledge Management 2012 Sep 07;16(5):740-753. [doi: 10.1108/13673271211262781]
36. Choi BK, Moon HK, Nae EY. Cognition- and affect-based trust and feedback-seeking behavior: the roles of value, cost, and goal orientations. J Psychol 2014;148(5):603-620. [doi: 10.1080/00223980.2013.818928] [Medline: 25087321]
37. Massey GR, Dawes PL. Personal characteristics, trust, conflict, and effectiveness in marketing/sales working relationships. European Journal of Marketing 2007 Sep 25;41(9/10):1117-1145. [doi: 10.1108/03090560710773372]
38. Parayitam S, Dooley RS. The interplay between cognitive- and affective conflict and cognition- and affect-based trust in influencing decision outcomes. Journal of Business Research 2009 Aug;62(8):789-796. [doi: 10.1016/j.jbusres.2008.02.006]
39. Wang J, Zhang Z, Jia M. Understanding How Leader Humility Enhances Employee Creativity. The Journal of Applied Behavioral Science 2016 Nov 23;53(1):5-31. [doi: 10.1177/0021886316678907]
40. Barker RT, Camarata MR. The Role of Communication in Creating and Maintaining a Learning Organization: Preconditions, Indicators, and Disciplines. Journal of Business Communication 1998 Oct 01;35(4):443-467. [doi: 10.1177/002194369803500402]
41. Schwaer C, Biemann T, Voelpel S. Antecedents of employee's preference for knowledge-sharing tools. The International Journal of Human Resource Management 2012 Oct;23(17):3613-3635. [doi: 10.1080/09585192.2011.639552]
42. Kim Y. Trust in health information websites: A systematic literature review on the antecedents of trust. Health Informatics J 2016 Jun;22(2):355-369. [doi: 10.1177/1460458214559432] [Medline: 25518944]
43. Hsiung HH, Tsai WC. The Joint Moderating Effects of Activated Negative Moods and Group Voice Climate on the Relationship between Power Distance Orientation and Employee Voice Behavior. Applied Psychology 2017 Mar 07;66(3):487-514. [doi: 10.1111/apps.12096]
44. Young GJ, Meterko MM, Mohr D, Shwartz M, Lin H. Congruence in the assessment of service quality between employees and customers: A study of a public health care delivery system. Journal of Business Research 2009 Nov;62(11):1127-1135. [doi: 10.1016/j.jbusres.2008.08.004]
45. Boekhorst JA. The Role of Authentic Leadership in Fostering Workplace Inclusion: A Social Information Processing Perspective. Hum Resour Manage 2014 Nov 18;54(2):241-264. [doi: 10.1002/hrm.21669]
46. Blau PM. Exchange and Power in Social Life. New York: Wiley; 1964.
47. Foa UG, Foa EB. Societal structures of the mind. Oxford: Charles C Thomas; 1974.
48. Foa EB, Foa UG. Resource theory of social exchange. In: Contemporary Topics in Social Psychology. Morristown: General Learning Press; 1976:99-131.
49. Barbouni A, Nalmpanti M, Gennimata D, Theodoridis D, Merakou K. Beliefs and practices of Greek doctors in relation to patients' adherence to antihypertensive medication. J Hum Hypertens 2017 May;31(5):341-346. [doi: 10.1038/jhh.2016.84] [Medline: 28032629]
50. Pappas JM, Flaherty KE. The effect of trust on customer contact personnel strategic behavior and sales performance in a service environment. Journal of Business Research 2008 Sep;61(9):894-902. [doi: 10.1016/j.jbusres.2007.09.017]

XSL•FO
RenderX

51. Moorman C, Zaltman G, Deshpande R. Relationships between Providers and Users of Market Research: The Dynamics of Trust within and between Organizations. Journal of Marketing Research 1992 Aug;29(3):314. [doi: 10.2307/3172742]
52. Ha H, John J, John JD, Chung Y. Temporal effects of information from social networks on online behavior. Internet Research 2016 Feb;26(1):213-235. [doi: 10.1108/IntR-03-2014-0084]
53. Rutten W, Blaas-Franken J, Martin H. The impact of (low) trust on knowledge sharing. J of Knowledge Management 2016 Apr 04;20(2):199-214. [doi: 10.1108/JKM-10-2015-0391]
54. Emerson RM. Social exchange theory. In: Social Psychology: Sociological Perspectives. New York: Basic Books; 1981.
55. Singh AG, Singh S, Singh PP. YouTube for information on rheumatoid arthritis--a wakeup call? J Rheumatol 2012 May;39(5):899-903. [doi: 10.3899/jrheum.111114] [Medline: 22467934]
56. Wiertz C, de Ruyter K. Beyond the Call of Duty: Why Customers Contribute to Firm-hosted Commercial Online Communities. Organization Studies 2016 Jun 30;28(3):347-376. [doi: 10.1177/0170840607076003]
57. Hoque R, Sorwar G. Understanding factors influencing the adoption of mHealth by the elderly: An extension of the UTAUT model. Int J Med Inform 2017 May;101:75-84. [doi: 10.1016/j.ijmedinf.2017.02.002] [Medline: 28347450]
58. Gefen D, Straub DW, Boudreau MC. Structural equation modeling and regression: Guidelines for research practice. Communications of the Association for Information Systems 2000;4(7):1-77.
59. Buysse HEC, Coorevits P, Van MG, Hutse A, Kaufman J, Ruige J, et al. Introducing telemonitoring for diabetic patients: development of a telemonitoring 'Health Effect and Readiness' Questionnaire. Int J Med Inform 2010 Aug;79(8):576-584. [doi: 10.1016/j.ijmedinf.2010.05.005] [Medline: 20599161]
60. Sardeshmukh S, Vandenberg RJ. Integrating moderation and mediation: a structural equation modeling approach. AMPROC 2013 Jan 25;2013(1):11839-11839. [doi: 10.5465/AMBPP.2013.92]
61. Bu X, You L, Li Y, Liu K, Zheng J, Yan T, et al. Psychometric Properties of the Kessler 10 Scale in Chinese Parents of Children With Cancer. Cancer Nurs 2017;40(4):297-304. [doi: 10.1097/NCC.0000000000000392] [Medline: 27257800]
62. Xiao Y, Li T, Xiao L, Wang S, Wang S, Wang H, et al. The Chinese version of Instrument of Professional Attitude for Student Nurses (IPASN): Assessment of reliability and validity. Nurse Educ Today 2017 Feb;49:79-83. [doi: 10.1016/j.nedt.2016.11.013] [Medline: 27889582]
63. Wong WS, Chen PP, Chow YF, Wong S, Fielding R. A Study of the Reliability and Concurrent Validity of the Chinese Version of the Pain Medication Attitude Questionnaire (ChPMAQ) in a Sample of Chinese Patients with Chronic Pain. Pain Med 2016 Jun;17(6):1137-1144. [doi: 10.1093/pm/pnv086] [Medline: 26843593]
64. Wimble M. Understanding Health and Health-Related Behavior of Users of Internet Health Information. Telemed J E Health 2016 Oct;22(10):809-815. [doi: 10.1089/tmj.2015.0267] [Medline: 27045569]
65. Kaiser HF. A second generation little jiffy. Psychometrika 1970 Dec;35(4):401-415. [doi: 10.1007/BF02291817]
66. Kurtuldu MK, Bulut D. Development of a Self-Efficacy Scale toward Piano Lessons. Educational Sciences: Theory and Practice 2017;17(3):835-857. [doi: 10.12738/estp.2017.3.0209]
67. Biasutti M, Frate S. A validity and reliability study of the Attitudes toward Sustainable Development scale. Environmental Education Research 2016 Feb 19;23(2):214-230. [doi: 10.1080/13504622.2016.1146660]
68. Erdogan M, Ok A, Marcinkowski TJ. Development and validation of Children's Responsible Environmental Behavior Scale. Environmental Education Research 2012 Aug;18(4):507-540. [doi: 10.1080/13504622.2011.627421]
69. Wu W, Tang F, Dong X, Liu C. Different identifications cause different types of voice: A role identity approach to the relations between organizational socialization and voice. Asia Pac J Manag 2014 Jun 21;32(1):251-287. [doi: 10.1007/s10490-014-9384-x]
70. Nuño-Solinís R, Berraondo ZI, Sauto AR, San MRL, Toro PN. Development of a questionnaire to assess interprofessional collaboration between two different care levels. Int J Integr Care 2013 Apr;13:e015 [FREE Full text] [Medline: 23882165]
71. Mu GM, Hu Y. Validation of the Chinese Version of the 12-Item Child and Youth Resilience Measure. Children and Youth Services Review 2016 Nov;70:332-339. [doi: 10.1016/j.childyouth.2016.09.037]
72. Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin 1980;88(3):588-606. [doi: 10.1037/0033-2909.88.3.588]
73. Steiger JH, Lind J. Statistically based tests for the number of common factors. 1980 Presented at: the annual meeting of the Psychometric Society; 1980; Iowa City.
74. Bentler PM, Kano Y. On the Equivalence of Factors and Components. Multivariate Behav Res 1990 Jan 01;25(1):67-74. [doi: 10.1207/s15327906mbr2501_8] [Medline: 26741971]
75. Cohen J. Statistical power analysis for the behavioral sciences. Journal of the American Statistical Association 1988;2(4):19-74. [doi: 10.1016/B978-0-12-179060-8.50012-8]
76. Lee Y, Lin JL. The effects of trust in physician on self-efficacy, adherence and diabetes outcomes. Soc Sci Med 2009 Mar;68(6):1060-1068. [doi: 10.1016/j.socscimed.2008.12.033] [Medline: 19162386]

Abbreviations

ABT: affect-based trust
CBT: cognition-based trust

XSL•FO
RenderX

SEM: structural equation modeling

Edited by G Eysenbach; submitted 08.11.17; peer-reviewed by PCI Pang, Q Yang, L Sbaffi; comments to author 15.03.18; revised version received 10.05.18; accepted 21.06.18; published 17.08.18.

Please cite as:
Lu X, Zhang R, Wu W, Shang X, Liu M
Relationship Between Internet Health Information and Patient Compliance Based on Trust: Empirical Study
J Med Internet Res 2018;20(8):e253
URL: http://www.jmir.org/2018/8/e253/
doi:10.2196/jmir.9364
PMID:30120087

©Xinyi Lu, Runtong Zhang, Wen Wu, Xiaopu Shang, Manlu Liu. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 17.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Cloud Computing for Infectious Disease Surveillance and Control: Development and Evaluation of a Hospital Automated Laboratory Reporting System

Mei-Hua Wang[1*], RN, MS; Han-Kun Chen[2*], MD; Min-Huei Hsu[1], MD, PhD; Hui-Chi Wang[1], MS; Yu-Ting Yeh[1,3], RN, PhD

[1]Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

[2]Department of General Surgery, Chi-Mei Medical Center, Tainan, Taiwan

[3]Information Technology Office, Shuang Ho Hospital, Taipei Medical University, New Taipei City, Taiwan

[*]these authors contributed equally

Corresponding Author:
Yu-Ting Yeh, RN, PhD
Graduate Institute of Biomedical Informatics
College of Medical Science and Technology
Taipei Medical University
250 Wuxing Street
Taipei,
Taiwan
Phone: 886 22490088 ext 8901
Email: yuting@tmu.edu.tw

Abstract

Background: Outbreaks of several serious infectious diseases have occurred in recent years. In response, to mitigate public health risks, countries worldwide have dedicated efforts to establish an information system for effective disease monitoring, risk assessment, and early warning management for international disease outbreaks. A cloud computing framework can effectively provide the required hardware resources and information access and exchange to conveniently connect information related to infectious diseases and develop a cross-system surveillance and control system for infectious diseases.

Objective: The objective of our study was to develop a Hospital Automated Laboratory Reporting (HALR) system based on such a framework and evaluate its effectiveness.

Methods: We collected data for 6 months and analyzed the cases reported within this period by the HALR and the Web-based Notifiable Disease Reporting (WebNDR) systems. Furthermore, system evaluation indicators were gathered, including those evaluating sensitivity and specificity.

Results: The HALR system reported 15 pathogens and 5174 cases, and the WebNDR system reported 34 cases. In a comparison of the two systems, sensitivity was 100% and specificity varied according to the reported pathogens. In particular, the specificity for *Streptococcus pneumoniae*, *Mycobacterium tuberculosis* complex, and hepatitis C virus were 99.8%, 96.6%, and 97.4%, respectively. However, the specificity for influenza virus and hepatitis B virus were only 79.9% and 47.1%, respectively. After the reported data were integrated with patients' diagnostic results in their electronic medical records (EMRs), the specificity for influenza virus and hepatitis B virus increased to 89.2% and 99.1%, respectively.

Conclusions: The HALR system can provide early reporting of specified pathogens according to test results, allowing for early detection of outbreaks and providing trends in infectious disease data. The results of this study show that the sensitivity and specificity of early disease detection can be increased by integrating the reported data in the HALR system with the cases' clinical information (eg, diagnostic results) in EMRs, thereby enhancing the control and prevention of infectious diseases.

(J Med Internet Res 2018;20(8):e10886) doi:10.2196/10886

KEYWORDS

laboratory autoreporting system; HALR; electronic medical records

XSL•FO
RenderX

Introduction

Electronic laboratory reporting (ELR) generally refers to the automated transmission of reportable laboratory findings from public health services, hospitals, and other laboratories to local or state public health agencies. Many communicable diseases currently under nationwide surveillance can be identified and confirmed by laboratory test results. Thus, ELR has become a critical part of the disease surveillance process. Previous studies have shown that ELR improves the timeliness, accuracy, and completeness of reported laboratory data, which, in turn, improves the effectiveness and efficiency of public health responses to outbreaks and cases of notifiable conditions [1-4]. It has been included as a meaningful use objective by the US Centers for Medicare and Medicaid Services' Electronic Health Records Incentive programs [5,6]. However, ELR often lacks the clinical information needed to satisfy a case definition, such as disease signs, symptoms, and diagnoses. Moreover, ELR is usually nonspecific, particularly in cases (such as acute hepatitis B) where diagnoses require integration of laboratory tests and clinical information [7-10]. The lack of specificity in ELR increases the workload for health departments compelled to investigate suggestive but nonspecific lab results [11].

An electronic medical record (EMR) is a systematized collection of computerized patient and population health information. An EMR supports secure, real-time, point-of-care, patient-centric information availability and is a resource for clinical care as well as research and education. Recently, EMR systems have become an increasingly pervasive technology in health care settings [12-14]. Several studies have shown that EMR systems can accelerate clinical information flow, facilitate health care data integration, and improve the efficiency and quality of medical services [15-17]. With the growing adoption of EMR systems in recent years, increasingly sophisticated data have become available in EMRs to support infection surveillance, prevention, and control [18-20]. Using EMR data for the detection and reporting of infectious diseases also has the potential to improve public health monitoring and reporting [21]. However, current EMR systems are primarily built to serve clinical practice and are not structured for public health use. Laboratory test data transmitted to EMR systems might not be as complete and as timely as data received directly from a laboratory information system (LIS) [22].

In Taiwan, as required by law [23], clinical laboratory units operated by hospitals, health agencies, and research institutes must report the outcomes of tests for pathogens, mostly bacteria and viruses, that meet the notification conditions of Taiwan's Centers of Disease Control (TCDC) in order to enable epidemiological surveillance and advanced alerting of communicable diseases. Thus, the TCDC has developed a Web-based Notifiable Disease Reporting (WebNDR) system to strengthen infectious disease control and surveillance. The system requires surveillance professionals to enter detailed information of notifiable diseases and laboratory tests manually if the diseases meet the TCDC's reporting definitions. Unfortunately, these reporting operations are time consuming and error prone [24]. This is also the case for unannounced or delayed notifications [25].

In order to reduce the workload of surveillance professionals, and address the lack of specificity of ELR, the TCDC launched a pilot project, called Automated Laboratory Reporting (ALR), in 2014 [26]. The ALR system enables a hospital to automatically transmit reportable cases to the TCDC when they meet TCDC's notifiable conditions. Reportable cases include not only detailed laboratory test data but also relevant EMR data. Currently, hospitals can join the project on a voluntary basis. The TCDC offers incentives to help develop a counterpart to the TCDC's ALR system in hospital settings. Although including EMR data into laboratory-reportable cases can improve the specificity of ELR, few studies have evaluated such systems. In this study, we developed a counterpart to the TCDC's ALR system for a hospital setting and evaluated its effectiveness in terms of sensitivity and specificity compared with the existing WebNDR system.

Methods

Settings

In this study, we implemented a counterpart of the TCDC's ALR system in a teaching affiliate hospital with about 1000 beds starting in August 2014 in Taiwan. The hospital has an LIS to manage laboratory orders and results, a computerized physician order entry (CPOE) system to support clinicians in their daily clinical practice including entering test orders and reviewing test results, and other EMR systems. It can also access the WebNDR system to report notifiable diseases manually.

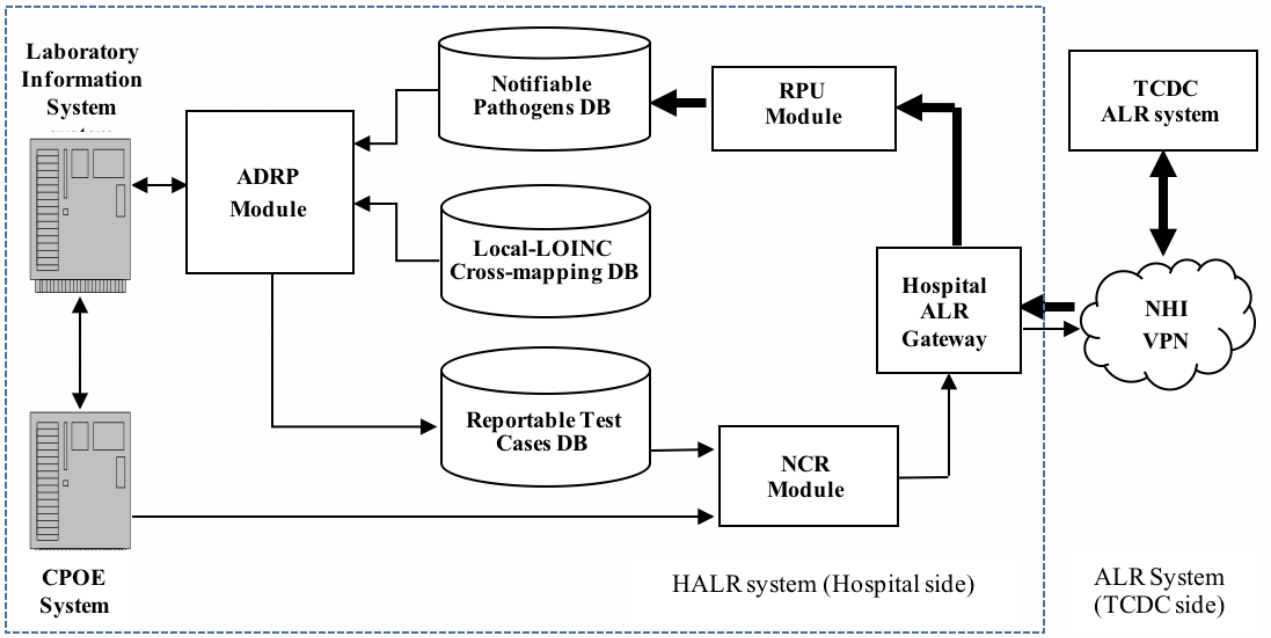
The Framework of Automated Laboratory Reporting System in a Hospital Setting

Based on the framework of the TCDC's ALR system, this study documents the implementation in a hospital setting, termed the Hospital Automated Laboratory Reporting (HALR) system (Figure 1). The HALR system consists of four major modular components: HALR Gateway, Reportable Pathogens Update (RPU), Automated Detection of Reportable Pathogens (ADRP), and Notifiable Case Reporting (NCR). The HALR Gateway is responsible for downloading the latest notifiable disease pathogens from the TCDC's ALR system and uploading reportable cases to the TCDC ALR system for further processing. Using the RPU module, the downloaded pathogens are then used to update the Notifiable Pathogen Database (NPDB) in the HALR system. This belongs to the out-of-hospital system data and is indicated by a thick line.

When the LIS releases a pathogen test result, the ADRP module checks the pathogen against the NPDB. LIS pathogen test results include patient ID, name, gender, test item, specimen type, and pathogen report content. If the pathogen can be found in the NPDB and its test result is positive, the test case is labeled as reportable. The patient's information is then included with the pathogen test result to form a reportable case. The laboratory test is encoded with a hospital code; however, the TCDC requires a laboratory test with a Logical Observation Identifiers Names and Codes (LOINC) code.

XSL•FO
RenderX

Figure 1. The framework of a Hospital Automated Laboratory Reporting (HALR) System. (The thick "-" corresponds to the out-of-hospital system; the thin "-" corresponds to the in-hospital system). ADRP: Automated Detection of Reportable Pathogens, ALR: Automated Laboratory Reporting, CPOE: computerized physician order entry, DB: database, LOINC: Logical Observation Identifiers Names and Codes, NCR: Notifiable Case Reporting, NHI: National Health Insurance, RPU: Reportable Pathogens Update, TCDC: Taiwan's Centers of Disease Control, VPN: virtual private network.



Thus, the ADRP module can translate the hospital code of a pathogen test into its corresponding LOINC code by referencing a local LOINC cross-mapping database. The reportable test cases detected by the ADRP module are then stored into the Reportable Test Case Database (RTCDB).

Finally, the NCR module retrieves a reportable test case from the RTCDB and links the test data with the patient's clinical information, such as disease name, diagnosis code, and the date when the condition was diagnosed, which are stored in the CPOE system. Then, the linked data are compiled into a laboratory-reportable case in a digital format such as XML that is defined by the TCDC. Subsequently, the HALR Gateway sends the laboratory-reportable case to the TCDC ALR system; this belongs to the in-hospital system data and is indicated by a thin line.

Mapping Between Pathogens and Diseases

When the LIS confirms a pathogen test result that meets a laboratory reporting condition, the pathogen test result and other information, such as order number, local code, specimen type, and time of result, will be automatically written into the RTCDB.

After retrieving new reporting data, the program converts the laboratory test items to LOINC codes and retrieves other patient information required by the TCDC from different hospital information systems and then submits the combined data to the hospital reporting module within the TCDC gateway in the required format.

Since the WebNDR system is dedicated to reporting communicable diseases, the five disease categories reported by this system are used as a basis of evaluating the HALR system. The TCDC provides a table for mapping between pathogens and their related diseases (Table 1).

System Evaluation

After the HALR system was implemented in a regional hospital in Taipei, we collected reported cases from the HALR and WebNDR systems for 6 months, from December 2014 to May 2015. Since the reported cases from the WebNDR system were confirmed by infectious control professionals, they served as a gold standard for evaluating the HALR system in this study. Thus, the sensitivity and specificity of the reported cases by the HALR system could be evaluated according to the following definitions:

Sensitivity (true positive rate): If a patient has a notifiable disease reported by the WebNDR system, the sensitivity is the probability that the HALR system would report the case. The numerator is the number of cases identified as positive and reported by both the HALR and WebNDR systems, whereas denominator is the total number of cases reported by the WebNDR system (# of reported cases by both the HALR and WebNDR systems)/(# of cases reported by the WebNDR system).

Specificity (true negative rate): If a patient does not have a notifiable disease and is not reported by the WebNDR system, the specificity is the probability that the HALR system would not report the case. The numerator is the number of cases in whom test results were identified as negative and were not reported by the WebNDR system, whereas the denominator is the total number of cases not reported by the WebNDR system (# of not reported cases by neither the HALR nor WebNDR system)/(# of cases not reported by the WebNDR).

XSL•FO
RenderX

Table 1. Mapping between pathogens and diseases (based on the Web-based Notifiable Disease Reporting system).

Pathogen	ICD-9-CM[a]	Disease
Streptococcus pneumoniae	481, 482, 485, 486, 038, 041, 320	Invasive pneumococcal disease
Mycobacterium tuberculosis complex	010-018	Tuberculosis
Influenza virus	487	Severe complicated influenza case
Hepatitis B virus	070.20,070.21,070.30, 070.31	Acute hepatitis B
Hepatitis C virus	070.41	Acute hepatitis C

[a]ICD-9-CM: The International Classification of Diseases, Ninth Revision, Clinical Modification.

Table 2. The predictive specificity and sensitivity values of tests evaluated for the Hospital Automated Laboratory Reporting (HALR) system (pathogen test result) compared with those for the Web-based Notifiable Disease Reporting (WebNDR) system (reported result).

Method	Reported by the WebNDR system	Not reported by the WebNDR system	Total
Reported (positive) by the HALR system	TP[a]	(FP[b])	TP+(FP)
Not reported (negative) by the HALR system	(FN[c])	TN[d]	(FN)+TN
Total	TP+(FN)	(FP)+TN	—

[a]TP: true positive.

[b]FP: false positive.

[c]FN: false negative.

[d]TN: true negative.

The cases reported by the WebNDR system were patients with notifiable diseases, whereas those reported by the HALR system were patients whose pathogen test results were positive. During his or her hospital stay, a patient with the same pathogen reported more than once (ie, a repeatedly reported case) was excluded.

Furthermore, since a pathogen may cause more than one disease, the HALR system includes clinical diagnosis information such as that from the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). Thus, the test results were combined with ICD-9-CM data to determine whether the HALR system performance could be improved.

- Sensitivity (with ICD): The numerator is the number of cases identified as a notifiable disease with a positive test result, which were automatically reported by the HALR and WebNDR systems. The denominator is the total number of cases reported by the WebNDR system.
- Specificity (with ICD): The numerator is the number of cases identified as negative and, therefore, not reported by either the HALR or WebNDR system. The denominator is the total number of cases not reported by the WebNDR system (Table 2).

Results

Pathogens Analysis

Based on the TCDC's reportable pathogens and reporting criteria, 15 pathogens were reported to the TCDC ALR system (Table 3). Among these, 5174 patients had positive test results and were reported by the HALR system.

Simultaneously, 34 cases in five disease categories were reported by the WebNDR system (Table 4). The number of reported cases from different disease categories was pulmonary tuberculosis (25), complications from severe influenza (3) and acute viral hepatitis C (3), invasive *Streptococcus pneumoniae* infection (2), and acute viral hepatitis B (1).

Sensitivity and Specificity Analysis

Table 5 shows the sensitivity and specificity analysis for the HALR system. In this study, the sensitivity for the HALR system reached 100%, but the specificity varied according to the pathogens. The specificity for *S. pneumoniae*, *Mycobacterium tuberculosis* complex, and hepatitis C virus were 99.8%, 96.6%, and 97.4%, respectively. In addition, the specificity for influenza and hepatitis B virus were only 79.9% and 47.1%, respectively.

Because the HALR system collects a reported case that includes not only laboratory test results but also clinical information (such as ICD-9-CM) from the TCDC ALR system, only cases with associated laboratory positive test results and an associated diagnosis code that meets the TCDC's notifiable disease definition are reported to the TCDC ALR system (Table 6). The sensitivity and specificity of cases reported by the HALR system can be recalculated (Table 6). The sensitivity performance remains unchanged, but the specificity is greatly improved, in particular for influenza (89.2%) and hepatitis B virus (99.1%). Thus, the inclusion of clinical information in the reported data can improve the specificity performance.

XSL•FO
RenderX

Table 3. The analysis of laboratory test results during the study period.

Name of the detected pathogen	Total number of subjects (N=57,511), N (%)	Pathogen test results, n (%)	
		Negative (n=52,337)	Positive (n=5174)
Notifiable diseases (law requirements)			
Salmonella spp	10,201 (17.8)	10,173 (19.4)	28 (0.5)
Streptococcus pneumoniae	9970 (17.3)	9946 (19.0)	24 (0.5)
Mycobacterium tuberculosis complex	1622 (2.8)	1542 (3.0)	80 (1.5)
Influenza virus	1980 (3.4)	1579 (3.0)	401 (7.8)
Enterovirus	N/A[a]	N/A	N/A
Hepatitis B virus	7917 (13.8)	3725 (7.1)	4192 (81)
Hepatitis C virus	5124 (8.9)	4986 (9.5)	138 (2.7)
Nonnotifiable diseases			
Streptococcus agalactiae, GBS[b]	10,576 (18.4)	10,356 (19.8)	220 (4.2)
Streptococcus pyogenes	9958 (17.3)	9898 (18.9)	60 (1.2)
Parainfluenza virus	N/A	N/A	N/A
Respiratory syncytial virus	90 (0.2)	80 (0.2)	10 (0.2)
Rotavirus	73 (0.1)	52 (0.1)	21 (0.4)
Yersinia enterocolitica	N/A	N/A	N/A
Campylobacter spp	N/A	N/A	N/A
Listeria monocytogenes	N/A	N/A	N/A

[a]N/A: not applicable.

[b]GBS: Group B Streptococcus.

Table 4. The analysis of the Web-based Notifiable Disease Reporting system during the study period.

Name of the detected pathogen	Total number of reported cases, n (%)
Streptococcus pneumoniae	2 (6)
Mycobacterium tuberculosis complex	25 (73)
Influenza virus	3 (9)
Hepatitis B virus	1 (3)
Hepatitis C virus	3 (9)

XSL•FO
RenderX

Table 5. The sensitivity and specificity of the Hospital Automated Laboratory Reporting (HALR) system.

Pathogen name and HALR system test result	WebNDR[a] system		Total	Sensitivity or specificity, %
	Reported, n (%)	Not reported, n (%)		
Streptococcus pneumoniae				
Positive (reported)	2 (100)	22 (0.2)	24	100[b]
Negative (not reported)	0 (0)	9946 (99.8)	9946	99.8[c]
Total	2 (100)	9968 (100)	9970	—
***Mycobacterium tuberculosis* complex**				
Positive (reported)	25 (100)	55 (3.4)	80	100[b]
Negative (not reported)	0 (0)	1542 (96.6)	1542	96.6[c]
Total	25 (100)	1597 (100)	1622	—
Influenza virus				
Positive (reported)	3 (100)	398 (20.1)	401	100[b]
Negative (not reported)	0 (0)	1579 (79.9)	1579	79.9[c]
Total	3 (100)	1977 (100)	1980	—
Hepatitis B virus				
Positive (reported)	1 (100)	4191 (52.9)	4192	100[b]
Negative (not reported)	0 (0)	3725 (47.1)	3725	47.1[c]
Total	1 (100)	7916 (100)	7917	—
Hepatitis C virus				
Positive (reported)	3 (100)	135 (2.6)	138	100[b]
Negative (not reported)	0 (0)	4986 (97.4)	4986	97.4[c]
Total	3 (100)	5121 (100)	5124	—

[a]WebNDR: Web-based Notifiable Disease Reporting.

[b]Refers to sensitivity.

[c]Refers to specificity.

XSL•FO
RenderX

Table 6. The analysis of sensitivity and specificity between the Hospital Automated Laboratory Reporting (HALR) and Web-based Notifiable Disease Reporting (WebNDR) systems with the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes.

Pathogens name (ICD-9-CM code) and result	WebNDR system		Total	Sensitivity or specificity (%)
	Reported, n (%)	Not reported, n (%)		
***Streptococcus pneumoniae* (ICD-9-CM: 485, 486)**				
Positive (reported)	2 (100)	12 (0.1)	14	100[a]
Negative (not reported)	0 (0)	9956 (99.9)	9956	99.9[b]
Total	2 (100)	9968 (100)	9970	—
***Mycobacterium tuberculosis* complex (ICD-9-CM: 011.00, 011.90)**				
Positive (reported)	25 (100)	12 (0.8)	37	100[a]
Negative (not reported)	0 (0)	1585 (99.2)	1585	99.2[b]
Total	25 (100)	1597 (100)	1622	—
Influenza virus (ICD-9-CM: 487.1)				
Positive (reported)	3 (100)	214 (10.8)	217	100[a]
Negative (not reported)	0 (0)	1763 (89.2)	1763	89.2[b]
Total	3 (100)	1977 (100)	1980	—
Hepatitis B virus (ICD-9-CM: 070.30)				
Positive (reported)	1 (100)	74 (0.9)	75	100[a]
Negative (not reported)	0 (0)	7842 (99.1)	7842	99.1[b]
Total	1 (100)	7916 (100)	7917	—
Hepatitis C virus (ICD-9-CM: 070.41, 070.51)				
Positive (reported)	3 (100)	9 (0.2)	12	100[a]
Negative (not reported)	0 (0)	5112 (99.8)	5112	99.8[b]
Total	3 (100)	5121 (100)	5124	—

[a]Refers to sensitivity.

[b]Refers to specificity.

Discussion

Principal Findings

In this study, we developed an HALR system for the automatic reporting of pathogen test results and clinical information. Once laboratory test results are released, the HALR system can automatically detect pathogens that meet notifiable conditions as defined by the TCDC and report the cases to the TCDC ALR system. Since the patients' laboratory test results are usually released to their physicians far in advance of the physicians' final diagnoses of any notifiable diseases, the HALR system can improve the timeliness for notifiable disease surveillance and control. Moreover, this system's effectiveness is also improved as long as the doctor has included a working diagnosis.

As indicated in previous studies [7-10], an ELR is usually nonspecific. The analysis reported in this study showed that if reported data only included the results of laboratory tests, the specificity of the reported cases by the HALR system for some diseases would be quite low. This could lead to an increase in reported cases as well as increase in workload related to the investigation of suggestive cases. However, if reported data were slightly augmented with clinical information, such as the clinical diagnosis code, the specificity of the reported cases could be greatly improved. The primary reason is that a given pathogen may cause several different diseases, which are not notifiable. For example, a hepatitis B virus infection may lead to acute or chronic hepatitis, but only acute hepatitis B (with ICD-9-CM diagnosis code 070.30) must be reported. Similarly, influenza viruses may cause acute respiratory infections, from mild to severe, but only cases with severe complications from influenza (ICD-9-CM 487.1) must be reported. However, because influenza is caused by a variety of pathogens and easily causes pandemics, the National Respiratory and Enteric Virus Surveillance System of the World Health Organization provides influenza-like illnesses with clear definitions and clinical symptom identification standards, which must be updated and revised according to emerging viruses and changes in patients' disease features. The HALR system reports not only laboratory test results but also relevant clinical information and, thus, can greatly improve its performance.

In response to emerging infectious diseases, the HALR system detects reportable cases based on the NPDB. This design allows

XSL•FO
RenderX

the TCDC to add, delete, or update the definitions of reportable pathogens as new emerging threats are identified. Hence, as this system can routinely download and update the NPDB, it can quickly respond to altered requirements.

The improvement of the sensitivity and specificity of reportable cases depends not only on timely and accurate laboratory test results but also on the availability of clinical information required to identify cases. This study simultaneously examined reporting data on the basis of test results and clinical evidence in order to enable data sharing between information systems. Through double verification, with this approach, hospitals can immediately confirm patients' conditions and submit reports when required. Thus, the sensitivity and specificity for early disease detection can be improved. The findings of this study can serve as a reference for disease prevention measures in clinical care.

Limitations

The following are some limitations of this system. First, the system will be limited to tests accepted by patients in the hospital. This system can only automatically link information when required clinical data are complete. Second, the notification condition of the test and the positive definition of the pathogen remain important factors affecting the accuracy of the content of the notification.

Conclusions and Future Directions

In this study, we developed an HALR system in a hospital to automatically and actively report pathogen test results and relevant clinical information to the TCDC ALR system. The study results show that the HALR system can improve the timeliness, sensitivity, and specificity of reported cases. Furthermore, it can provide the flexibility to integrate frequent changes to the definitions of notifiable cases when the TCDC finds new, emerging threats or diseases.

The improvement of the sensitivity and specificity of reportable cases depends not only on timely and accurate laboratory test results but also on the availability of clinical information that may be required for identified cases. This study recommends the following feasible solutions: integrate automatic pathogen reporting by the HALR system with related medical data in patients' EMRs through the automatic system determination, with EMR data provided according to the TCDC clinical standards specifications for infectious diseases (eg, according to a "Dengue Fever" diagnosis, clinical data can include a patient's fever [>38°C] and at least one of the following symptoms: retro-orbital pain, myalgia, arthralgia, rash, leukopenia, and hemorrhagic manifestations.)

Conflicts of Interest

None declared.

References

1. Vogt RL. Laboratory reporting and disease surveillance. J Public Health Manag Pract 1996;2(4):28-30. [Medline: 10186690]
2. Wurtz R, Cameron BJ. Electronic laboratory reporting for the infectious diseases physician and clinical microbiologist. Clin Infect Dis 2005 Jun 01;40(11):1638-1643. [doi: 10.1086/429904] [Medline: 15889362]
3. Overhage JM, Grannis S, McDonald CJ. A comparison of the completeness and timeliness of automated electronic laboratory reporting and spontaneous reporting of notifiable conditions. Am J Public Health 2008 Feb;98(2):344-350. [doi: 10.2105/AJPH.2006.092700] [Medline: 18172157]
4. Backer HD, Bissell SR, Vugia DJ. Disease reporting from an automated laboratory-based reporting system to a state health department via local county health departments. Public Health Rep 2001;116(3):257-265 [FREE Full text] [doi: 10.1093/phr/116.3.257] [Medline: 12034915]
5. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med 2010 Aug 05;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]
6. Centers for Medicare & Medicaid Services. US Government Publishing Office. Medicare and Medicaid Programs; Electronic Health Record Incentive Program URL: https://www.gpo.gov/fdsys/pkg/FR-2012-09-04/pdf/2012-21050.pdf [accessed 2018-06-26] [WebCite Cache ID 70SSahLPs]
7. Panackal AA, M'ikanatha NM, Tsui F, McMahon J, Wagner MM, Dixon BW, et al. Automatic electronic laboratory-based reporting of notifiable infectious diseases at a large health system. Emerg Infect Dis 2002 Jul;8(7):685-691 [FREE Full text] [doi: 10.3201/eid0807.010493] [Medline: 12095435]
8. M'ikantha NM, Southwell B, Lautenbach E. Automated laboratory reporting of infectious diseases in a climate of bioterrorism. Emerg Infect Dis 2003 Sep;9(9):1053-1057 [FREE Full text] [doi: 10.3201/eid0909.020486] [Medline: 14519239]
9. Klompas M, Platt R, Hou X, Campion FX, Kruskal B, Haney G, et al. MMWR Weekly. 2008 Apr 11. Automated Detection and Reporting of Notifiable Diseases Using Electronic Medical Records Versus Passive Surveillance—Massachusetts, June 2006-July 2007 URL: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5714a4.htm [accessed 2018-06-26] [WebCite Cache ID 70STTKxkC]
10. Allen-Dicker J, Klompas M. Comparison of electronic laboratory reports, administrative claims, and electronic health record data for acute viral hepatitis surveillance. J Public Health Manag Pract 2012;18(3):209-214. [doi: 10.1097/PHH.0b013e31821f2d73] [Medline: 22473112]

XSL•FO
RenderX

11. McHugh LA, Semple S, Sorhage FE, Tan CG, Langer AJ. MMWR Weekly. 2008 Jan 8. Effect of Electronic Laboratory Reporting on the Burden of Lyme Disease Surveillance --- New Jersey, 2001--2006 URL: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5702a4.htm [accessed 2018-06-26] [WebCite Cache ID 70STvwr9q]
12. Shiffman RA, Brandt CA, Liaw Y, Corb GJ. A design model for computer-based guideline implementation based on information management services. J Am Med Inform Assoc 1999;6(2):99-103 [FREE Full text] [Medline: 10094062]
13. Hillestad R, Bigelow J, Bower A, Girosi F, Meili R, Scoville R, et al. Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. Health Aff (Millwood) 2005;24(5):1103-1117. [doi: 10.1377/hlthaff.24.5.1103] [Medline: 16162551]
14. Simborg DW. Promoting electronic health record adoption. Is it the correct focus? J Am Med Inform Assoc 2008 Mar;15(2):127-129 [FREE Full text] [doi: 10.1197/jamia.M2573] [Medline: 18096904]
15. O'Connor PJ, Crain AL, Rush WA, Sperl-Hillen JM, Gutenkauf JJ, Duncan JE. Impact of an electronic medical record on diabetes quality of care. Ann Fam Med 2005 Jul;3(4):300-306 [FREE Full text] [doi: 10.1370/afm.327] [Medline: 16046561]
16. Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. Int J Med Inform 2008 May;77(5):291-304. [doi: 10.1016/j.ijmedinf.2007.09.001] [Medline: 17951106]
17. Atreja A, Gordon SM, Pollock DA, Olmsted RN, Brennan PJ, Healthcare Infection Control Practices Advisory Committee. Opportunities and challenges in utilizing electronic health records for infection surveillance, prevention, and control. Am J Infect Control 2008 Apr;36(3 Suppl):S37-S46. [doi: 10.1016/j.ajic.2008.01.002] [Medline: 18374211]
18. Wright MO, Fisher A, John M, Reynolds K, Peterson LR, Robicsek A. The electronic medical record as a tool for infection surveillance: successful automation of device-days. Am J Infect Control 2009 Jun;37(5):364-370. [doi: 10.1016/j.ajic.2008.11.003] [Medline: 19269712]
19. Pokorny L, Rovira A, Martín-Baranera M, Gimeno C, Alonso-Tarrés C, Vilarasau J. Automatic detection of patients with nosocomial infection by a computer-based surveillance system: a validation study in a general hospital. Infect Control Hosp Epidemiol 2006 May;27(5):500-503. [doi: 10.1086/502685] [Medline: 16671032]
20. Lo Y, Lee W, Chen G, Liu C. Improving the work efficiency of healthcare-associated infection surveillance using electronic medical records. Comput Methods Programs Biomed 2014 Nov;117(2):351-359. [doi: 10.1016/j.cmpb.2014.07.006] [Medline: 25154644]
21. Calderwood MS, Platt R, Hou X, Malenfant J, Haney G, Kruskal B, et al. Real-time surveillance for tuberculosis using electronic health record data from an ambulatory practice in eastern Massachusetts. Public Health Rep 2010;125(6):843-850 [FREE Full text] [doi: 10.1177/003335491012500611] [Medline: 21121229]
22. Samoff E, Fangman MT, Fleischauer AT, Waller AE, Macdonald PDM. Improvements in timeliness resulting from implementation of electronic laboratory reporting and an electronic disease surveillance system. Public Health Rep 2013;128(5):393-398 [FREE Full text] [doi: 10.1177/003335491312800510] [Medline: 23997286]
23. Laws & Regulations Database of The Republic of China. 2016 Jul 07. Regulations Governing the Implementation of the Epidemiological Surveillance and Advance-Alert System for Communicable Disease URL: http://law.moj.gov.tw/Eng/LawClass/LawContent.aspx?PCODE=L0050006 [accessed 2018-06-26] [WebCite Cache ID 70SUz3QNQ]
24. Yuping H, Jialin L, Lun L, Qiming Z, Xiang Z, Jen-Hsiang C. Reporting Notifiable Infectious Diseases by Using Hospital Electronic Medical Record Systems: An Effectiveness Assessment. The Journal of Taiwan Association for Medical Informatics 2016;25(1):23-31 [FREE Full text]
25. Lo H, Yang S, Chou P, Chuang J, Chiang C. Completeness and timeliness of tuberculosis notification in Taiwan. BMC Public Health 2011 Dec 12;11:915 [FREE Full text] [doi: 10.1186/1471-2458-11-915] [Medline: 22151346]
26. Taiwan's CODC. Operation Manual for the Automatic Laboratory Reporting System and a Platform for Laboratory Information Exchange across Hospitals (V3.1) (in Chinese) URL: http://www.cdc.gov.tw/professional/info.aspx?treeid=4c19a0252bbef869&nowtreeid=1a8c498af5f8af5d&tid=763A28C245DF7CC6

Abbreviations

ADRP: Automated Detection of Reportable Pathogens
ALR: Automated Laboratory Reporting
CPOE: computerized physician order entry
ELR: electronic laboratory reporting
EMR: electronic medical record
FN: false negative
FP: false positive
GBS: Group B Streptococcus
HALR: Hospital Infectious Disease Laboratory Autoreporting
ICD-9-CM: The International Classification of Diseases, Ninth Revision, Clinical Modification
LIS: Laboratory Information System
LOINC: Logical Observation Identifiers Names and Codes

XSL•FO
RenderX

NCR: Notifiable Case Reporting
NPDB: Notifiable Pathogen Database
RPU: Reportable Pathogens Update
RTCDB: Reportable Test Case Database
TCDC: Taiwan's Centers of Disease Control
TN: true negative
TP: true positive
WebNDR: Web-based Notifiable Disease Reporting

Edited by G Eysenbach; submitted 02.05.18; peer-reviewed by LC Yu, IC Lin, MIK Almani, C Fincham; comments to author 31.05.18; revised version received 04.06.18; accepted 19.06.18; published 08.08.18.

Please cite as:
Wang MH, Chen HK, Hsu MH, Wang HC, Yeh YT
Cloud Computing for Infectious Disease Surveillance and Control: Development and Evaluation of a Hospital Automated Laboratory Reporting System
J Med Internet Res 2018;20(8):e10886
URL: http://www.jmir.org/2018/8/e10886/
doi:10.2196/10886
PMID:30089608

©Mei-Hua Wang, Han-Kun Chen, Min-Huei Hsu, Hui-Chi Wang, Yu-Ting Yeh. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 08.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Forecasting the Maturation of Electronic Health Record Functions Among US Hospitals: Retrospective Analysis and Predictive Model

Hadi Kharrazi[1*], MHI, MD, PhD; Claudia P Gonzalez[2*], MBA; Kevin B Lowe[3*], PhD; Timothy R Huerta[4,5*], MS, PhD; Eric W Ford[6*], PhD, MPH

[1]Center for Population Health IT, Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[2]Strategic Management Program, Foster School of Business, University of Washington, Seattle, WA, United States

[3]The University of Sydney Business School, Sydney, Australia

[4]Department of Family Medicine, College of Medicine, The Ohio State University, Columbus, OH, United States

[5]Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, OH, United States

[6]Department of Health Care Organization and Policy, School of Public Health, University of Alabama Birmingham, Birmingham, AL, United States

[*]all authors contributed equally

Corresponding Author:
Hadi Kharrazi, MHI, MD, PhD
Center for Population Health IT
Department of Health Policy and Management
Johns Hopkins Bloomberg School of Public Health
624 N Broadway
Office 606
Baltimore, MD, 21205-2103
United States
Phone: 1 443 287 8264
Fax: 1 410 955 0470
Email: kharrazi@jhu.edu

Abstract

Background: The Meaningful Use (MU) program has promoted electronic health record adoption among US hospitals. Studies have shown that electronic health record adoption has been slower than desired in certain types of hospitals; but generally, the overall adoption rate has increased among hospitals. However, these studies have neither evaluated the adoption of advanced functionalities of electronic health records (beyond MU) nor forecasted electronic health record maturation over an extended period in a holistic fashion. Additional research is needed to prospectively assess US hospitals' electronic health record technology adoption and advancement patterns.

Objective: This study forecasts the maturation of electronic health record functionality adoption among US hospitals through 2035.

Methods: The Healthcare Information and Management Systems Society (HIMSS) Analytics' Electronic Medical Record Adoption Model (EMRAM) dataset was used to track historic uptakes of various electronic health record functionalities considered critical to improving health care quality and efficiency in hospitals. The Bass model was used to predict the technological diffusion rates for repeated electronic health record adoptions where upgrades undergo rapid technological improvements. The forecast used EMRAM data from 2006 to 2014 to estimate adoption levels to the year 2035.

Results: In 2014, over 5400 hospitals completed HIMSS' annual EMRAM survey (86%+ of total US hospitals). In 2006, the majority of the US hospitals were in EMRAM Stages 0, 1, and 2. By 2014, most hospitals had achieved Stages 3, 4, and 5. The overall technology diffusion model (ie, the Bass model) reached an adjusted R-squared of .91. The final forecast depicted differing trends for each of the EMRAM stages. In 2006, the first year of observation, peaks of Stages 0 and 1 were shown as electronic health record adoption predates HIMSS' EMRAM. By 2007, Stage 2 reached its peak. Stage 3 reached its full height by 2011, while Stage 4 peaked by 2014. The first three stages created a graph that exhibits the expected "S-curve" for technology diffusion, with inflection point being the peak diffusion rate. This forecast indicates that Stage 5 should peak by 2019 and Stage 6 by 2026.

XSL•FO
RenderX

Although this forecast extends to the year 2035, no peak was readily observed for Stage 7. Overall, most hospitals will achieve Stages 5, 6, or 7 of EMRAM by 2020; however, a considerable number of hospitals will not achieve Stage 7 by 2035.

Conclusions: We forecasted the adoption of electronic health record capabilities from a paper-based environment (Stage 0) to an environment where only electronic information is used to document and direct care delivery (Stage 7). According to our forecasts, the majority of hospitals will not reach Stage 7 until 2035, absent major policy changes or leaps in technological capabilities. These results indicate that US hospitals are decades away from fully implementing sophisticated decision support applications and interoperability functionalities in electronic health records as defined by EMRAM's Stage 7.

(J Med Internet Res 2018;20(8):e10458) doi:10.2196/10458

KEYWORDS

electronic health records; United States; hospitals; HIMSS EMRAM; Bass diffusion model

Introduction

Background

Technology policy in health care has profoundly affected service delivery and operational efficiencies [1,2]. The period from 2006 to 2016 saw a dramatic increase in electronic health record (EHR) adoption as well as expansion of its functionality [3]. These improvements are attributable to dual environmental pressures [4]. On one hand, the US government put policies into place that provided financial benefits to hospitals for adopting EHRs that met certain criteria [5,6]; on the other hand, internal pressures to adopt EHRs were significant as health systems sought to establish a competitive advantage through operational benefits associated with EHRs [7].

Research in the health care field has closely linked EHR technology adoption to business and clinical outcomes [8-10]. As a result, traditional innovation diffusion analysis, when applied to health care, is complicated by the dynamics found as multiple and varied EHR functionalities are introduced over time. These dynamics represent an opportunity to explore alternative conceptual and analytic approaches to examining technology diffusion and policy interactions.

Adoption of Electronic Health Records Among US Hospitals

The Health Information Technology for Economic and Clinical Health (HITECH) Act [11] was signed into law with the dual aims of accelerating EHR adoption and promoting their "meaningful use" (MU) by US hospitals [12]. HITECH appropriated billions of dollars to create financial incentives for hospitals that implement EHRs, which meet certain criteria designed to have a meaningful impact on care quality and cost [13]. Hospitals had to attest to MU Stage 2 program eligibility by 2016 to qualify and participate in the reward payments schema [14].

The impact of the HITECH Act has been evaluated in the health services literature so that policymakers can assess the extent to which their intended EHR adoption goals are being realized [15]. Indeed, the HITECH policy resulted in a rapid adoption of EHRs among nonfederal hospitals, increasing the adoption rate from 9.4% in 2008 to 83.8% in 2015 [3]. However, the EHR adoption rates were not equally distributed among all types of hospitals (eg, rural vs urban hospitals) [16], and certain functionalities were adopted earlier than others (eg, MU-mandated functions vs more advanced functions) [17,18].

Challenges With Using Meaningful Use Data to Assess Electronic Health Record Adoption

In the research literature that focuses on EHR technology adoption, analyses frequently rely on MU data for measuring current use percentages in a binary fashion [3]. In particular, the extant literature on EHR adoption has focused on the transition from paper to electronic data collection or the adoption of a specific function within an EHR [19]. As the adoption of a basic EHR became commonplace, researchers began to frame EHR adoption in terms of its ability to support specific tasks (eg, integrating clinical decision support [CDS] into clinical workflow; automating the collection of patient-reported outcomes; capturing high-quality data for clinical trials; and integrating population health management efforts) [20-26].

Concurrently, hospital planners adopted maturity models that sought to frame EHR implementation as a journey rather than an endpoint. The Healthcare Information and Management Systems Society (HIMSS) Analytics' Electronic Medical Record Adoption Model (EMRAM) [27] was developed by information technology (IT) and care delivery experts based on the observation that best practices in the industry were path dependent [28,29]. The EMRAM model identifies technological waypoints along an organization's adaptation journey that are sequential, specific, and measurable [27,28]. For example, closed-loop medical administration requires that decision support software be implemented prior to installing bar code readers that match patients to the prescription drugs they are receiving (ie, need for one level of technology before another level can be adopted as required and measured by the EMRAM model).

Using Diffusion of Innovation Models to Predict Electronic Health Record Maturation

Diffusion of innovation model produces "technology sophistication forecasts" that predict the degree to which a market or sector has and will adopt sequentially higher levels of functionality in the near future [29,30]. The diffusion of innovation literature and associated methods are critical to understanding and prediction of adoption dynamics. Taken together, these methodological approaches and conceptual frameworks offer a foundation upon which researchers can study the diffusion of innovation in cases where the supporting infrastructure is not replaced, a frequent condition of the Bass

XSL•FO
RenderX

model [31]. Furthermore, IT platforms such as EHRs, one where the hardware requirements become secondary to the software innovation, represent a new domain for modeling adoption dynamics using common diffusion models [32].

The purpose of this study was to explore when hospitals will achieve critical EHR functionality. HIMSS Analytics' EMRAM data and Bass diffusion models were used to assess current EHR capability levels and forecast future diffusion of EHR functionality levels.

Methods

Overview

In this study, we explored US hospitals' EHR technology adoption and implementation patterns accounting for functionality and application upgrades. We used the HIMSS EMRAM data to observe the granular change or progression of EHR functionality among hospitals. The same dataset was used to train the Bass diffusion model and then predict the EMRAM score (ie, the level of EHR functionality) for each hospital. The forecasted scores were aggregated across all hospitals within each future year to depict a national picture of EHR functionality improvements until 2035. We assumed no change in future policies that would affect health IT efforts or EHR functionality (eg, no new MU incentives). Similarly, no dramatic advancement in the technology itself is modeled (eg, effective Natural Language Processing or Artificial Intelligence) as such innovations would change the diffusion curves.

Data Sources

We used the HIMSS Analytics' EMRAM data since it provides an MU-comparable EHR adoption measure that takes a more granular approach to assessing functionality uptake (Table 1) [27]. EMRAM data are collected annually across all participating hospitals and are made publically available to interested researchers. HIMSS promulgates its "Annual Study," which is designed to capture a realistic portrait of the hospital's IT landscape. The data are submitted via a Web-based portal, phone, or spreadsheets [27]. Given the benchmarking value of these reports, a growing number of hospitals have participated in EMRAM's Annual Study since 2006. In 2014, over 5402 hospitals (86% of total US hospitals) completed the Annual Study. See Multimedia Appendix 1 for additional details about the EMRAM model and its stages of EHR maturation.

Theoretical Justification

Modeling the EHR diffusion using an adaptation approach requires two components. First, the technology must track progression through the diffusion stages as set by the "Diffusion of Innovation" theorem, resulting in an "S-curve" to measure the functional form of analysis appropriately [33]. Second, the assumption that new technologies completely displace prior generations needs to be relaxed [34]. Under these two conditions, the Bass "BB-01" model is an appropriate analytic approach to evaluate the diffusion of EHR as it adheres to these requirements [35]. See Multimedia Appendix 2 for additional details of the theoretical justification of using the Bass model to forecast EHR functionality improvement among hospitals.

Statistical Analysis

The EMRAM data were used as the basis of BB-01 statistical analyses, with estimates calculated in Microsoft Excel using nonlinear regression estimates. Visual Basic, Solver, and the SAS Model Procedure were also used to train and estimate several parameters used by the Bass model [34,35]. The algorithms and macros are publicly available [36]. See Multimedia Appendix 2 for additional details of algorithms used to train the model and predict EHR functionality adoption rates (ie, aggregated EMRAM scores).

Table 1. Summary of Electronic Medical Record Adoption Model (EMRAM) stages.

Stage	Description
Stage 0	The organization has not installed all of the three key ancillary department systems (laboratory, pharmacy, and radiology).
Stage 1	All three major ancillary clinical systems are installed (ie, pharmacy, laboratory, and radiology).
Stage 2	Major ancillary clinical systems feed data to a clinical data repository (CDR) that provides physician access for reviewing all orders and results.
Stage 3	Clinical documentation is implemented and integrated with the CDR for at least one inpatient service in the hospital. The Electronic Medication Administration Record application is implemented. Medical image access from picture archive and communication systems (PACS) is available for access by physicians outside the radiology department.
Stage 4	Computerized Practitioner Order Entry for use by any clinician licensed to create orders is added to the nursing and CDR environment along with the second level of clinical decision support (CDS) capabilities related to evidence-based medicine protocols.
Stage 5	A full complement of radiology PACS systems provides medical images to physicians via an intranet and displaces all film-based images.
Stage 6	Full physician documentation with structured templates and discrete data is implemented for at least one inpatient area. Level 3 of CDS provides guidance for all clinician activities. The closed-loop medication administration with bar-coded unit is fully implemented.
Stage 7	The hospital no longer uses paper charts to deliver and manage patient care and has a mixture of discrete data, document images, and medical images within its EHR environment. Clinical information can be readily shared via standardized electronic transactions with all entities that are authorized to treat the patient or a health information exchange.

XSL•FO
RenderX

Results

Study Populations and Base Adoption Rates

On average, approximately 5200 hospitals were represented in the EMRAM data across the years studied (2006-2014). The percentage of hospitals achieving various EMRAM stages varied across years (Figure 1). More than 96% of hospitals were identified to be in Stage 3 or below in 2006, while this number decreased to approximately 31% in 2014. Less than 4% of hospitals were in Stage 4 or higher in 2004, while this number dramatically increased across the consequent years: ~6% in 2008, ~20% in 2010, ~38% in 2012, and ~68% in 2014.

Model Performance

The overall model produced an adjusted R-squared of .91, suggesting a high model fit. Table 2 provides the estimates for the external motivation coefficient (p) and internal motivation coefficient (q) used in the final model (see Multimedia Appendix 2 for additional details). The two motivation coefficients show trends moving in opposite directions. For the earlier stages (ie, EMRAM Stages 1-3), the external influence is the primary motivation for EHR adoption. Starting with Stage 3, the internal influence metric begins to play a more impactful role, and eventually, it becomes the more important factor for EHR functionality adoption in Stages 4 and 5. Given the small number of hospitals that have achieved Stage 6 or 7, interpretation of the motivation coefficients was not undertaken for these stages.

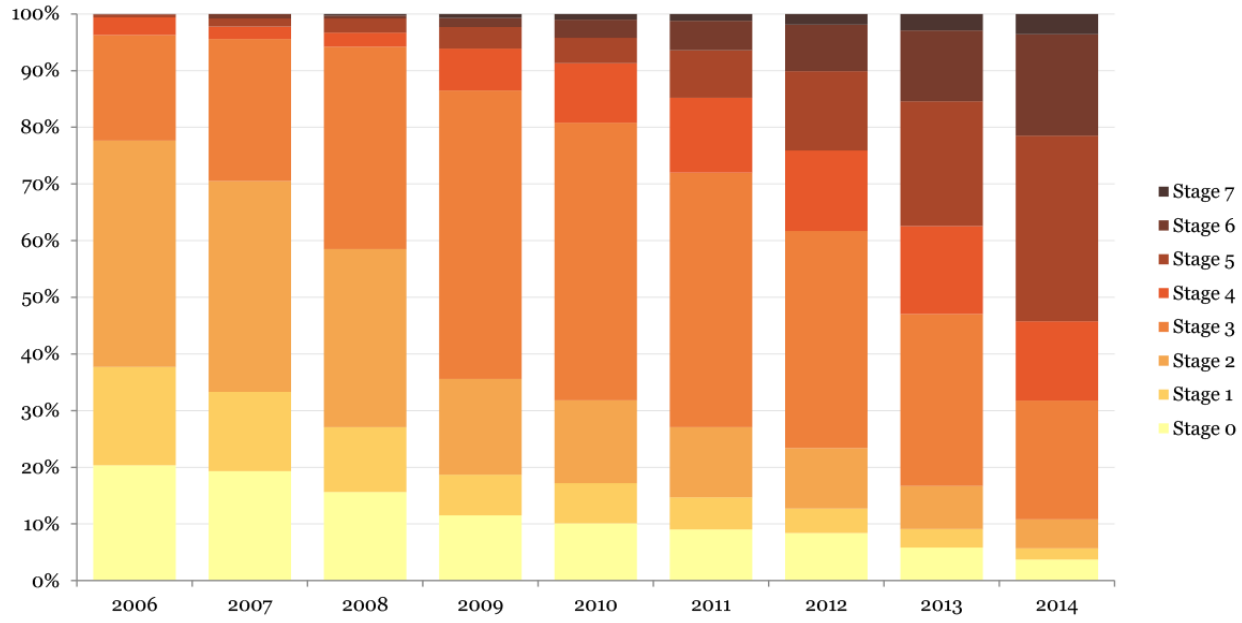
Electronic Health Record Maturation Forecast

The forecast used EMRAM data from 2006 to 2014 to estimate adoption levels to the year 2035. Table 3 offers a high-level snapshot of the forecasted EHR functionality progression from 2006 to 2035.

Figure 1 depicts the forecasted EHR functionality (ie, EMRAM stages) among US hospitals, assuming no major policy or technological changes in the future. Stages 0 and 1 seem to reach their peaks in the first year of observation as the use of EHRs predates 2006 (when HIMSS began to collect adoption data). By 2007, Stage 2 reaches its peak as well. Stage 3 reaches its peak by 2011, while Stage 4 reaches its peak in 2014. The first three stages create a graph that exhibits an "S-curve," with inflection point being the peak diffusion rate. Assuming current diffusion trajectories, the forecast predicts that Stage 5 will reach its peak by the year 2019 and Stage 6 by the year 2026. Although this forecast extends to the year 2035, no peak was readily observed for Stage 7. A considerable number of hospitals (800+) will stall their EHR adoption at Stage 5, while a higher number of hospitals (2200+) will remain in Stage 6 over an expanded period of time until 2035 (Figure 2).

Figure 3 depicts the cumulative volume of hospitals adopting EHRs with higher levels of functionality over the forecasted years. The cumulative volume of hospitals in Stage 4 is constant between the years 2010 and 2014; however, the volume of Stage 5 continues to grow. This is a clear indication of *leapfrogging*, suggesting that adopters either skipped Stage 4 or moved concomitantly with technology adoption for both Stages 4 and 5. Based on the analysis, most hospitals will be focused on the higher stages (Stages 5, 6, and 7) by the year 2025. It is also clear that Stage 7 will not reach a maximum or plateau by the end of the forecast window (Figure 3).

Figure 1. Historical Electronic Medical Record Adoption Model stages among US hospitals from 2006 to 2014.



XSL•FO
RenderX

Table 2. Parameter estimation and model performance.

Parameter	Stage 0	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6	Stage 7
M[a]	1606	888	2953	100	20	10	5	0
p[b]	0.964	0.234	0.680	0.255	0.015	0.043	0.064	0.026
q[c]	1	1E-9[d]	1E-9	0.120	0.505	0.354	1E-9	0.001

[a]M: market (sample) size for each stage.

[b]p: external motivation coefficient.

[c]q: external motivation coefficient.

[d]1E-9: 0.000 000 001.

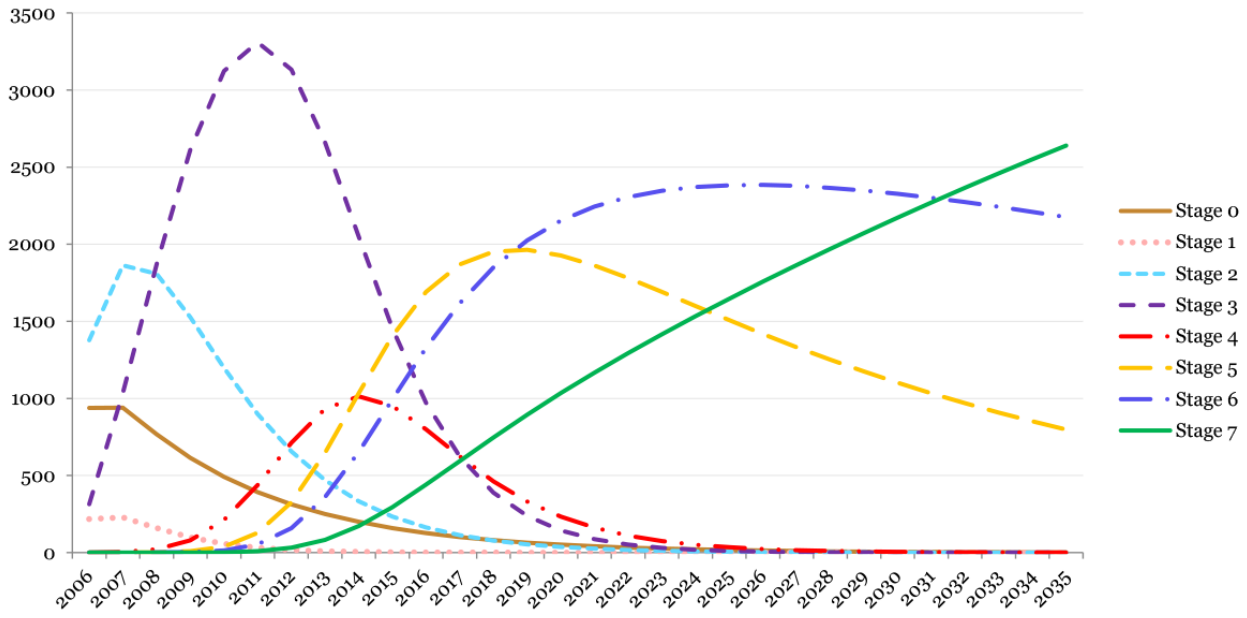
Table 3. Electronic Health Record (EHR) adoption milestones based on Electronic Medical Record Adoption Model stages.

Rate	Stage 0	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6	Stage 7
50% Year[a]	2008	2006	2008	2010	2014	2021	2025	2027
Max Year	2007	2007	2007	2011	2014	2019	2026	2035[b]

[a]Year that each EHR maturation stage reaches its mid-point.

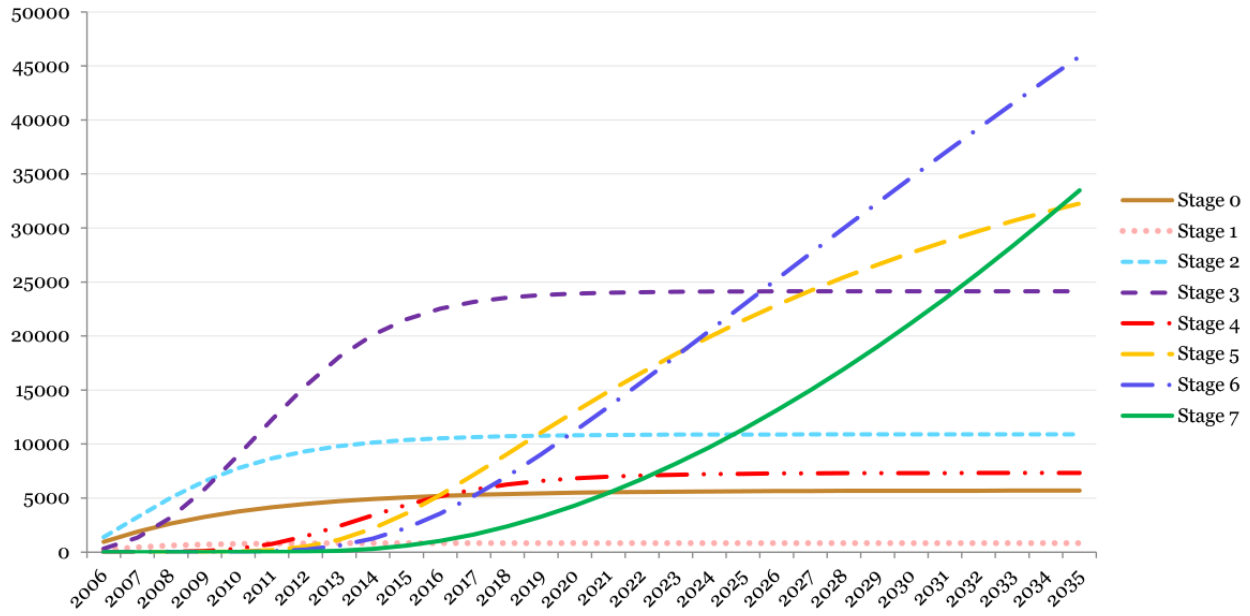
[b]Stage 7 did not reach a peak in any year until 2035.

Figure 2. Electronic health record functionality-level adoption among US hospitals using the Electronic Medical Record Adoption Model maturation stages (2014-2035 years are forecasted using the Bass model; vertical-axis represents the number of hospitals).



XSL•FO
RenderX

Figure 3. Cumulative electronic health record functionality-level adoption among US hospitals using the Electronic Medical Record Adoption Model maturation stages (2014-2035 years are forecasted using the Bass model; vertical-axis represents the cumulative number of hospitals).



Discussion

Principal Findings

EHR is a technology platform that allows for the integration of both hardware and software applications designed to improve care quality and increase operational efficiency. To those ends, the United States has introduced policies designed to promote EHRs' increasingly sophisticated functions. EHRs within hospitals are a prime example of a technology that is adopted and, then, repeatedly updated. This study seeks to outline a direction for future research critical to understanding the dynamics that drive EHR innovation among US hospitals.

We utilized the HIMSS Analytic EMRAM data to assess the EHR functionality levels retrospectively and train the Bass model of diffusion to forecast the adoption of new EHR functionality among US hospitals for the next two decades. The Bass model generated a good explanatory power, and the external and internal influence coefficients mapped closely to the existing regulatory environment. The forecast estimates were also consistent with other literature.

The findings can be evaluated and interpreted in two temporal categories: the retrospective pattern of EHR functionalities identified among the hospitals and the forecasted adoption pattern of EHR functionalities prospectively.

Retrospective Diffusion Pattern of Electronic Health Record Functionalities (2006-2016)

Given that EHRs had been discussed at the national level for decades before 2006, having Stage-2 as the most prevalent stage in that year is reasonable; however, it is interesting to note that the curve for Stage 1 never exceeded the curve for Stage 0. This phenomenon is the hallmark of leapfrogging and suggests that hospitals moved from Stage 0 directly to Stage 2 or Stage 3 (ie, hospitals adopted multiple generations of functionalities simultaneously rather than adapting them in separate phases). There are two potential explanations for the simultaneous, multistage adoption in the lower EMRAM levels: EHR vendors integrating multiple functions upfront and hospitals being overtly motivated by external factors (eg, MU incentives).

First, EHR vendors may have introduced multiple functions at once. As part of the MU program, the US government introduced an EHR vendor certification regime [37]. Its purpose was to assure hospitals that the EHR platforms would be capable of accommodating future innovations that were likely to be made mandatory features. All of the functionalities and clinical applications delineated in the HIMSS Analytics' EMRAM Stages 1 through 3 were required components in order for an EHR vendor to be successfully certified [38].

Second, hospitals may have wanted to move through the early stages quickly (Table 2). Considering that the external motivation measures for EMRAM Stages 1 through 3 (ie, p-coefficient: Stage 1 = .234; Stage 2 = .680; and Stage 3 = .255; Table 2) were higher than the internal motivation measures (ie, q-coefficient: Stage 1 = .000; Stage 2 = .000; and Stage 3 = .120; Table 2), the federal government's MU rewards appeared to have played a significant role in accelerating EHR diffusion. Many hospitals and health systems were incentivized to purchase all or most of the required EHR functionalities from a single vendor rather than having to acquire them separately and in multiple phases [39-41]. This strategy made it possible for hospitals to complete multiple levels at once, allowing them to collect the reward payments in a shorter period [7].

Additionally, hospitals with more recent EHR adoptions may have taken a simultaneous, multistage form in that product vendors began bundling functionalities and clinical capabilities together in a more holistic fashion [41,42]. Similar patterns have occurred in other technologies such as personal computers or mobile phones. Originally, personal computers were sold

XSL•FO
RenderX

with little more than an operating system. Consumers had to buy software programs (eg, internet browsers) to be able to use the machine. Later, personal computers came with many preinstalled applications so that consumers could start using their new machines "out-of-the-box." The net effect is that state-of-the-art information platforms' minimum feature sets encompass multiple generations of earlier innovations as a technology matures. It is likely that EHR vendors followed a similar pattern of increased technological sophistication as a matter of normal business prior to 2006 [43].

Prospective Diffusion Pattern of Electronic Health Record Functionalities (2016-2035)

The government's MU program did not provide rewards or incentives for the later EMRAM stages (ie, Stages 4 through 7). As a result, one of the previously noted major external motivations for adopting higher EHR functionalities was not in play. The BB-01 model effectively controls for this change in motivational factors, suggesting that internal motivation measures play a significantly large role in EHR functionality and clinical application adoption (Table 2). This can be interpreted as hospitals reaching the later EMRAM stages because they are "mission driven" (ie, internally motivated) to adopt the more sophisticated functionalities into their EHR platforms.

The lack of additional EHR incentives in this period will potentially cause internal factors to become the main driver for hospitals to adopt new EHR features. In this scenario, hospitals should observe the imminent need to request and adopt new EHR functionalities to achieve their higher-order goals (eg, quality improvement). For example, EMRAM's Stage 6 of EHR maturation requires the full adoption of CDS systems across the entire health care system for a variety of clinical practice guidelines. However, if the desired outcomes of a health system, either cost or clinical outcomes, are not aligned with such decision support enhancements in the underlying EHR platform, the hospitals may not have the internal pressure or desire to adopt the new EHR functionalities. Indeed, a complex series of internal factors may disincentivize such progression through EHR functionalities, specifically in a volatile health care market (eg, the cost of aggregating data and embedding a full array of CDS in clinical workflow may outpace the immediate benefits for the hospital). Hence, a considerable number of hospitals are forecasted not to reach Stage 6 by 2035 (Figures 2 and 3). In such a context, EMRAM's Stage 7 requirements can be harder to achieve as it further pushes the tradeoff between internal factors and expected outcomes by introducing more sophisticated EHR functionalities such as centralized data warehouses that can be readily used for analytical purposes as well as fully interoperable EHRs across hospitals (Table 1; see Multimedia Appendix 1).

Stage 7 of EMRAM requires the development of EHR-derived centralized data warehouses along with extensive analytic infrastructure by hospitals. Although the need for data analytics has grown tremendously among health care providers over the last decade [44], the value of such efforts is not clear for all types of hospitals [17]. On one end of this spectrum, academic medical centers and integrated or value-based delivery systems have realized the need for advanced analytics to push forward with their academic research agenda and quality improvement efforts, hence, accepting or planning for the development of centralized EHR-derived data warehouses. However, on the other end of this spectrum, with fewer internal incentives, smaller critical access and rural or community hospitals may not see the added value of investment in developing complex and often expensive EHR-derived data warehouses, unless the EHR vendors offer it as part of their basic or routine updates without additional charges (eg, EHR vendors attempting to keep their market share). The lack of immediate need for advanced EHR-derived analytics should be further investigated as a potential factor in impeding the attainment of EMRAM's Stage 7 among underresourced hospitals.

Another major milestone of EMRAM's Stage 7 for EHR maturation is the interoperability of EHRs among health care providers as well as integration of EHRs with local and regional health information exchanges (Table 1). The challenge of achieving wide interoperability in the health care sector, including hospitals, is a well-known fact, and a variety of causes have been studied (eg, lack of clear guidelines in the MU program) [7,45]. The federal government has extensively persuaded health care providers to adopt interoperability by providing roadmaps and facilitating the development and adoption of new information exchange standards [46]; however, hospital-based EHRs are still largely not interoperable with other settings [45,47,48]. Not reaching Stage 7 of EHR maturity by 2035 is concerning as the continued lack of interoperability may adversely affect patient safety, clinical outcomes, and population health management efforts [48,49]. Future studies should investigate and measure the levels of EHR interoperability among US hospitals and attempt to identify internal and external factors that may impede them from reaching—or drive them to reach—the highest EMRAM score. In addition, hospitals have varying level of capital assets, resources, and IT-driven mindsets that may lead to different adoption patterns of EHR functionalities. Future research should also investigate and discover EHR maturation patterns that are unique to specific hospital groups.

Limitations

Theoretical

Bass model [35] has been used to forecast technology diffusion in a variety of scientific domains [50-52]. The model positions the adoption of technology as either focused on consumers' replacement of existing products or the adoption of a new technology [34,53]. Furthermore, the more recent Bass "BB-01 Generations" analytic framework, which was used in this study, can be used to model the technological diffusion rates for repeated adoptions where customers upgrade a product as it undergoes rapid technological improvements [35]. However, similar to other simulation studies, the theoretical limits of the Bass model also limit the validity of the results and, consequently, the generalizability of the study [54].

Assumptions

We assumed no change in future policies or external factors that may affect EHR functionality advancements or health IT

XSL•FO
RenderX

adoption generally (eg, no new MU incentives; stable EHR market for hospital settings) [55]. New health IT policies may change the adoption rate of new EHR functionalities, specifically when incentives are directed for hospitals that are predicted to not achieve the higher stages of EHR maturation [43,56]. Therefore, the findings of this study should be interpreted within the limits of these assumptions and should be updated regularly when new EMRAM data becomes available after the roll out of such policies (eg, Centers for Medicare and Medicaid policies) [57]. Nonetheless, the likelihood of an exogenous factor supporting and reinforcing the adoption of advanced EHR functions is higher than that of a factor demoting the adoption of such functions (eg, more affordable IT infrastructure such as cloud-based EHRs).

Data Source

EMRAM does not include EHR adoption data for 2004 and 2005 when health IT policies started to take effect [7]. As this study used data starting in 2006, we could not observe some of the early dynamics that were derived from policies enacted before 2006. Furthermore, this study relies on the definitions and order of stages as defined by HIMSS Analytics in EMRAM (Table 1). Future studies can explore the impact on EHR functionality forecasts if some of these stages were collapsed into fewer categories, if external datasets are merged and used (eg, American Hospital Association's IT survey) [58], or if new methods are applied to break down the challenges of adopting EHRs into more refined internal or external factors [18].

Setting

This study only focuses on inpatient hospital settings and excludes the potential effect of EHR adoption trends in outpatient setting on hospitals. Future studies should investigate the interaction regarding adopting new EHR functionalities between inpatient and outpatient settings [59] (eg, hospitals joining a value-based network may require adopting new EHR functionalities such as higher interoperability with other participating health care providers).

Conclusion

This study sought to examine when more advanced features of EHRs will be adopted by US hospitals. Using the HIMSS EMRAM data and Bass diffusion models, we were able to forecast the adoption of EHR capabilities from a paper-based environment (Stage 0) to an environment where only electronic information is used to document and direct care delivery (Stage 7). According to the forecast, the majority of hospitals will not reach Stage 7 of EHR maturity by 2035, given that there are no major policy changes.

Acknowledgments

We acknowledge the support of the Healthcare Information and Management Systems Society (HIMSS) in providing the research team with the EMRAM data.

Conflicts of Interest

None declared.

Multimedia Appendix 1

HIMSS Analytics' Electronic Medical Record Adoption Model Stages.

[PDF File (Adobe PDF File), 48KB - jmir_v20i8e10458_app1.pdf]

Multimedia Appendix 2

Evolution of the diffusion model.

[PDF File (Adobe PDF File), 60KB - jmir_v20i8e10458_app2.pdf]

References

1. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 16;144(10):742-752. [Medline: 16702590]
2. Thune T, Mina A. Hospitals as innovators in the health-care system: A literature review and research agenda. Research Policy 2016 Oct;45(8):1545-1557. [doi: 10.1016/j.respol.2016.03.010]
3. Henry J, Pylypchuk Y, Searcy T, Patel V. The Office of the National Coordinator for Health Information Technology. 2016 May. Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2015 URL: https://dashboard.healthit.gov/evaluations/data-briefs/non-federal-acute-care-hospital-ehr-adoption-2008-2015.php [accessed 2018-03-20] [WebCite Cache ID 6y3xObNYE]
4. Geels FW, Schot J. Typology of sociotechnical transition pathways. Research Policy 2007 Apr;36(3):399-417. [doi: 10.1016/j.respol.2007.01.003]
5. Blind K, Petersen SS, Riillo CA. The impact of standards and regulation on innovation in uncertain markets. Research Policy 2017 Feb;46(1):249-264. [doi: 10.1016/j.respol.2016.11.003]

XSL•FO
RenderX

6. Werfel SH, Jaffe AB. Induced innovation and technology trajectory: Evidence from smoking cessation products. Research Policy 2013 Feb;42(1):15-22. [doi: 10.1016/j.respol.2012.04.013]
7. Jones SS, Rudin RS, Perry T, Shekelle PG. Health information technology: an updated systematic review with a focus on meaningful use. Ann Intern Med 2014 Jan 7;160(1):48-54. [doi: 10.7326/M13-1531] [Medline: 24573664]
8. Ashraf S, Moore C, Gupta V, Chowdhury A, Azad AK, Singh N, et al. Overview of a multi-stakeholder dialogue around Shared Services for Health: the Digital Health Opportunity in Bangladesh. Health Res Policy Syst 2015 Dec 09;13:74 [FREE Full text] [doi: 10.1186/s12961-015-0063-2] [Medline: 26646372]
9. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011 Jan;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
10. García-Romero A, Escribano Á, Tribó JA. The impact of health research on length of stay in Spanish public hospitals. Research Policy 2017 Apr;46(3):591-604. [doi: 10.1016/j.respol.2017.01.006]
11. US Department of Health and Human Services. Health Information Technology for Economic and Clinical Health (HITECH) Act, Title XIII of Division A and Title IV of Division B of the American Recovery and Reinvestment Act of 2009 (ARRA). 2009. URL: https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveredentities/hitechact.pdf [accessed 2018-07-18] [WebCite Cache ID 710Hmw1OH]
12. Maxson E, Jain S, Kendall M, Mostashari F, Blumenthal D. The regional extension center program: helping physicians meaningfully use health information technology. Ann Intern Med 2010 Nov 16;153(10):666-670. [doi: 10.7326/0003-4819-153-10-201011160-00011] [Medline: 21079224]
13. Buntin MB, Burke MF, Hoaglin MC, Blumenthal D. The benefits of health information technology: a review of the recent literature shows predominantly positive results. Health Aff (Millwood) 2011 Mar;30(3):464-471 [FREE Full text] [doi: 10.1377/hlthaff.2011.0178] [Medline: 21383365]
14. Office of the National Coordinator for Health Information Technology (ONC), Department of Health and Human Services. Health information technology: standards, implementation specifications, and certification criteria for electronic health record technology, 2014 edition; revisions to the permanent certification program for health information technology. Final rule. Fed Regist 2012 Sep 04;77(171):54163-54292 [FREE Full text] [Medline: 22946139]
15. Millenson ML. The long wait for medical excellence. Health Aff (Millwood) 2013 Oct;32(10):1757-1760. [doi: 10.1377/hlthaff.2013.0567] [Medline: 24101065]
16. DesRoches CM, Worzala C, Joshi MS, Kralovec PD, Jha AK. Small, nonteaching, and rural hospitals continue to be slow in adopting electronic health record systems. Health Aff (Millwood) 2012 May;31(5):1092-1099 [FREE Full text] [doi: 10.1377/hlthaff.2012.0153] [Medline: 22535503]
17. Adler-Milstein J, DesRoches CM, Kralovec P, Foster G, Worzala C, Charles D, et al. Electronic Health Record Adoption In US Hospitals: Progress Continues, But Challenges Persist. Health Aff (Millwood) 2015 Dec;34(12):2174-2180. [doi: 10.1377/hlthaff.2015.0992] [Medline: 26561387]
18. Chan KS, Kharrazi H, Parikh MA, Ford EW. Assessing electronic health record implementation challenges using item response theory. Am J Manag Care 2016 Dec 01;22(12):e409-e415 [FREE Full text] [Medline: 27982669]
19. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
20. Bae J, Huerta TR, Ford EW. Cardiovascular disease documentation and counseling in electronic medical records. Prev Med 2016 Jul;88:27-32. [doi: 10.1016/j.ypmed.2016.03.017] [Medline: 27002255]
21. Fiks AG, Grundmeier RW, Mayne S, Song L, Feemster K, Karavite D, et al. Effectiveness of decision support for families, clinicians, or both on HPV vaccine receipt. Pediatrics 2013 Jun;131(6):1114-1124 [FREE Full text] [doi: 10.1542/peds.2012-3122] [Medline: 23650297]
22. Bae J, Ford EW, Kharrazi HHK, Huerta TR. Electronic medical record reminders and smoking cessation activities in primary care. Addict Behav 2018 Feb;77:203-209. [doi: 10.1016/j.addbeh.2017.10.009] [Medline: 29065376]
23. Wu AW, Kharrazi H, Boulware LE, Snyder CF. Measure once, cut twice--adding patient-reported outcome measures to the electronic health record for comparative effectiveness research. J Clin Epidemiol 2013 Aug;66(8 Suppl):S12-S20 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.04.005] [Medline: 23849145]
24. Kharrazi H, Chi W, Chang HY, Richards TM, Gallagher JM, Knudson SM, et al. Comparing Population-based Risk-stratification Model Performance Using Demographic, Diagnosis and Medication Data Extracted From Outpatient Electronic Health Records Versus Administrative Claims. Med Care 2017 Aug;55(8):789-796. [doi: 10.1097/MLR.0000000000000754] [Medline: 28598890]
25. Kharrazi H, Weiner JP. IT-enabled Community Health Interventions: Challenges, Opportunities, and Future Directions. EGEMS (Wash DC) 2014;2(3):1117 [FREE Full text] [doi: 10.13063/2327-9214.1117] [Medline: 25848627]
26. Kharrazi H, Wang C, Scharfstein D. Prospective EHR-based clinical trials: the challenge of missing data. J Gen Intern Med 2014 Jul;29(7):976-978 [FREE Full text] [doi: 10.1007/s11606-014-2883-0] [Medline: 24839057]
27. The Healthcare Information and Management Systems Society (US). 2015. ELECTRONIC MEDICAL RECORD ADOPTION MODEL: HIMSS Analytics EMR Adoption Model (EMRAM) URL: http://www.himssanalytics.org/emram [accessed 2018-03-20] [WebCite Cache ID 6y3ySJZxD]

XSL•FO
RenderX

28. Chen C, Hsu C, Worley J. Evaluation systems of the electronic medical records: a comparative study. Paper presented at 2013 IEEE 4th International Conference on Software Engineering and Service Science; Beijing, China 2013:558-564 [FREE Full text] [doi: 10.1109/ICSESS.2013.6615371]
29. Vergne J, Durand R. The path of most persistence: An evolutionary perspective on path dependence and dynamic capabilities. Organization Studies 2011 Mar 24;32(3):365-382. [doi: 10.1177/0170840610397485]
30. Sood A, James GM, Tellis GJ, Zhu J. Predicting the path of technological innovation: SAW vs. Moore, Bass, Gompertz, and Kryder. Marketing Science 2012 Nov;31(6):873-1025. [doi: 10.1287/mksc.1120.0739]
31. Hauser J, Tellis GJ, Griffin A. Research on innovation: a review and agenda for marketing science. Marketing Science 2006 Nov;25(6):687-717. [doi: 10.1287/mksc.1050.0144]
32. Jiang B, Jerath K, Srinivasan K. Firm strategies in the “mid tail” of platform-based retailing. Marketing Science 2011 Sep;30(5):757-994. [doi: 10.1287/mksc.1110.0656]
33. Rogers EM. Diffusion of Innovations, 5th Edition. New York: Free Press; 2003.
34. Bass FM. A New Product Growth for Model Consumer Durables. Management Science 2004 Dec;50(12_supplement):1763-1893. [doi: 10.1287/mnsc.1040.0264]
35. Bass P, Bass F. Diffusion of technology generations: A model of adoption and repeat sales. 2001. URL: http://www.bassbasement.org/F/N/FMB/Pubs/Bass%20and%20Bass%202001.pdf [accessed 2018-07-18] [WebCite Cache ID 6y3z6qtGr]
36. Bass's BRI. Available at. The Bass Model URL: http://www.bassbasement.org [accessed 2018-07-18] [WebCite Cache ID 6y3zE53PV]
37. D'Amore JD, Mandel JC, Kreda DA, Swain A, Koromia GA, Sundareswaran S, et al. Are Meaningful Use Stage 2 certified EHRs ready for interoperability? Findings from the SMART C-CDA Collaborative. J Am Med Inform Assoc 2014;21(6):1060-1068 [FREE Full text] [doi: 10.1136/amiajnl-2014-002883] [Medline: 24970839]
38. Krist AH, Beasley JW, Crosson JC, Kibbe DC, Klinkman MS, Lehmann CU, et al. Electronic health record functionality needed to better support primary care. J Am Med Inform Assoc 2014;21(5):764-771. [doi: 10.1136/amiajnl-2013-002229] [Medline: 24431335]
39. Gardner E. Big bang. Small hospital tackles big HIS project. Health Data Manag 2010 Feb;18(2):56-59. [Medline: 20180457]
40. Huerta TR, Thompson MA, Ford EW, Ford WF. Implementing electronic lab order entry management in hospitals: Incremental strategies lead to better productivity outcomes. International Journal of Information Management 2013 Feb;33(1):40-47. [doi: 10.1016/j.ijinfomgt.2012.05.008]
41. Ford EW, Menachemi N, Huerta TR, Yu F. Hospital IT adoption strategies associated with implementation success: implications for achieving meaningful use. J Healthc Manag 2010;55(3):175-188; discussion 188. [Medline: 20565034]
42. Nelson RR, Buterbaugh K, Perl M, Gelijns AC. How medical know-how progresses. Research Policy 2011 Dec;40(10):1339-1344. [doi: 10.1016/j.respol.2011.06.014]
43. Barrett AK. Electronic Health Record (EHR) Organizational Change: Explaining Resistance Through Profession, Organizational Experience, and EHR Communication Quality. Health Commun 2018 Apr;33(4):496-506. [doi: 10.1080/10410236.2016.1278506] [Medline: 28157382]
44. Brennan PF, Chiang MF, Ohno-Machado L. Biomedical informatics and data science: evolving fields with significant overlap. J Am Med Inform Assoc 2018 Jan 01;25(1):2-3. [doi: 10.1093/jamia/ocx146] [Medline: 29267964]
45. Adler-Milstein J, Embi PJ, Middleton B, Sarkar IN, Smith J. Crossing the health IT chasm: considerations and policy recommendations to overcome current challenges and enable value-based care. J Am Med Inform Assoc 2017 Sep 01;24(5):1036-1043. [doi: 10.1093/jamia/ocx017] [Medline: 28340128]
46. The OOTNCFHIT. The Office of the National Coordinator for Health Information Technology (ONC). Connecting health and care for the nation: a shared nationwide interoperability roadmap URL: https://www.healthit.gov/sites/default/files/hie-interoperability/nationwide-interoperability-roadmap-final-version-1.0.pdf [WebCite Cache ID 6y3zf6UZD]
47. Swain M, Charles D, Patel V, Searcy T. ONC Data Brief. 2015 Apr. Health Information Exchange among U.S. Non-federal Acute Care Hospitals: 2008-2014 URL: https://www.healthit.gov/sites/default/files/data-brief/ONC_DataBrief24_HIE_Final.pdf [accessed 2018-03-20] [WebCite Cache ID 6y3ziATeU]
48. Samarath A, Sorace J, Patel V, Boone E, Kemper N, Rafiqi F, Yencha R, Kharrazi H. 2016. Measurement of interoperable electronic health care records utilization URL: https://aspe.hhs.gov/system/files/pdf/255526/EHRUtilizationReport.pdf [accessed 2018-06-18] [WebCite Cache ID 6y3zlhNya]
49. Kharrazi H, Lasser EC, Yasnoff WA, Loonsk J, Advani A, Lehmann HP, et al. A proposed national research and development agenda for population health informatics: summary recommendations from a national expert workshop. J Am Med Inform Assoc 2017 Jan;24(1):2-12 [FREE Full text] [doi: 10.1093/jamia/ocv210] [Medline: 27018264]
50. Easingwood CJ, Mahajan V, Muller E. A nonuniform influence innovation diffusion model of new product acceptance. Marketing Science 1983 Aug;2(3):203-317. [doi: 10.1287/mksc.2.3.273]
51. Bawa K, Shoemaker R. The effects of free sample promotions on incremental brand sales. Marketing Science 2004 Aug;23(3):275-467. [doi: 10.1287/mksc.1030.0052]
52. Manchanda P, Xie Y, Youn N. The role of targeted communication and contagion in product adoption. Marketing Science 2008 Nov;27(6):949-1136. [doi: 10.1287/mksc.1070.0354]

XSL•FO
RenderX

53. Sillup GP. Forecasting the adoption of new medical technology using the Bass model. J Health Care Mark 1992 Dec;12(4):42-51. [Medline: 10123584]
54. Norton JA, Bass FM. A diffusion theory model of adoption and substitution for successive generations of high-technology products. Management Science 1987 Sep;33(9):1069-1207. [doi: 10.1287/mnsc.33.9.1069]
55. Butler M. Journal of AHIMA. CMS to end Meaningful Use in 2016 URL: http://journal.ahima.org/2016/01/13/cms-to-end-meaningful-use-in-2016 [accessed 2018-03-20] [WebCite Cache ID 6y409zUpu]
56. Heath ML. IT cultural conflict, trust belief in IT artifacts, and user involvement: conceptual study of Meaningful Use resistance. In: MWAIS 2015 Proceedings. 2015 Presented at: The Midwest United States Association for Information Systems; 2015; Pittsburg, KS URL: http://aisel.aisnet.org/mwais2015/14/
57. Centers for Medicare and Medicaid Services. MACRA: Delivery system reform, Medicare payment reform. URL: https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/value-based-programs/macra-mips-and-apms/macra-mips-and-apms.html [accessed 2018-07-18] [WebCite Cache ID 6y40KtOos]
58. American Hospital Association. AHA Data Products URL: https://www.aha.org/data-insights/aha-data-products [accessed 2018-03-20] [WebCite Cache ID 6y40VEoyg]
59. Sholler D. Resistance to mandatory technology use - the case of EMR in the United States healthcare industry. In: MCIS 2016 Proceedings. 2016 Presented at: Mediterranean Conference on Information Systems; 2016; Paphos, Cyprus URL: https://aisel.aisnet.org/mcis2016/38/

Abbreviations

CDR: clinical data repository
CDS: clinical decision support
EHR: electronic health record
EMRAM: Electronic Medical Record Adoption Model
HIMSS: Healthcare Information and Management Systems Society
IT: information technology
MU: Meaningful Use
PACS: picture archive and communication systems

Edited by G Eysenbach; submitted 20.03.18; peer-reviewed by C Jung, K Ford-Eickhoff; comments to author 23.05.18; revised version received 01.06.18; accepted 16.06.18; published 07.08.18.

Please cite as:
Kharrazi H, Gonzalez CP, Lowe KB, Huerta TR, Ford EW
Forecasting the Maturation of Electronic Health Record Functions Among US Hospitals: Retrospective Analysis and Predictive Model
J Med Internet Res 2018;20(8):e10458
URL: http://www.jmir.org/2018/8/e10458/
doi:10.2196/10458
PMID:30087090

©Hadi Kharrazi, Claudia P Gonzalez, Kevin B Lowe, Timothy R Huerta, Eric W Ford. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 07.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Developing Embedded Taxonomy and Mining Patients' Interests From Web-Based Physician Reviews: Mixed-Methods Approach

Jia Li[1], PhD; Minghui Liu[1], MB; Xiaojun Li[2], PhD; Xuan Liu[1], PhD; Jingfang Liu[3], PhD

[1]School of Business, East China University of Science and Technology, Shanghai, China

[2]Xi'an Research Institute of Hi-Tech, Xi'an, China

[3]School of Management, Shanghai University, Shanghai, China

Corresponding Author:
Jingfang Liu, PhD
School of Management
Shanghai University
99 Shangda Road
Shanghai, 200444
China
Phone: 86 21 66133790 ext 803
Fax: 86 21 66134284
Email: jingfangliu2014@hotmail.com

Abstract

Background: Web-based physician reviews are invaluable gold mines that merit further investigation. Although many studies have explored the text information of physician reviews, very few have focused on developing a systematic topic taxonomy embedded in physician reviews. The first step toward mining physician reviews is to determine how the natural structure or dimensions is embedded in reviews. Therefore, it is relevant to develop the topic taxonomy rigorously and systematically.

Objective: This study aims to develop a hierarchical topic taxonomy to uncover the latent structure of physician reviews and illustrate its application for mining patients' interests based on the proposed taxonomy and algorithm.

Methods: Data comprised 122,716 physician reviews, including reviews of 8501 doctors from a leading physician review website in China (haodf.com), collected between 2007 and 2015. Mixed methods, including a literature review, data-driven-based topic discovery, and human annotation were used to develop the physician review topic taxonomy.

Results: The identified taxonomy included 3 domains or high-level categories and 9 subtopics or low-level categories. The physician-related domain included the categories of medical ethics, medical competence, communication skills, medical advice, and prescriptions. The patient-related domain included the categories of the patient profile, symptoms, diagnosis, and pathogenesis. The system-related domain included the categories of financing and operation process. The F-measure of the proposed classification algorithm reached 0.816 on average. Symptoms (Cohen d=1.58, Δu=0.216, t=229.75, and P<.001) are more often mentioned by patients with acute diseases, whereas communication skills (Cohen d=−0.29, Δu=−0.038, t=−42.01, and P<.001), financing (Cohen d=−0.68, Δu=−0.098, t=−99.26, and P<.001), and diagnosis and pathogenesis (Cohen d=−0.55, Δu=−0.078, t=−80.09, and P<.001) are more often mentioned by patients with chronic diseases. Patients with mild diseases were more interested in medical ethics (Cohen d=0.25, Δu 0.039, t=8.33, and P<.001), operation process (Cohen d=0.57, Δu 0.060, t=18.75, and P<.001), patient profile (Cohen d=1.19, Δu 0.132, t=39.33, and P<.001), and symptoms (Cohen d=1.91, Δu=0.274, t=62.82, and P<.001). Meanwhile, patients with serious diseases were more interested in medical competence (Cohen d=−0.99, Δu=−0.165, t=−32.58, and P<.001), medical advice and prescription (Cohen d=−0.65, Δu=−0.082, t=−21.45, and P<.001), financing (Cohen d=−0.26, Δu=−0.018, t=−8.45, and P<.001), and diagnosis and pathogenesis (Cohen d=−1.55, Δu=−0.229, t=−50.93, and P<.001).

Conclusions: This mixed-methods approach, integrating literature reviews, data-driven topic discovery, and human annotation, is an effective and rigorous way to develop a physician review topic taxonomy. The proposed algorithm based on Labeled-Latent Dirichlet Allocation can achieve impressive classification results for mining patients' interests. Furthermore, the mining results reveal marked differences in patients' interests across different disease types, socioeconomic development levels, and hospital levels.

(J Med Internet Res 2018;20(8):e254) doi:10.2196/jmir.8868

XSL•FO
RenderX

KEYWORDS

labeled-LDA; physicians; topic modeling; topic taxonomy; Web-based review

Introduction

Background

With the popularity of the internet, more and more people search Web-based information when they make decisions regarding health care providers. Among those sources, Web-based physician reviews are most frequently cited. Physician review websites (PRWs) permit patients and third-party reviewers to grade both physicians and hospitals in popular Web-based forums. Therefore, Web-based physician reviews reduce uncertainty surrounding the experience and serve as a valuable source of information for patients making choices [1-3]. Examples of PRWs include Healthgrades.com [4], Vitals.com [5], RateMDs.com [6], and a host of smaller, less-organized websites. Despite the subjectivity of evaluations in the reviews, Web-based reviews are an important source of Web-based information because they are perceived as more reliable and trustworthy than traditional information sources [7]. In addition, the Web-based physician review not only provides valuable information for patients who want to make a wise choice among health care providers but also provides some insights into physicians and hospitals who intend to improve their services in the future [1,8]. In summary, Web-based physician reviews are invaluable gold mines that merit further investigation [2,8,9].

Literature Review

Most extant studies on physician reviews only use numeric variables, such as volume (number of reviews) and valence (rating score), in their empirical analysis and fail to consider the information in the review text; for example, Hao [10] examined the development of the Web-based doctor review practice in China, focusing on the number of doctors and specialty areas available for Web-based review, the number of Web-based reviews for these doctors, the specialty areas where doctors are more likely to be reviewed, and the quantitative rating score distribution. Li et al [11] examined how the proportion and position of negative reviews on such websites influence readers' willingness to choose the reviewed physician and found that an increase in the proportion of negative reviews led to a reduced willingness to use the physician's services. In addition, a primacy effect was found for negative reviews, such that readers were less willing to use the physician's services when negative reviews were presented before the positive reviews. Yang et al [1] explored the effect of the patient- and system-generated information on patients' Web-based searches, evaluations, and decisions, suggesting that the positive patient- and system-generated information on physicians' service quality positively impacted patients' reactions at different stages. Moreover, they found that synergies between the patient-generated and the system-generated information positively associated with patients' decisions to consult a physician.

However, the information from numeric ratings is very limited compared with the whole review text, leading to a substantial loss of valuable information. The description of a medical consultation is multifaceted [12]. Therefore, a single number, such as a rating score of satisfaction or attitude, might not be adequate for patients to identify entire information relevant to physician choice. In addition, Web-based rating scores may not accurately capture or serve as a proxy for the physician quality. Recently, Daskivich et al [13] indicated that Web-based ratings of specialist physicians fail to predict objective measures of the quality of care or peer assessment of the clinical performance. Furthermore, Web-based ratings tend to be exaggerated at the upper or lower ends of the quality spectrum [14]. Therefore, recent studies on physician reviews focused more on the rich information embedded in the review text; for example, Hao and Zhang [9] automatically extracted hidden topics from Web-based physician reviews using text mining techniques to examine what Chinese patients said about their doctors and whether these topics differ across various specialties. Hao et al [15] compared the positive and negative reviews of obstetrics and gynecology doctors from the two most popular Web-based doctor rating websites in the United States and China. Grabner-Kräuter and Waiguny [2] explored how certain characteristics of physician reviews affected the evaluation of the review and users' attitude toward the rated physician and suggested a positive main effect of the number of reviews as well as an interaction effect with the style of the review. If the physician received only a few reviews, fact-oriented reviews induced a more favorable attitude toward the physician compared with emotional reviews; however, there was no such effect when the physician received many reviews. Lockie et al [16] investigated which textual and content elements were related to the perceived usefulness of Web-based reviews for doctors (general practitioners) and indicated that reviews with a more narrative or experiential style were generally perceived as more useful than more fact-based or very short reviews.

The prior research most related to this study is Hao and Zhang [9], in which authors extracted hidden topics from Web-based physician reviews using the Latent Dirichlet Allocation (LDA). However, there are several differences between this study and that by Hao and Zhang [9]. First, in the study conducted by Hao and Zhang [9], only the data-driven approach was used to derive the topics. However, in this study, we use a mixed-methods approach consisting of a literature review, data-driven-based topic discovery, and human annotation approaches to rigorously and systematically develop a physician review taxonomy. Second, there is no theoretical basis in the study conducted by Hao and Zhang [9]. In this study, Maslow's hierarchy of needs theory [17] is used to build a theoretical framework to guide the research questions and the whole paper.

Third, only LDA is used in the study conducted by Hao and Zhang [9]. In this study, both LDA and labeled-LDA are used. As an unsupervised machine-learning approach, LDA is used to find the initial topics embedded in the reviews, whereas labeled-LDA, as a semisupervized machine-learning approach, is used to classify the reviews into topics. Fourth, Hao and Zhang [9] made a comparison across 4 specialty areas (ie, internal medicine, surgery, obstetrics-pediatric, and Chinese

XSL•FO
RenderX

medicine) as well as between Chinese doctor reviews and American doctor reviews. In this study, we focus on patients' interests across different disease types and hospital levels.

Research Questions

In this study, we are first interested in developing a systematic topic taxonomy embedded in the physician reviews. Although many studies have explored the text information of physician reviews, very few have focused on developing a systematic topic taxonomy embedded in the physician reviews. Web-based physician reviews are multifaceted in nature. Without such a topic taxonomy, the findings of the physician review text mining research are hard to compare directly. Hence, the first step toward mining physician reviews is to determine how the natural structure or dimension is embedded in reviews. Of note, extant studies that involve physician review topics are *ad hoc* in nature. There is a lack of research that develops the topic taxonomy in a rigorous and systematic manner. Therefore, the first research question is proposed as follows.:

> RQ1: What topics are embedded within the Web-based physician review?

Second, patients' interests and needs may not be universal across different disease types and hospital levels. According to Maslow's hierarchy of needs theory [17], low-level needs, such as physiological requirements and safety, must be satisfied before high-level needs, such as self-fulfillment, are pursued. When a need is mostly satisfied, it no longer motivates, and the next higher need takes its place. Therefore, some topics that reflect low-level needs should be important for low-level hospitals, and some other topics that reflect high-level needs should be important for high-level hospitals. Hence, in this study, we are further interested in investigating different interests and needs of different patients. Therefore, the second research question is proposed as follows:

> RQ2: Do different patients have different interests and needs? If so, how do their interests and needs differ across different disease types and hospital levels?

In summary, we aim to develop a hierarchical topic taxonomy to uncover the latent structure of physician reviews and illustrate its application for mining patients' interests based on the proposed taxonomy and algorithm in this study.

Data

In this study, we focus on a leading Web-based PRW, "Good Doctor" (haodf.com [18]), in China. China has its own hospital categorization system. According to the facilities and technique strength, Chinese hospitals are classified into three levels, with the A-level being the best and the C-level the worst. A-level hospitals have the best physicians and medical equipment; these not only provide specialized medical services but also undertake many teaching and research tasks. However, C-level hospitals focus on mass coverage and only provide basic medical services for community members.

The website haodf.com [18] was set up to help Chinese consumers to find good and appropriate specialists for their health care problems based on Web-based reviews [10]. As of July 2017, the platform had 507,365 registered doctors and 2,745,304 physician reviews. The physicians who received reviews on haodf.com [18] cover all major specialty areas. Because haodf.com [18] was designed to find good doctors, more physicians are from high-level hospitals (eg, A-level hospitals) than low-level hospitals (eg, C-level hospitals), especially from the largest and most famous hospitals in Beijing and Shanghai. Anyone who visited a physician can write a review on the website. The review process is anonymous because the website masks the reviewers' username. Other personal information about reviewers is also unavailable to the public. In addition, reviewers can voluntarily disclose their sociodemographic information and health status. Similar to other health rating systems, such as healthgrades.com [4], users are allowed to rate a physician with scores and comments. Writing a review is voluntary. For more information about the development of Web-based physician reviews in China and haodf.com [18], please refer to the analysis of Hao [10].

A total of 122,716 physician reviews of 8501 doctors from haodf.com [18] were collected by a Web spider. The dataset covers the most frequently reviewed top 9 diseases (diabetes, gastric cancer, hypertension, hyperthyroidism, infantile diarrhea, infantile pneumonia, infertility, influenza, and liver cancer) from 2007 to 2015 on the website; this distribution is different from that of Western countries. For example, the incidence rates for liver and gastric cancer are higher in China than those in the United States [19]. Table 1 summarizes the descriptive statistics information for the review data.

Table 1. Descriptive statistics of the review data.

Disease	Reviews (N=122,716), n (%)	Doctors (N=10,764), n (%)	Reviews per physician	Average length per review (words)
Infertility	71,556 (58.31)	3573 (33.19)	20.0	444
Infantile pneumonia	21,839 (17.80)	2148 (19.96)	10.2	435
Infantile diarrhea	711 (0.60)	275 (2.60)	2.6	499
Influenza	1796 (1.46)	643 (6.00)	2.8	408
Hyperthyroidism	3028 (2.47)	872 (8.10)	3.5	383
Diabetes	20,849 (16.99)	2627 (24.41)	7.9	360
Liver cancer	1679 (1.37)	288 (2.70)	5.8	569
Gastric cancer	1053 (0.86)	231 (2.20)	4.6	560
Hypertension	205 (0.20)	107 (1.00)	1.9	369

XSL•FO
RenderX

Methods

Framework

Figure 1 shows the proposed methodology to discover the hidden topics and build the taxonomy. The framework consists of three major steps as follows: summarizing topics from the literature; discovering hidden topics using a data-driven approach; and finalizing the topic taxonomy with human annotations.

Step 1: Literature Review

The literature was reviewed to determine the initial topic taxonomy. We used 5 investigators to search the MEDLINE, EBSCO, and Web of Science databases with keywords "physician review topic," "physician review," "doctor review," "topic taxonomy," and "topic dimension" between 2013 and 2016. A total of 65 papers were found. Then, each investigator scored 65 papers on the relevance, 5 being the robust score and 1 being the weakest score. In addition, a paper was considered relevant if it included a taxonomy developed from the physician reviews. The agreement between any two investigators on the relevance score ranged from 0.6 to 0.81. Finally, we ranked the papers based on the score. After scoring each paper, we identified 7 relevant papers whose scores were >20 [20-26]. The 7 papers were carefully reviewed to identify the potential topics. Table 2 summarizes the topics identified (topics with different names but the same meaning were combined). The topics listed in Table 2 provided a good starting point to build the final topic taxonomy and helped to interpret the output of the LDA algorithm in step 2. As will be discussed later, the identified topics might be classified into 3 domains or high-level categories, including physician-related categories, patient-related categories, and system-related categories.

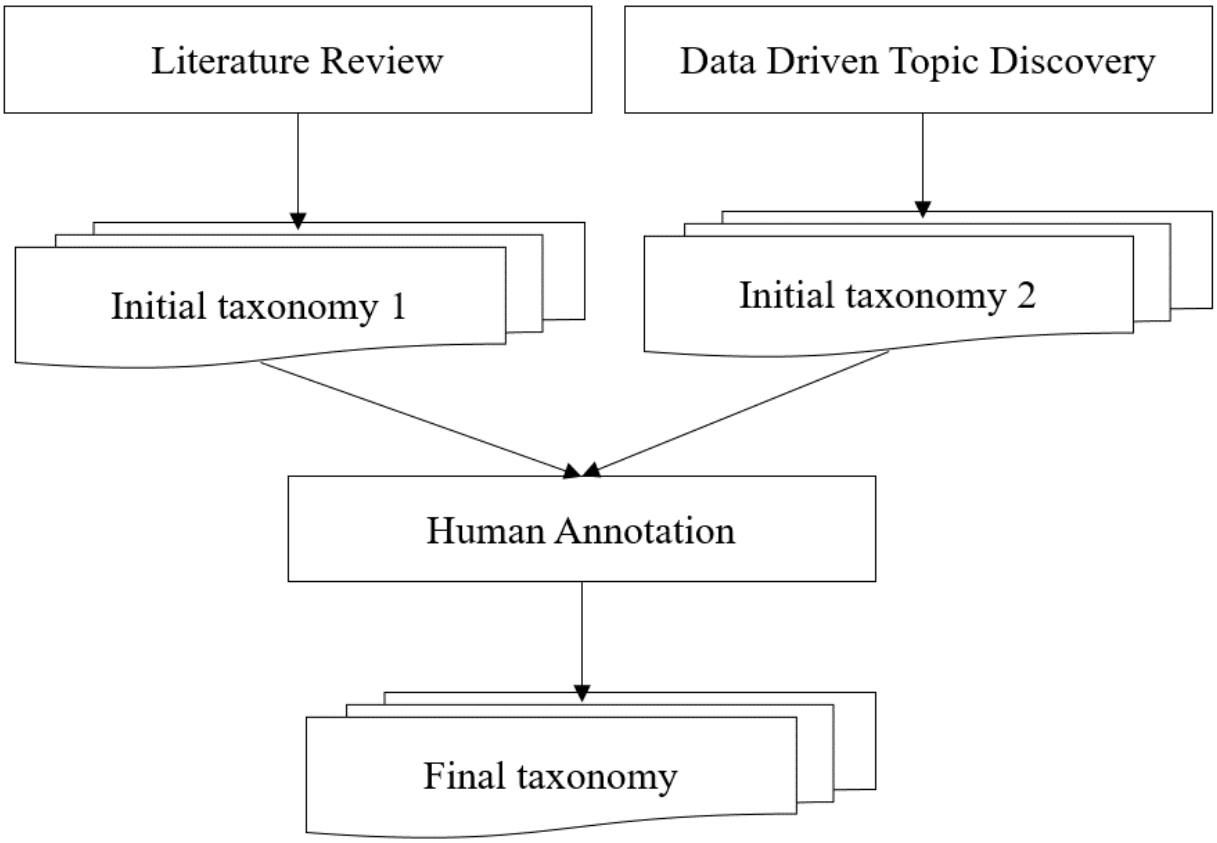
Step 2: Data-Driven Analysis

The data-driven approach uses the LDA algorithm to explore the hidden topics among physician reviews. The approach consists of the following 2 steps.

Text Preprocessing

The preprocessing consists of several steps such as word segmentation, part-of-speech tagging, word stopping, and word replacement. Because the physician reviews are downloaded from a Chinese website and the Chinese words are not delimited, word segmentation is a necessary step. In this study, HanLP [27] was used to segment the Chinese text into a vector of words. HanLP provides part-of-speech tagging for each output token. Only meaningful phrases, such as nouns, verbs, adjectives, and adverbs, are retained after word segmentation. Therefore, the whole sentence is transformed into a vector with meaningful phrases.

Figure 1. Framework for the development of the physician review topic taxonomy.



XSL•FO
RenderX

Table 2. Physician review topics identified from the literature with related descriptions and references.

Topics	Description and reference
Physician knowledge and skill	• Physician knowledge [20-23] • Professional competence [21-23] • Satisfaction with treatment [24]
Medical ethics (relational conduct)	• Friendliness and caring attitude [24] • Interpersonal style [21-23] • Punctuality [20,25] • Time spent with the patient [24]
Medicine and advice	• Information and advice [24] • Medicine and Pain control [26]
Communication	• Communication attributes [22,24] • Communication with patients [25] • Communication with doctors [26]
Environment	• Environment [26] • Condition and equipment of a doctor's office [25]
Business process	• Appointment [25] • Availability and accessibility [22] • Other Staff [20,26] • Responsiveness [26] • Systems issues [21,23,26]
Financing	• Cost of medical advice [22] • Financing [26]

Another important step in preprocessing is word replacement. There are two reasons for word replacement. First, a synonym, a word having the same or nearly the same meaning as another word. Synonyms make the meaning more difficult to capture because the same meaning might have different forms. Synonyms are identified with the help of a synonym thesaurus (Harbin Institute of Technology IR-Lab Tongyici Cilin [28]). Second, to increase the number of instances for some named entities, such as body temperature, age, number, height, weight, and so on. For example, both 39°C and 40°C may refer to the body temperature. Therefore, we replaced them with the same special symbol, such as #BODY_TEMPERATURE#, to increase the number of cases for body temperature. Furthermore, a rule-based name entity recognition task was performed to recognize the important entities of the MUC-7 (Message Understanding Conference-7) framework, such as date, location, money, organization, percentage, person, and time [29].

Explorative Topic Discovery

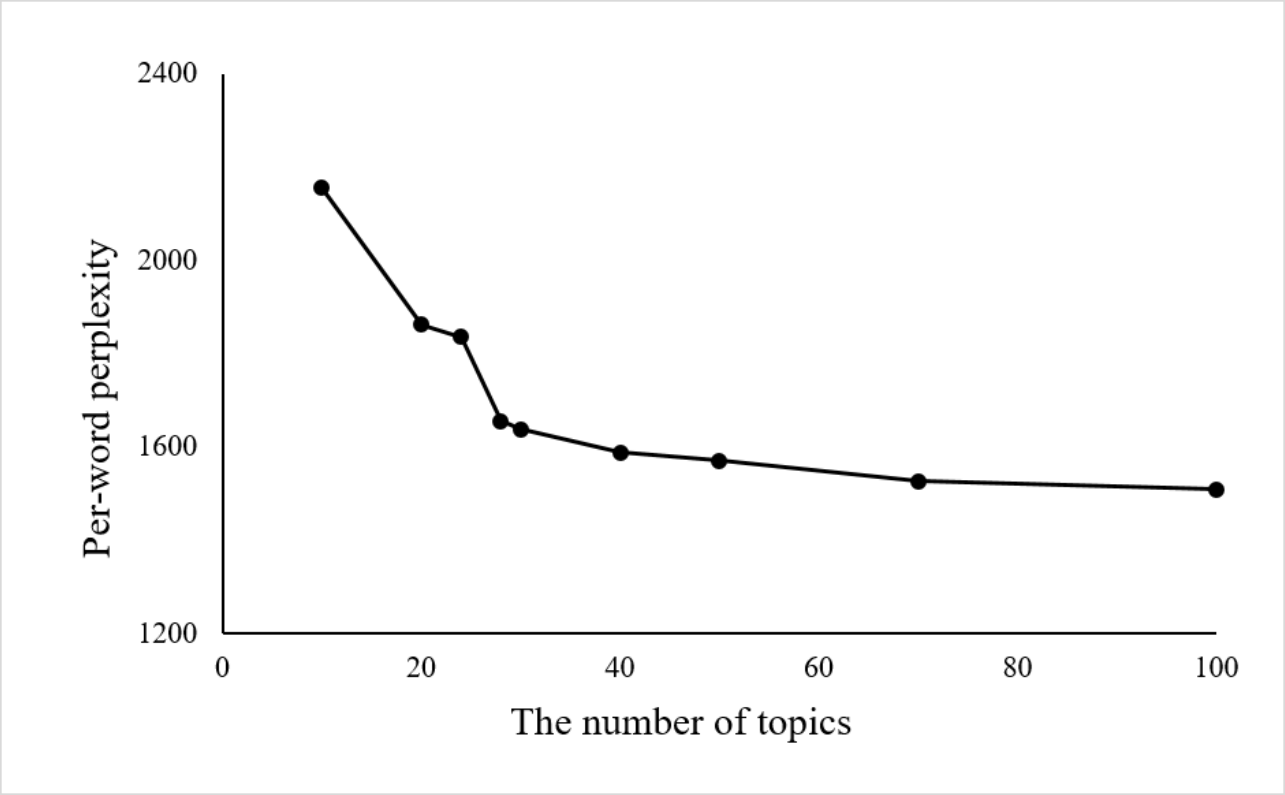
Topic modeling was used to explore the hidden structure of physician review texts. Topic modeling is a type of statistical model for discovering the abstract "topics" that occur in a collection of documents. In this study, LDA was used as the tool for topic discovery [30,31]. LDA is a generative statistical model that allows sets of observations to be explained by unobserved categories that explain why some parts of the data are similar; for example, if observations are words collected into documents, LDA posits that each document is a mixture of a small number of topics and that each word's creation is attributable to one of the document's topics. In LDA, each document may be viewed as a mixture of various topics, where each document is considered to have a set of topics that are assigned to it via LDA. Moreover, a piece of text is generated as random mixtures over latent topics, where each topic is characterized by a distribution over words. LDA has been widely used to explore topics from the medical text [32-34].

An important question in LDA (and perhaps, for most cluster algorithms) is to determine the best number of categories (or clusters). In this study, the perplexity was used to determine the best category number. The perplexity measures the predictive power of competing models in language modeling. For a collection D of M reviews, the per-word perplexity is defined as:

The perplexity can be understood as the predicted average number of equally likely words in certain positions, and it is a monotonically decreasing function of the log-likelihood [35,36]. Therefore, a lower perplexity over a held-out text means a higher log-likelihood, that is, better predictive performance. Figure 2 presents the predictive power of the LDA model in terms of the per-word perplexity for different numbers of topics.

XSL•FO
RenderX

Figure 2. Per-word perplexity as a function of the number of topics.



As shown in Figure 2, the perplexity monotonically decreases with an increase in the number of topics, eventually tending to converge to a fixed value. Therefore, a higher number of topics is more preferred than a lower number. However, LDA often learns some topics that are difficult to interpret, and the model provides no tools for tuning the generated topics to suit an end-use application, even when time and resources exist to provide document labels. It is very difficult to interpret the medical or managerial meaning if the topic number is too large. Thus, there should be a balance between the perplexity and the interpretability. In this study, we chose the number of topics to be 30 because we observed that the perplexity decreases much more slowly when the topic number is >30. Table 3 presents the results of 30 identified topics; each topic is represented by a group of keywords. In addition, Table 3 provides the interpretation for each topic.

Step 3: Human Annotation

After the data-driven approach for topic discovery, a human annotation was performed to finalize the topic taxonomy [37,38]. The topics identified from the literature (shown in Table 2) and the data (shown in Table 3) provided an initial topic taxonomy. We independently employed 10 graduate students who majored in information systems as coders to annotate 500 reviews each. They were asked to try to classify each review into the topics identified either in Table 2 or Table 3. Before the annotation started, they were asked to read another 200 reviews to familiarize themselves with the text. In addition, a training session was introduced to make sure each coder understood the meanings of the topics. When they encountered a review that could not be included in any previously identified topic, a new topic was created [39]. In this phase, the patient profile (PP) was created. Because each review was coded by 3 annotators, the discrepancies among the 3 coders were discussed until all conflicts were resolved. The agreement between any 2 coders on their initial review ranged from 0.6 to 0.81. After human annotation, some topics with similar content but different names (eg, topics 22 and 28 in Table 3) were combined. Table 4 presents the final topic taxonomy; it is a hierarchical taxonomy consisting of 3 high-level categories, namely physician-related categories, system-related categories, and patient-related categories. In addition, the explanations, keywords, and examples for each subtopic are provided.

Automatic Topic Classification Algorithm

In this study, labeled-LDA was used to classify review texts into topics identified in the previous section. Labeled-LDA is a generative model for multilabeled corpora [40]. As a natural extension of both LDA (by incorporating supervised learning methods) and multinomial naïve Bayes (by incorporating a mixture model), it performs well in solving the problem of topic identification in multilabeled texts with improved the interpretability over LDA. The competitive advantage of labeled-LDA over a strong baseline discriminative classifier, such as a support vector machine, on multilabel text classification tasks has been validated by previous studies [40]. Multimedia Appendix 1 provides further details about labeled-LDA.

XSL•FO
RenderX

Table 3. Physician review topics discovered by the Latent Dirichlet Allocation.

Topic	Keywords	Interpretation
Topic 1	mood, confidence, from the heart, prospect, encourage, comfort, warm, relax, pressure, psychological distress, kind	Physician knowledge and skill (positive treatment evaluation)
Topic 2	state of an illness, patience, treatment, asking patients questions, carefully, detailed asking, serious and responsible, situation, interpretation, diagnosis, explain	Communication (asking and listening)
Topic 3	thanks, age, mother, daughter, health, son, sincerely, help, opportunity, once again, baby	Physician knowledge and skill (thanks for the positive results)
Topic 4	operate, success, test tube baby, transplant, eggs, failure, artificial insemination, thanks, natural, embryo, give up	Therapeutic schedule
Topic 5	thanks, work, a good man, grateful to, happiness, smooth life, health, express, family, wish, blessing	Physician knowledge and skill (thanks and positive results)
Topic 6	treatment, effect, symptoms, disease, obvious, turn for the better, combined therapy of Chinese and Western medicine, cure, ease one's pain, acupuncture	Physician knowledge and skill (treatment effect)
Topic 7	at that time, in the heart, tell, really, no, worry about, know, be afraid of, nervous, feeling, happy	Communication (asking and listening)
Topic 8	good, no, online, really, more, looking for, give it a try, want, evaluation, have a try, experts	Business process (make an appointment)
Topic 9	see a doctor, successful, methods, treatment, hyperthyroidism, normal, indicators, drug, test, take the medicine, advice	Therapeutic schedule
Topic 10	good, attitude, better, special, patience, medical skill, kind, satisfaction, curative effect, technology, rare	Medical ethics (Relational conduct)
Topic 11	patient, time, experts, compare, need to do, experience, feeling, think, go to a doctor, trust, choose	Business process (make an appointment)
Topic 12	Traditional Chinese Medicine, toning your body, medicine, effect, Western medicine, proprietary Chinese medicine, drink, how long, adhere to, body	Medicine
Topic 13	check, operation, B ultrasound, report, inspection result, problem, reexamination, advice, test, draw blood, requirements	Medical examination
Topic 14	pregnancy, age, conceive, no, get married, treatment, infertility, friend, thank you, check, sperm	Disease symptoms (eg, infertility)
Topic 15	see a doctor, every time, time, patient, looking for, good, body, every day, disease, often, pay attention to	Medical ethics
Topic 16	surgery, in the hospital, out of the hospital, restore, thank you, treatment, arrange, excision, time, team, admitted to hospital	Business process (responsiveness, ward management)
Topic 17	patients, patience, problem, consulting, explain, thank you, touch, outpatient service, reply, online, encounter	Communication (explaining)
Topic 18	surgery, laparoscopic, the fallopian tubes, uterus, imaging, hysteroscopy, lining, found, uterine fibroids, adhesion, cyst	Therapeutic schedule
Topic 19	looking for, see a doctor, better, disease, attitude, good, introduce, friend, effect, cure	Physician knowledge and skill (treatment effect)
Topic 20	menstruation, pregnancy, follicles, ovulation, normal, monitoring, abortion, ovary, hormone, stimulate ovulation	Disease symptoms (eg, infertility)
Topic 21	hope, good, believe, really, feeling, confidence, better, best, think, as soon as possible, must be	Physician knowledge and skill (treatment effect, confidence)
Topic 22	cough, pneumonia, catch a cold, have a fever, how long, infusion, medicine, take the medicine, age, take an injection, turn for the better	Disease diagnosis (eg, infantile pneumonia and influenza)
Topic 23	medicine, prescribing, take the medicine, no, expensive, effect, cheap, capsule, disorderly, prescribe medicine disorderly, test	Medicine and prescribing
Topic 24	ask questions, do not, no, problem, impatient, want, speak, directly, medical records, to see a doctor, why	Communication (listening and explaining)
Topic 25	medical skill, medical ethics, noble, patients, good, exquisite, technology, worth, amiable, trust, enthusiasm	Physician skill and medical ethic

XSL•FO
RenderX

Topic	Keywords	Interpretation
Topic 26	patient, attitude, seriously, patience, responsible, better, kindly, to see a doctor, careful, work, good	Medical ethics (Relational conduct)
Topic 27	do not, know, want, think, bad, why, problem, do not know, compare, comfortable, what do I do	Communication (listening and explaining)
Topic 28	treatment, diabetes, blood sugar, control, drug, insulin, diet, how long, stable, the state of illness, adjust	Disease diagnosis (eg, diabetes)
Topic 29	registered, time, line up, make an appointment, outpatient service, a plus sign, time, to see a doctor, difficult, particular requirement, expert registered ticket	Business process (make an appointment)
Topic 30	do not, good, no, how long, pain, serious, at a time, diarrhea, blood, already, appear	Physician knowledge and skill (pain control)

The corpus annotated in the last section was used to provide supervised information for labeled-LDA. The output was the word distribution for each topic as well as the topic distribution for each physician review. Multimedia Appendix 2 shows the top words learned by labeled-LDA (the Chinese words were translated into English). Each topic was illustrated using a word cloud map, in which the font size was proportional to the probability of the word occurring in the topic.

Algorithm Evaluation

An evaluation was performed to estimate the performance of the proposed algorithm. The blind test bed consisted of 200 reviews covering all 9 diseases mentioned above. The algorithm output was compared with the human annotation results. Because the output of labeled-LDA was a distribution over all topics, we decided to label a topic only when its probability was higher than 1/L (*L* is the number of topics).

Table 5 presents the evaluation results. Precision, recall, and *F*-measure were used as the evaluation metrics [41]. Precision answers the following question:

> Given all retrieved responses, what is the probability that the retrieved responses are relevant?

Recall answers the following question:

> Given all relevant responses, what is the probability that the relevant responses are retrieved?

For classification tasks, the terms true positives (*tp*), true negatives (*tn*), false positives (*fp*), and false negatives (*fn*) compare the results of the classifier under test with trusted external judgments. The terms positive and negative refer to the classifier's prediction (sometimes known as the expectation), and the terms true and false refer to whether that prediction corresponds to the external judgment (sometimes known as the observation). Precision and recall are then defined as:

Because precision and recall are inversely related, the *F*-measure (also known as *F*-score) was used to evaluate the trade-off between them. In this case, an unweighted *F*-score was used. The unweighted *F*- score is the harmonic mean of precision and recall and is defined as:

As shown in Table 5, the algorithm achieved impressive results with an average *F*-measure of 0.816 and with the highest score for hypertension (0.904) and the lowest score for infertility (0.682).

XSL•FO
RenderX

Table 4. The taxonomy of Web-based physician reviews (final).

Topic	
Physician-related topics	
Subtopic: Medical ethics	
Explanation	Physician's interpersonal manners and behaviors perceived in the patient-physician interaction, including politeness, decency, patient participation in the treatment process, caring, listened to, understood, taken seriously.
Keywords	Kind, good, attitude, medical ethics, patient, considerate, noble, careful, polite, show a patient every consideration, respect, indifference, pressing, unfriendly
Examples	1. This doctor always gets to the treatment room on time and treats me with good manners. 2. At one point, our family member had inappropriate bleeding from several sites on the body which (a doctor) dismissed as unimportant. Eventually we were able to convince other people in the intensive care unit to take care of the situation.
	LDA[a] topic: 10, 15, 25, 26
Original topic	Topic from the literature: Medical ethics (Relational conduct)
Frequency (n)	2592
Subtopic: Medical competence	
Explanation	Patient's perceptions of the doctor's competence, experience and knowledge, including accuracy of the diagnostic process and treatment, safe practices or outcomes, observations of missed or correct care, and pain control.
Keywords	Good, medical skill, curative effect, turn for the better, improved significantly, high degree of medical skill, recovered, exacerbation, aggravation
Examples	1. His medical skill is so superb that I was soon discharged from the hospital. 2. Doctor Lin is very knowledgeable. He provided me with alternative treatment plans.
Original topic	LDA topic: 1, 3, 5, 6, 19, 21, 25, 30 Topic from the literature: Physician knowledge and skill
Frequency (n)	2391
Subtopic: Medical advice and prescription	
Explanation	The treatment solution given by a physician. It also includes the side effects of medications and treatments.
Keywords	Treatment, drug, antibiotics, surgery, insulin, take the medicine, injection, oral drugs, infusion, azithromycin, chemotherapy
Examples	1. Doctor Wang gave Gemcitabine, 5-fluorouracil for the treatment and explained it in detail in view of the different symptoms. 2. I did not get fat. My body shape is normal. No other side effects.
Original topic	LDA topic: 4, 9, 12, 18, 23 Topic from the literature: Medicine and advice
Frequency (n)	1910
Subtopic: Communication skills	
Explanation	Communication skills such as listening skills, asking skills and explaining skills. Make sure the patient is understood and address each patient's question well.
Keywords	Patiently answering question, explicate, cannot understand, do not get it, state of illness, problem, ask, detailed, explain, urge again and again, enjoin, analyze
Examples	What doctor Wang said was easily understood. He will patiently explain what you do not understand.
Original topic	LDA topic: 2, 7, 17, 24, 27 Topic from the literature: Communication
Frequency (n)	2133
System-related topics	
Subtopic: Financing	
Explanation	The price for similar services, confusion about billing issues, the stress of dealing with billing departments and unexpected out-of-pocket costs.
Keywords	Expenses, money, expensive, spend money, cheap, registration fee, waste money, inexpensive, difference, price, expenses for medicine, affordable price

XSL•FO
RenderX

Topic	
Examples	The service is good, but it's expensive. Your insurance may not cover most of it. [...] Call the hospital and ask for the cost first before you commit to any specialty sessions [...]
Original topic	Topic from the literature: Financing
Frequency (n)	1903
Subtopic: Operation process	
Explanation	Discharge information, responsiveness, clinical environment and equipment, and make an appointment. Discharge information allows patients to report their feeling and experiences related to discharge, such as perceived diagnostic errors and revisits to the emergency department. Responsiveness paid attention to whether the physician was accessible when needed or not, such as when seeking emergency department care or immediate care, when the primary doctor is unavailable. Clinical environment and equipment focus on patients' perception of the comfort of the environment of the hospital, quality and scarcity of equipment, and efficiency.
Keywords	Difficult, make an appointment, see a doctor, flow, service, ward, consulting, wait in line, beds, reexamination, need to be, out of the hospital, online consulting
Examples	After I got out of the hospital, doctor Liu called me several times. I told him my recent conditions. It took a long time before the ambulance reached the hospital. Fortunately, the doctors rescued the little boy from the jaws of death. I was very comfortable in the emergency room because at least one expert was on duty every night who regularly checked the patients. The doctor's office was clean, and the hospital was not crowded. I was comfortable.
Original Topic	LDA topic: 8, 11, 13, 16, 29 Topic from the literature: Environment, Business process
Frequency (n)	1934
Patient-related topics	
Subtopic: Patient profile	
Explanation	Patient's demographic information, including age, sex, address, occupation, diet, hobbies, etc
Keywords	Gender, height, age, born, weight, address, habits, allergic, hair
Examples	My aunt feels uncomfortable. She is 60 years old.
Original Topic	Human annotation
Frequency (n)	2339
Subtopic: Symptom	
Explanation	Symptom changes and sign changes. Symptom changes refer to subjective discomfort, abnormal feelings, and obvious pathological changes. Sign changes refer to the anomalous changes which could be diagnosed with objective tools.
Keywords	Cough, symptoms, transfer, have a fever, abnormal, virus, high blood pressure, diarrhea, serious, catch a cold, thin, swelling, lymph node, bloodshot
Examples	I have flu-like symptoms, for example, headache, cough, fever and rhinorrhea.
Original topic	LDA topic: 14, 20
Frequency (n)	1606
Subtopic: Diagnosis and pathogenesis	
Explanation	Disease types, features and pathogenesis.
Keywords	Pathogenesis, influencing factor, pneumonia, diabetes, blood pressure, fallopian tubes, function, cause, infertility, congenital, gastric cancer, genetic abnormality
Examples	I added a side dish of vegetables for my son. Soon he developed severe diarrhea, 10 times a day.
Original topic	LDA topic: 22, 28
Frequency (n)	1666

[a]LDA: Latent Dirichlet Allocation.

XSL•FO
RenderX

Table 5. The topic classification performance.

Disease	Precision	Recall	*F*-score
Infertility	0.672	0.738	0.682
Liver cancer	0.942	0.815	0.862
Influenza	0.742	0.795	0.762
Hypertension	0.863	0.958	0.904
Hyperthyroidism	0.947	0.774	0.821
Diabetes	0.883	0.878	0.871
Gastric cancer	0.927	0.704	0.795
Infantile pneumonia	1.000	0.722	0.820
Infantile diarrhea	0.967	0.745	0.826
Average	0.882	0.792	0.816

Results

Patient Listing Analysis Framework

Listening to patients is very important for health care providers to understand their customer needs and increase their satisfaction. The major results of patients' interesting mining are summarized below. Welch's *t* tests were conducted to test significant differences recorded on the topics for different groups of patients [42]. The *t* test can be used to determine if two datasets are significantly different from each other. Here is an example to illustrate the procedure for testing a given topic's (eg, medical ethics [ME]) ratio differences for patients with acute diseases versus chronic diseases. First, we obtained the topic ratio for each message for topic ME from labeled-LDA algorithm's output. For example, "message 1=0.245" indicates that message 1 contains 24.5% of the ME topic. To reduce unnecessary noise, topic ratios below 1/9 were set to 0. Then, we calculated relevant variables, such as the mean, SD, and sample size, for each group (ie, acute diseases vs chronic diseases), which were further used as inputs for the *t* test to determine whether there is a marked difference in the topic ratio for acute diseases versus chronic diseases.

Acute Versus Chronic Diseases

Influenza, infantile pneumonia, infantile diarrhea, and hyperthyroidism were categorized as acute diseases, whereas hypertension, diabetes, and infertility were categorized as chronic diseases. Figure 3 shows the interests of patients with acute and chronic diseases; the length of the bar represents the percentage of messages that contain a given topic; for example, a value of 0.141 for acute disease with topic ME means that 14.1% (14.1/100) of the messages for the acute disease group contained topic ME. The effect size (eg, small, medium, large, very large, and huge) was labeled for each topic to indicate the magnitude of the difference. The effect size was first measured by Cohen *d* [43] and then interpreted as small (<0.01), medium (0.01-0.20), large (0.20-0.50), very large (0.50-0.80), and huge (0.80-2.0) according to the values of Cohen *d* [44].

As shown in Figure 3, patients with acute diseases were more interested in symptoms (Cohen d=1.58, Δu=0.216, t=229.75, and P<.001). Meanwhile, patients with chronic diseases were more interested in communication skills (CS; Cohen d=−0.29, Δu=−0.038, t=−42.01, and P<.001), financing (F; Cohen d=−0.68, Δu=−0.098, t=−99.26, and P<.001), and diagnosis and pathogenesis (DAP; Cohen d=−0.55, Δu=−0.078, t=−80.09, and P<.001).

Patients with acute diseases were more interested in symptoms likely because symptoms are the most important factor for a patient describing an illness experience or during a consultation. In contrast, patients with chronic diseases were more concerned with CS, financing, and DAP. Because chronic diseases cannot be quickly cured and usually have an extended duration, patients are very familiar with their treatments and prescriptions. Therefore, medical competence (MC) and medical advice and prescription (MAP) are not the focus of their reviews. However, self-management and financial burden are major concerns for patients with chronic disease. These findings indicated that patients with different disease development rates (acute vs chronic) indeed exhibited different concerns to be addressed. Therefore, the training focus for physicians should be tailored to accommodate these distinct needs.

Mild Versus Serious Diseases

Influenza is categorized as a mild disease, whereas liver cancer and gastric cancer are categorized as serious diseases. Figure 4 shows the interests of patients with mild and serious diseases. Patients with mild diseases were more interested in ME (Cohen d=0.25, Δu=0.039, t=8.33, and P<.001), operation process (OP; Cohen d=0.57, Δu=0.060, t=18.75, and P<.001), PP (Cohen d=1.19, Δu=0.132, t=39.33, and P<.001), and symptoms (S; Cohen d=1.91, Δu=0.274, t=62.82, and P<.001). Patients with serious diseases were more interested in MC (Cohen d=−0.99, Δu=−0.165, t=−32.58, and P<.001), MAP (Cohen d=−0.65, Δu=−0.082, t=−21.45, and P<.001), financing (F; Cohen d=−0.26, Δu=−0.018, t=−8.45, and P<.001), and DAP (Cohen d=−1.55, Δu=−0.229, t=−50.93, and P<.001).

Patients with mild disease were more interested in ME, OP, PP, and symptoms because mild diseases are usually simple, nonlife-threatening, and easy to cure. Therefore, these patients placed more attention on the service quality and patient empowerment. In contrast, we observed that patients with serious diseases were more concerned with MC, MAP,

XSL•FO
RenderX

financing, and DAP. Because serious diseases are usually complicated, life-threatening, and carry a heavy financial burden, the disease pathogenesis, treatment method, and financial issues are the major concerns for these patients. These findings demonstrated that patients presenting with different disease severities exhibited different interests or concerns to be addressed.

High- Versus Low-Level Hospitals

A-level hospitals were categorized as high-level hospitals, whereas B- and C-level hospitals were categorized as low-level hospitals. The topic ratios were compared with the hospital level, as seen in Figure 5. Patients at high-level hospitals were more interested in OP (Cohen d=0.08, Δu=0.014, t=6.73, and P<.001), CS (Cohen d=0.08, Δu=0.012, t=8.33, and P<.001), and financing (F; Cohen d=0.10, Δu=0.016, t=7.07, and P<.001).

Figure 3. Topic ratio for acute and chronic diseases.Small (< 0.01), medium (0.01 - 0.20), large (0.20 - 0.50), very large (0.50 - 0.80) and huge (0.80 - 2.0) are effect sizes according to the magnitudes of Cohen's d. ME: Medical ethics; MC: Medical competence; MAP: Medical advice and prescription; CS: Communication skills; F: Financing; OP: Operation process; PP: Patient profile; S: Symptoms; DAP: Diagnosis and pathogenesis.

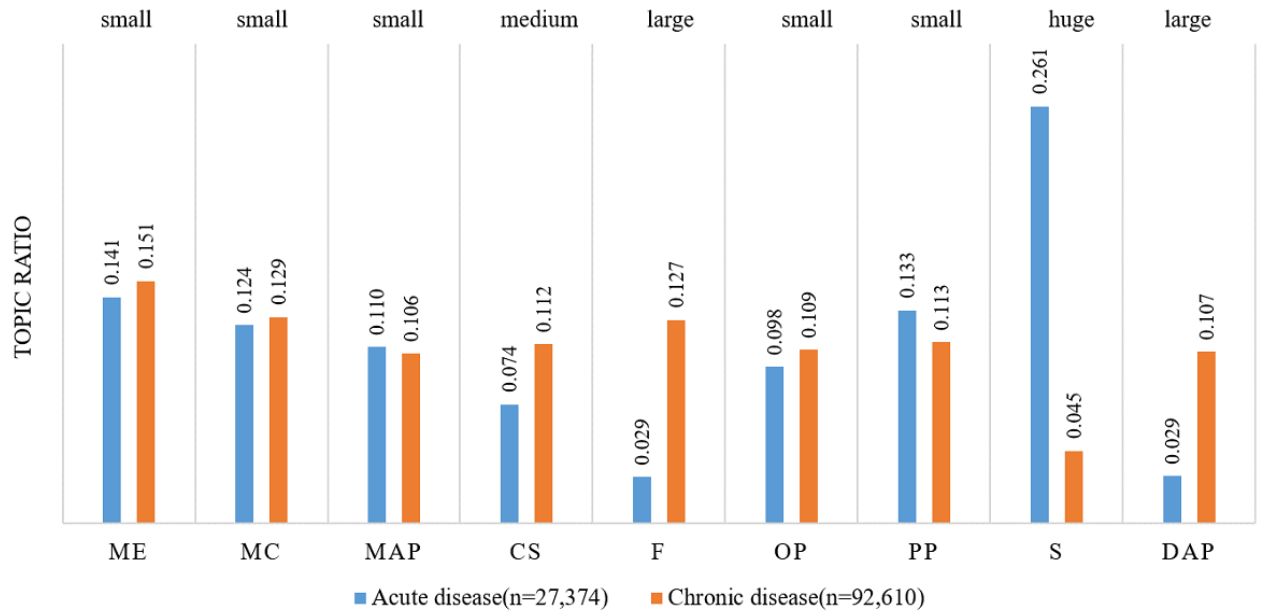
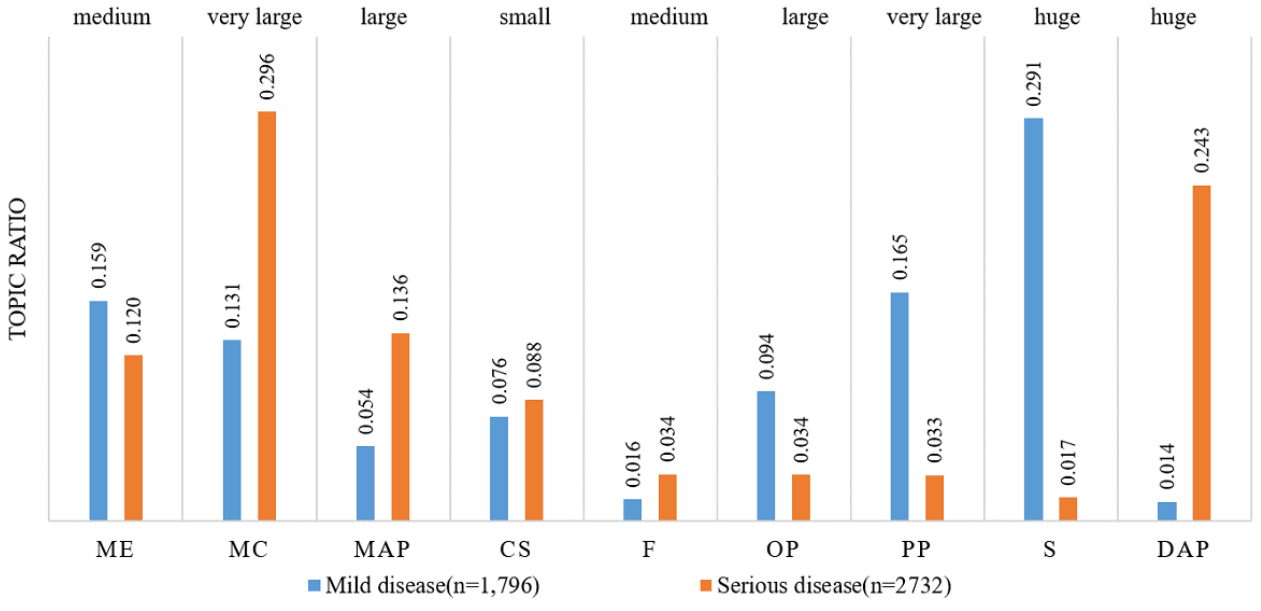
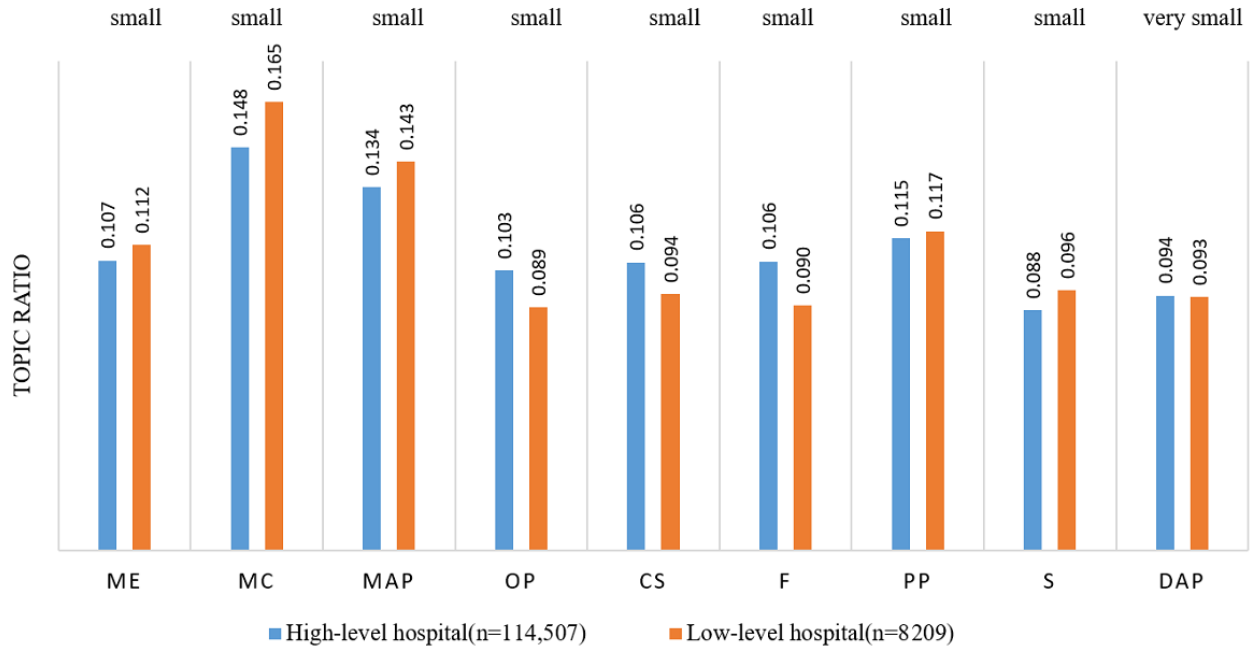


Figure 4. Topic ratio for mild and serious diseases. ME: Medical ethics; MC: Medical competence; MAP: Medical advice and prescription; CS: Communication skills; F: Financing; OP: Operation process; PP: Patient profile; S: Symptoms; DAP: Diagnosis and pathogenesis.



XSL•FO
RenderX

Figure 5. Topic ratio for high level and low level hospitals. ME: Medical ethics; MC: Medical competence; MAP: Medical advice and prescription; CS: Communication skills; F: Financing; OP: Operation process; PP: Patient profile; S: Symptoms; DAP: Diagnosis and pathogenesis.



However, the effect sizes for those differences were small. Patients at low-level hospitals were more interested in ME (Cohen d=−0.09, Δu=−0.005, t=−7.58, and P<.001), MC (Cohen d=−0.05, Δu=−0.017, t=−4.48, and P<.001), MAP (Cohen d=−0.03, Δu=−0.009, t=−2.88, and P<.01), and symptoms (S; Cohen d=−0.04, Δu=−0.008, t=−3.67, and P<.001); however, the effect sizes for those differences were also small.

It is clear that customer demand in high-level hospitals likely differed from that in low-level hospitals. As shown in Figure 5, ME and MC are the two most outstanding concerns for patients from low-level hospitals. In addition, MC is a concern because physician competence at low-level hospitals might be reduced compared with high-level hospitals, thus worrying patients. ME is another concern because customers of low-level hospitals are mostly low-income patients with little education, making conflicts with doctors or nurses more likely. Patient distrust in low-level hospitals, such as small clinics, village health clinics, community health service stations, and neighborhood health centers, in China has been reported in several prior studies [45]. Multimedia Appendix 3 describes complete comparisons of topic ratio differences across the disease type and hospital level.

Discussion

Principal Findings

In this study, we build a taxonomy that includes 3 domains or high-level categories and 9 subtopics or low-level categories using mixed methods. Then, a classification algorithm based on labeled-LDA is proposed. The evaluation result shows that the F-measure of the proposed classification algorithm reaches 0.816 on average. Our analysis on large review corpus suggests that patients with different disease types or hospital levels have different concerns to be addressed.

Comparison With Prior Work on the Taxonomy Development Method

Prior work on developing a physician review taxonomy mainly follows one of the three approaches. The first approach is a literature review [22]; for example, Boquiren et al [22] reviewed empirical studies that were published from 2000 to November 2013 to determine the primary domains underlying the patient satisfaction with the doctor. However, it is very hard to identify any new dimension by summarizing the literature, which is a major challenge for our fast-changing environment. The second approach is content analysis [24-26]; for example, Emmert et al [24] did a content analysis of 3000 randomly selected narrative comments from a German PRW and developed a theoretical categorization framework addressing physician-, staff-, and practice-related patient concerns. However, the manual nature of the content analysis makes it very hard to process large amounts of reviews, which is another challenge of the information age. The third approach is the algorithm or data-driven approach [9,20,21]; for example, Hao and Zhang [9] applied the topic extraction algorithm LDA to >500,000 textual reviews from >75,000 Chinese doctors across 4 major specialty areas to identify the dimensions inside the physician reviews. However, the output of the algorithm usually depends on the data, and many categories are hard to explain in medical practice. Some studies from other domains have also used the algorithm approach; for example, Guo et al [42] identified the key dimensions of customer service voiced by hotel visitors with a data mining approach LDA; they uncovered 19 controllable key dimensions that are important for managing hotels' interactions with visitors.

The three approaches have both advantages and disadvantages. In this study, we propose a new method that combines the three approaches mentioned above. We start with the literature review approach to form an initial taxonomy. Then, an algorithm-based data-driven approach is used to explore and find more

XSL•FO
RenderX

categories. Finally, a human annotation approach is used to finalize the taxonomy. This study demonstrates the plausibility and validity of the proposed mixed method.

Comparison With Prior Work on Patients' Interests Mined from Web-Based Reviews

Some researchers have manually coded and identified patients' interests from the Web-based reviews; for example, López et al [38] summarized two topic categories as global themes (which included the overall excellence, recommendation, negative sentiment, intent not to return, and professionalism) and specific factors (including interpersonal manner, technical competence, and system issues). Espinel et al [23] built a taxonomy for physician comments through an analysis of Web-based physician reviews, physician-related and system-related. In detail, physician-related topics included 2 subtopics, interpersonal style and technical skills and knowledge and preparation, whereas system-related topics included scheduling, wait time, parking, location, and cleanliness. Boquiren et al [22] revealed 5 broad domains underlying the patient satisfaction with the doctor—communication attributes, relational conduct, technical skill and knowledge, personal qualities, and availability and accessibility. Tymiński et al [25] identified patients' criteria for assessment of doctors—kindness and propriety, punctuality, communication with patients, condition and equipment of a doctor's office, length of the appointment, and cost of the medical advice. Davis and Hanauer [46] identified the key themes associated with positive and negative patient reviews. Themes that emerged from the high- and low-scoring reviews were similar in content but opposite in valence. Notably, physician-specific themes included temperament, knowledge and competency, physical examination, communication abilities, and mindfulness of cost. Practice-specific themes included scheduling, staff temperament, office cleanliness, waiting room, and insurance.

Some other researchers have applied the algorithm to extract patients' interests from the Web-based reviews automatically; for example, Hao and Zhang [9] applied LDA to >500,000 textual reviews of >75,000 Chinese doctors across 4 major specialty areas. They found the following important topics from the reviews: treatment effects, technical skills, appreciate the surgery result, story of treatment, story of surgery, bedside manner, story of registration, story of finding doctors, general appreciation, description of symptoms, and concern about children's health. In addition, Wallace et al [21] analyzed a corpus comprising nearly 60,000 such reviews with a state-of-the-art probabilistic model of text factorial LDA; they suggested three important topics of patients' interests as systems, technical, and interpersonal. There are also researchers who identified users' interests from Web-based health communities; for example, Lu et al [47] developed a new content analysis method using text mining techniques to determine hot topics of concern. They identified 5 significantly different health-related topics: symptom, examination, drug, procedure, and complication.

Conclusions

This study explores the internal dimensions of Web-based physician reviews, proposes an automatic classification algorithm based on labeled-LDA, and uses patient listening as an application to illustration the value of physician review mining. The identified taxonomy includes three high-level domains or categories and many subcategories or subtopics. The evaluation of the result of the proposed classification algorithm achieved impressive results with an *F*-measure of 0.816 on average with the highest *F*-measure for hypertension (0.904) and the lowest for infertility (0.682). The mining results indicate that symptoms are more often mentioned by patients with acute diseases, whereas CS, financing, and DAP are more often mentioned by patients with chronic diseases. Patients with mild diseases are more interested in ME, OP, PP, and symptoms. Meanwhile, patients with serious diseases are more interested in MC, MAP, financing, and DAP.

This study has some practical implications. First, this study provides an efficient and cost-effective way to analyze large amounts of physician reviews automatically. With the popularity of PRWs, the information overload has become a major concern for users. The taxonomy and method proposed in this study provide a convenient way to listen to patients, which is a very important step toward patient-centered care. In addition, the illustrated findings from physician review mining indicate that patients with different diseases and from different hospital levels might have different concerns that need to be addressed. Second, the taxonomy and algorithm proposed in this study also provide the bases for building decision aid tools that help patients make better decisions regarding the physician choice; for example, the decision aid tools might describe a physician with prominent tags such as good communication and high competence. Furthermore, the system can visualize different dimensions of a physician in a graph or compare multiple physicians across different dimensions in a matrix.

Although the taxonomy developed in this study occurred within a Chinese context, it is still generalizable to other countries and languages for the following two reasons. First, the initial taxonomy that serves as the starting point for this study was derived from the literature of Western countries. Therefore, the final taxonomy should also have a strong connection with those countries or languages. Second, China is a very large country. The economy in the eastern district is prosperous, whereas the western district is far behind. The intracountry diversity of China is quite high, rivaling or exceeding the intercountry differences of some continents (eg, Europe). Therefore, we believe that China is a good example that reflects the medical needs of both developed and developing countries.

This research also has three limitations. First, this study is descriptive in nature. Therefore, the correlation between important variables (eg, satisfaction and topics) cannot be revealed in this study. Second, the algorithm proposed in this study relies more or less on counting the prevalence of words rather than evaluating them positively or negatively. Some parts of the reviews may only reflect a mere description of an encounter. Often, people describe an encounter and in the end, give an evaluation of some specific aspects of that encounter. If the algorithm only counts frequencies, it may not get to the bottom of the real review's motivation in this respect. Third, the results of this study could be biased because of fake reviews because anyone visiting a physician can write a review on the

XSL•FO
RenderX

website, and the review process is anonymous. Although fake reviews are a common limitation to Web-based review studies, a fake review detection algorithm (eg, Yelp's fake review filter) can be applied in the future to increase study rigor.

Future research might include identifying hygienic and other motivating factors underlying for patient satisfaction. One avenue of research may be to explore a two-factor theory that states that certain factors cause user satisfaction, whereas a separate set of factors cause dissatisfaction. Because we can collect the user satisfaction data from PRWs, such an exploratory study is now practical in conjunction with additional sentiment analyses to determine a valence score for each dimension. The basic idea of the two-factor theory is that factors that lead to satisfaction or dissatisfaction are different. Some factors only relate to satisfaction, whereas others only relate to dissatisfaction. Understanding dissatisfying factors that demotivate and satisfying factors that motivate is important information for health care providers who want to improve user satisfaction in a cost-effective and patient-centered manner.

Acknowledgments

This research was supported by the National Natural Science Foundation of China grant number 71371005, 71471064, and 91646205.

Authors' Contributions

J Li and X Li were involved in the conception of the research and study protocol design. ML, X Liu, and J Liu executed the study and collected the data. All authors contributed to drafting the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Details for Labeled LDA.

[PDF File (Adobe PDF File), 108KB - jmir_v20i8e254_app1.pdf]

Multimedia Appendix 2

Word cloud visualization of topics (n=122,716).

[PNG File, 961KB - jmir_v20i8e254_app2.png]

Multimedia Appendix 3

The topic ratio differences across disease types, city levels and hospital levels.

[PDF File (Adobe PDF File), 44KB - jmir_v20i8e254_app3.pdf]

References

1. Yang H, Guo X, Wu T, Ju X. Exploring the effects of patient-generated and system-generated information on patients' online search, evaluation and decision. Electronic Commerce Research and Applications 2015 May;14(3):192-203. [doi: 10.1016/j.elerap.2015.04.001]
2. Grabner-Kräuter S, Waiguny MKJ. Insights into the impact of online physician reviews on patients' decision making: randomized experiment. J Med Internet Res 2015 Apr 09;17(4):e93 [FREE Full text] [doi: 10.2196/jmir.3991] [Medline: 25862516]
3. Mostaghimi A, Crotty BH, Landon BE. The availability and nature of physician information on the internet. J Gen Intern Med 2010 Nov;25(11):1152-1156 [FREE Full text] [doi: 10.1007/s11606-010-1425-7] [Medline: 20544300]
4. Healthgrades. URL: https://www.healthgrades.com/ [accessed 2018-05-14] [WebCite Cache ID 6zOlfqxNA]
5. Vitals. URL: https://www.vitals.com/ [accessed 2018-05-14] [WebCite Cache ID 6zOlpiH1u]
6. RateMDs. URL: https://www.ratemds.com/tx/spring/ [accessed 2018-05-14] [WebCite Cache ID 6zOlykETK]
7. Bickart B, Schindler R. Internet forums as influential sources of consumer information. Journal of Interactive Marketing 2001 Jan;15(3):31-40. [doi: 10.1002/dir.1014]
8. Kilaru AS, Meisel ZF, Paciotti B, Ha YP, Smith RJ, Ranard BL, et al. What do patients say about emergency departments in online reviews? A qualitative study. BMJ Qual Saf 2016 Jan;25(1):14-24. [doi: 10.1136/bmjqs-2015-004035] [Medline: 26208538]
9. Hao H, Zhang K. The Voice of Chinese Health Consumers: A Text Mining Approach to Web-Based Physician Reviews. J Med Internet Res 2016 May 10;18(5):e108 [FREE Full text] [doi: 10.2196/jmir.4430] [Medline: 27165558]

XSL•FO
RenderX

10. Hao H. The development of online doctor reviews in China: an analysis of the largest online doctor review website in China. J Med Internet Res 2015 Jun 01;17(6):e134 [FREE Full text] [doi: 10.2196/jmir.4365] [Medline: 26032933]
11. Li S, Feng B, Chen M, Bell R. Physician review websites: effects of the proportion and position of negative reviews on readers' willingness to choose the doctor. J Health Commun 2015 Apr;20(4):453-461. [doi: 10.1080/10810730.2014.977467] [Medline: 25749406]
12. Ghose A, Ipeirotis P, Li B. Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content. Marketing Science 2012 May;31(3):493-520. [doi: 10.1287/mksc.1110.0700]
13. Daskivich TJ, Houman J, Fuller G, Black JT, Kim HL, Spiegel B. Online physician ratings fail to predict actual performance on measures of quality, value, and peer review. J Am Med Inform Assoc 2018 Apr 01;25(4):401-407. [doi: 10.1093/jamia/ocx083] [Medline: 29025145]
14. Gao GG, Greenwood BN, Agarwal R, McCullough, PhD J. Vocal Minority and Silent Majority: How Do Online Ratings Reflect Population Perceptions of Quality? MIS Quarterly 2015;39(3):565-589. [doi: 10.2139/ssrn.2629837]
15. Hao H, Zhang K, Wang W, Gao G. A tale of two countries: International comparison of online doctor reviews between China and the United States. Int J Med Inform 2017 Dec;99:37-44. [doi: 10.1016/j.ijmedinf.2016.12.007] [Medline: 28118920]
16. Lockie M, Waiguny M, Grabner-Kräuter S. How style, information depth and textual characteristics influence the usefulness of general practitioners' reviews. Australasian Marketing Journal (AMJ) 2015 Aug;23(3):168-178. [doi: 10.1016/j.ausmj.2015.01.005]
17. Maslow A. A theory of human motivation. Psychological Review 1943;50(4):370-396.
18. The Good Doctor. URL: http://www.haodf.com/ [accessed 2018-05-14] [WebCite Cache ID 6zOlAkrsR]
19. Wang Y, Wei L, Liu J, Li S, Wang Q. Comparison of Cancer Incidence between China and the USA. Cancer Biol Med 2012 Jun;9(2):128-132 [FREE Full text] [doi: 10.3969/j.issn.2095-3941.2012.02.009] [Medline: 23691468]
20. Sharma RD, Tripathi S, Sahu SK, Mittal S, Anand A. Predicting Online Doctor Ratings from User Reviews Using Convolutional Neural Networks. International Journal of Machine Learning and Computing 2016;6(2):149-154.
21. Wallace BC, Paul MJ, Sarkar U, Trikalinos TA, Dredze M. A large-scale quantitative analysis of latent factors and sentiment in online doctor reviews. J Am Med Inform Assoc 2014;21(6):1098-1103 [FREE Full text] [doi: 10.1136/amiajnl-2014-002711] [Medline: 24918109]
22. Boquiren V, Hack T, Beaver K, Williamson S. What do measures of patient satisfaction with the doctor tell us? Patient Educ Couns 2015 Jun 10;98(12):1465-1473 (forthcoming). [doi: 10.1016/j.pec.2015.05.020] [Medline: 26111500]
23. Espinel AG, Shah RK, Beach MC, Boss EF. What parents say about their child's surgeon: parent-reported experiences with pediatric surgical physicians. JAMA Otolaryngol Head Neck Surg 2014 May;140(5):397-402. [doi: 10.1001/jamaoto.2014.102] [Medline: 24651973]
24. Emmert M, Meier F, Heider A, Dürr C, Sander U. What do patients say about their physicians? an analysis of 3000 narrative comments posted on a German physician rating website. Health Policy 2014 Oct;118(1):66-73. [doi: 10.1016/j.healthpol.2014.04.015] [Medline: 24836021]
25. Tymi?ski R, Walczewski M, Wieczorek M. Identifying patients' criteria for assessment of doctors on Polish physician rating websites. JMedSci 2016;84(3):189-196.
26. Bardach N, Lyndon A, Asteria-Peñaloza R, Goldman L, Lin G, Dudley R. From the closest observers of patient care: a thematic analysis of online narrative reviews of hospitals. BMJ Qual Saf 2016 Dec;25(11):889-897 [FREE Full text] [doi: 10.1136/bmjqs-2015-004515] [Medline: 26677215]
27. HanLP. 2018 URL: http://hanlp.linrunsoft.com/ [accessed 2018-05-14] [WebCite Cache ID 6zOm6Wz0x]
28. HIT IR-Lab Tongyici Cilin. 2018 URL: http://ir.hit.edu.cn/demo/ltp/Sharing_Plan.htm [accessed 2018-05-14] [WebCite Cache ID 6zOmS3sM5]
29. Sekine S, Nobata C. Definition, dictionaries and tagger for Extended Named Entity Hierarchy. 2004 Presented at: Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004); May 26-28, 2004; Lisbon, Portugal p. 1977-1980.
30. Lu X, Zheng B, Velivelli A, Zhai C. Enhancing text categorization with semantic-enriched representation and training data augmentation. J Am Med Inform Assoc 2006;13(5):526-535 [FREE Full text] [doi: 10.1197/jamia.M2051] [Medline: 16799127]
31. Blei D, Ng A, Jordan M. Latent Dirichlet Allocation. Journal of Machine Learning Research 2003;3:993-1022.
32. Han D, Wang S, Jiang C, Jiang X, Kim H, Sun J, et al. Trends in biomedical informatics: automated topic analysis of JAMIA articles. J Am Med Inform Assoc 2015 Nov;22(6):1153-1163 [FREE Full text] [doi: 10.1093/jamia/ocv157] [Medline: 26555018]
33. Perotte A, Ranganath R, Hirsch JS, Blei D, Elhadad N. Risk prediction for chronic kidney disease progression using heterogeneous electronic health record data and time series analysis. J Am Med Inform Assoc 2015 Jul;22(4):872-880 [FREE Full text] [doi: 10.1093/jamia/ocv024] [Medline: 25896647]
34. Roberts K, Demner-Fushman D. Interactive use of online health resources: a comparison of consumer and professional questions. J Am Med Inform Assoc 2016 Dec;23(4):802-811 [FREE Full text] [doi: 10.1093/jamia/ocw024] [Medline: 27147494]

XSL•FO
RenderX

35. Bao Y, Datta A. Simultaneously Discovering and Quantifying Risk Types from Textual Risk Disclosures. Management Science 2014;60(6):1371-1391.
36. Azzopardi L, Girolami M, Van Risjbergen K. Investigating the relationship between language model perplexity and IR precision-recall measures. 2003 Presented at: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval (SIGIR '03); July 28 - August 01, 2003; Toronto, Canada p. 369-370.
37. Borzekowski D, Schenk S, Wilson J, Peebles R. e-Ana and e-Mia: A content analysis of pro-eating disorder Web sites. Am J Public Health 2010;100(8):1526-1534.
38. López A, Detz A, Ratanawongsa N, Sarkar U. What patients say about their doctors online: a qualitative content analysis. J Gen Intern Med 2012;27(6):685-692.
39. Trochim W, Donnelly J. Research methods knowledge base, 3rd Edition. United States: Atomic Dog; 2006.
40. Ramage D, Hall D, Nallapati R, Manning C. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. Stroudsburg, PA, USA: Association for Computational Linguistics; 2009 Presented at: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP '09); August 06 - 07, 2009; Singapore p. 248-256.
41. Salton G, Harman D. Information retrieval. In: Encyclopedia of Computer Science. Chichester, UK: John Wiley and Sons Ltd; 2003:858-863.
42. Guo Y, Barnes SJ, Jia Q. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichllocation. Tourism Management 2017;59:467-483.
43. Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. J Exp Psychol Gen 2012 Feb;141(1):2-18. [doi: 10.1037/a0024338] [Medline: 21823805]
44. Sawilowsky S. New effect size rules of thumb. Journal of Modern Applied Statistical Methods 2009;8(2):467-474.
45. Duckett J, Hunt K, Munro N, Sutton M. Does distrust in providers affect health-care utilization in China? Health Policy Plan 2016 Oct;31(8):1001-1009 [FREE Full text] [doi: 10.1093/heapol/czw024] [Medline: 27117483]
46. Davis MM, Hanauer DA. Online Measures of Physician Performance: Moving Beyond Aggregation to Integration. JAMA 2017 Dec 13;317(22):2325-2326. [doi: 10.1001/jama.2017.3979] [Medline: 28609521]
47. Lu Y, Wu Y, Liu J, Li J, Zhang P. Understanding health care social media use from different stakeholder perspectives: a content analysis of an online health community. J Med Internet Res 2017;19(4):e109.

Abbreviations

CS: communication skills
DAP: diagnosis and pathogenesis
LDA: Latent Dirichlet Allocation
MAP: medical advice and prescription
MC: medical competence
ME: medical ethics
OP: operation process
PP: patient profile
PRW: physician review websites
RQ: research question

Edited by G Eysenbach; submitted 30.08.17; peer-reviewed by U Sander, M Lavin, S Bidmon; comments to author 02.12.17; revised version received 08.02.18; accepted 21.06.18; published 16.08.18.

Please cite as:
Li J, Liu M, Li X, Liu X, Liu J
Developing Embedded Taxonomy and Mining Patients' Interests From Web-Based Physician Reviews: Mixed-Methods Approach
J Med Internet Res 2018;20(8):e254
URL: http://www.jmir.org/2018/8/e254/
doi:10.2196/jmir.8868
PMID:30115610

©Jia Li, Minghui Liu, Xiaojun Li, Xuan Liu, Jingfang Liu. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 16.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete

XSL•FO
RenderX

bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Proposal

Internet of Things Buttons for Real-Time Notifications in Hospital Operations: Proposal for Hospital Implementation

Peter Ray Chai[1,2], MD, MS; Haipeng Zhang[3], DO, MMSc; Christopher W Baugh[3], MD, MBA; Guruprasad D Jambaulikar[3], MBBS, MPH; Jonathan C McCabe[3], BSN, RN; Janet M Gorman[3], RN; Edward W Boyer[1,2], MD, PhD; Adam Landman[4], MD

[1]Division of Medical Toxicology, Department of Emergency Medicine, Brigham and Women's Hospital, Boston, MA, United States

[2]The Fenway Institute, Boston, MA, United States

[3]Innovation Hub, Brigham and Women's Hospital, Boston, MA, United States

[4]Partners Information Systems, Sommerville, MA, United States

Corresponding Author:
Peter Ray Chai, MD, MS
Division of Medical Toxicology
Department of Emergency Medicine
Brigham and Women's Hospital
75 Francis street
Boston, MA,
United States
Phone: 1 6177325640
Email: pchai@bwh.harvard.edu

Abstract

Background: Hospital staff frequently performs the same process hundreds to thousands of times a day. Customizable Internet of Things buttons are small, wirelessly-enabled devices that trigger specific actions with the press of an integrated button and have the potential to automate some of these repetitive tasks. In addition, IoT buttons generate logs of triggered events that can be used for future process improvements. Although Internet of Things buttons have seen some success as consumer products, little has been reported on their application in hospital systems.

Objective: We discuss potential hospital applications categorized by the intended user group (patient or hospital staff). In addition, we examine key technological considerations, including network connectivity, security, and button management systems.

Methods: In order to meaningfully deploy Internet of Things buttons in a hospital system, we propose an implementation framework grounded in the Plan-Do-Study-Act method.

Results: We plan to deploy Internet of Things buttons within our hospital system to deliver real-time notifications in public-facing tasks such as restroom cleanliness and critical supply restocking. We expect results from this pilot in the next year.

Conclusions: Overall, Internet of Things buttons have significant promise; future rigorous evaluations are needed to determine the impact of Internet of Things buttons in real-world health care settings.

(J Med Internet Res 2018;20(8):e251) doi:10.2196/jmir.9454

KEYWORDS

Internet of Things; operations; hospital systems; health care

Introduction

Background

Simple repetitive tasks when done manually are often time consuming and can be overlooked or simply forgotten. In a hospital setting, staff perform multiple parallel processes hundreds to thousands of times a day that often require minimal margin of error [1]. Execution of these tasks may be interdependent; without a notification or completed process in an operating cascade, final completion of a task may be delayed or the task may remain incomplete. One possible way to improve hospital staff efficiency and reduce the chance of error with these repetitive tasks is to leverage Internet of Things (IoT) devices. IoT is the “interconnection via the internet of computing devices embedded in everyday objects, enabling them to send

XSL•FO
RenderX

and receive data" [2]. An IoT-enabled button can send automatic, reliable, just-in-time notifications or trigger one or more tasks when pressed.

A well-known example of IoT buttons is Amazon's "Dash" button that enables consumers to quickly reorder specific products through Amazon. Amazon and product manufacturers expect that consumers will place Dash buttons at the location where products are used; when a product's supply is depleted, a simple press of the Dash button orders a refill. For example, an Amazon Dash button for laundry detergent might be attached to the consumer's washing machine. When the laundry detergent is running low, the consumer presses the Dash button and laundry detergent is automatically reordered.

IoT buttons can be configured to perform a wide range of actions extending beyond internet product ordering [3,4]. In the hospital, IoT buttons may be a cost-efficient, intuitive, and scalable method to automate repetitive, commonplace hospital tasks and provide real-time insight into daily hospital operations. For example, IoT buttons can be configured to deliver messages to housekeeping, nurse managers, and administrators upon patient discharge, coordinating an efficient bed turnover process while recording each step of the process. These recorded data can then be analyzed to identify further process improvement opportunities.

Little data and almost no protocols exist regarding the real-world usage and operationalization of IoT buttons in hospitals. In this paper, we describe potential applications of IoT buttons to streamline and evaluate hospital operations as well as describe technological considerations, such as data security requirements. We also propose a framework that health care systems can use for IoT button deployment.

Internet of Things Buttons

IoT buttons are small and unobtrusive (approximately the size of a stick of gum; Figure 1) and are available through a variety of commercial vendors using a range of technologies. IoT buttons can be used for very specific purposes, such as a button located in a patient's room that, when pressed, indicates that room cleaning is needed. Alternatively, buttons can have more general purposes, such as a patient call button that summons assistance, but does not specify the precise reason for the assistance.

Pressing the IoT button sends a preprogrammed message through a network (often wireless) to a server that can send customizable notifications. Notifications might generate a standard short message service text message, an email, or a page (Figure 2). Additionally, notifications can be extended almost infinitely by calling application programming interfaces (APIs), software methods that allow computer systems to exchange information.

A single IoT button may also have the ability to perform multiple, distinct actions through different types of button presses (single button press, double button press, and long press). For example, a single press of the button may send a notification via email, a double press can send a different message to a pager, whereas a long press may log an event into the electronic medical record (EMR) through an API. Using the patient discharge and room cleaning application discussed earlier, a single button press may notify housekeeping that the room needs to be cleaned, whereas a double press may record when housekeeping has completed the task. Finally, a long press can call the EMR's API to update the room's status.

The IoT buttons can also provide feedback to users in real time. One IoT button vendor embeds a multicolor indicator light on the device. The light flashes white to indicate that a button press was detected and then changes to green to signal successful delivery of the notification. If a notification is not delivered, the light flashes red to alert the user to an error. In order to prevent inadvertent or deliberate repetitive presses, buttons should have the ability to lock out after a defined number of presses or period of time.

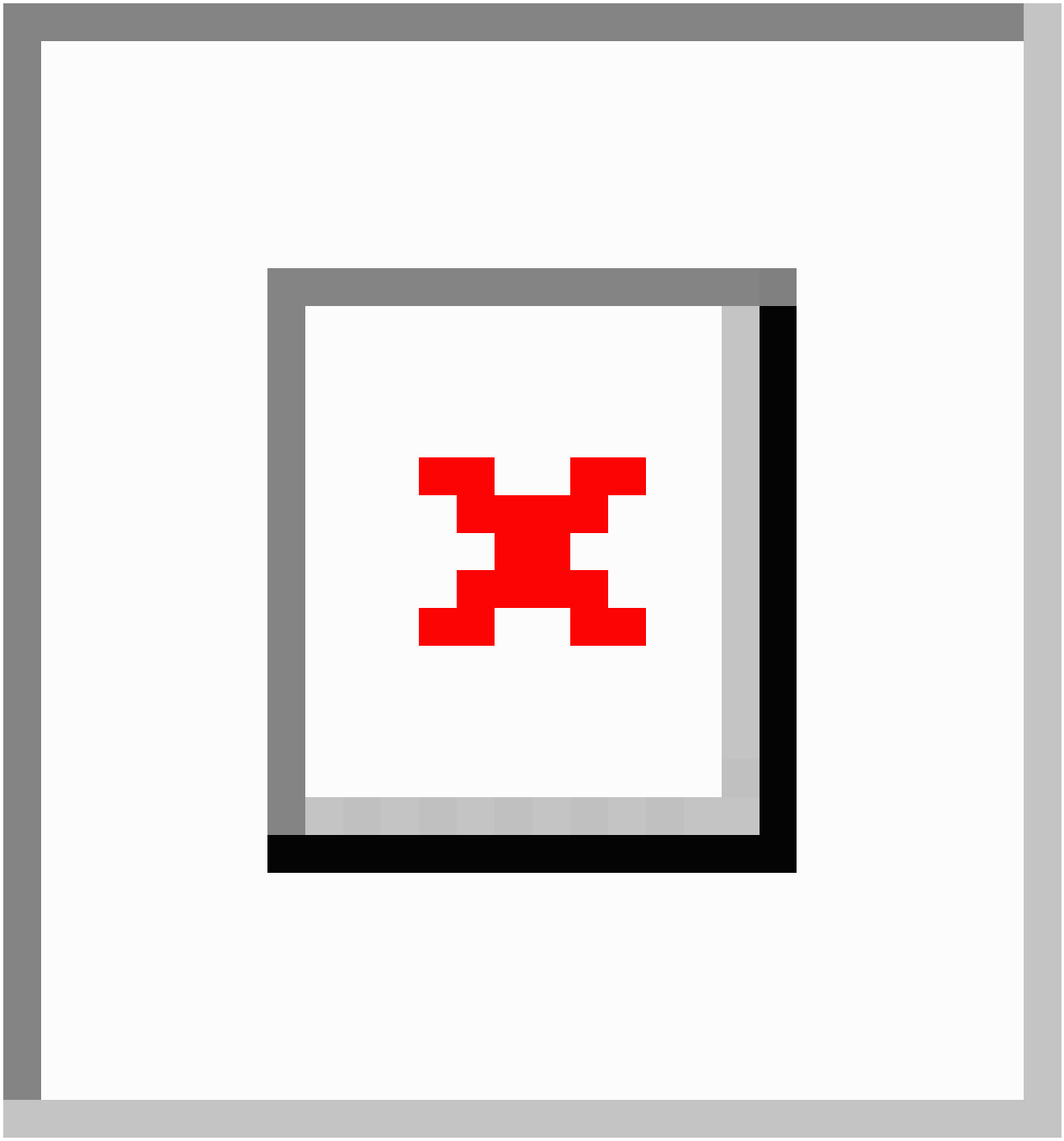
Button presses can be logged in a database and subsequently used for data visualization and analytics. The database could record which button was pressed, type of press, location of the button, and date or time of action. Off-the-shelf analytics tools can facilitate data summarization and visualization. Further, the data could be used for more complex analytics. Returning to the earlier room cleaning example, the stored data could be analyzed to determine the average room cleaning time by comparing the time of the initial room cleaning request (single button press) with the time of the housekeeper response (double button press). These analyses may in turn help influence future staffing decisions, such as the number of staff, location of work, and task schedules.

Existing Hospital Notification Systems

Many different notification systems currently exist in health care systems (Table 1). These notification systems may be devices like a patient call button, quick response code, or technological measures like a hospital paging portal. The widespread conversion of hospital medical records into EMRs has also fueled the development of EMR rules and dashboards to manage EMR-based notifications. IoT buttons are a highly adaptable notification system that can potentially supplement or replace other existing hospital notification systems. Buttons can be placed in various environments and programmed to deliver custom messages in response to specific tasks. Although the IoT button is a physical device, activating the button can trigger a cascade of tasks in other systems like the medical record or a Web-based paging portal. Additionally, IoT buttons can measure their own usage through usage logs.

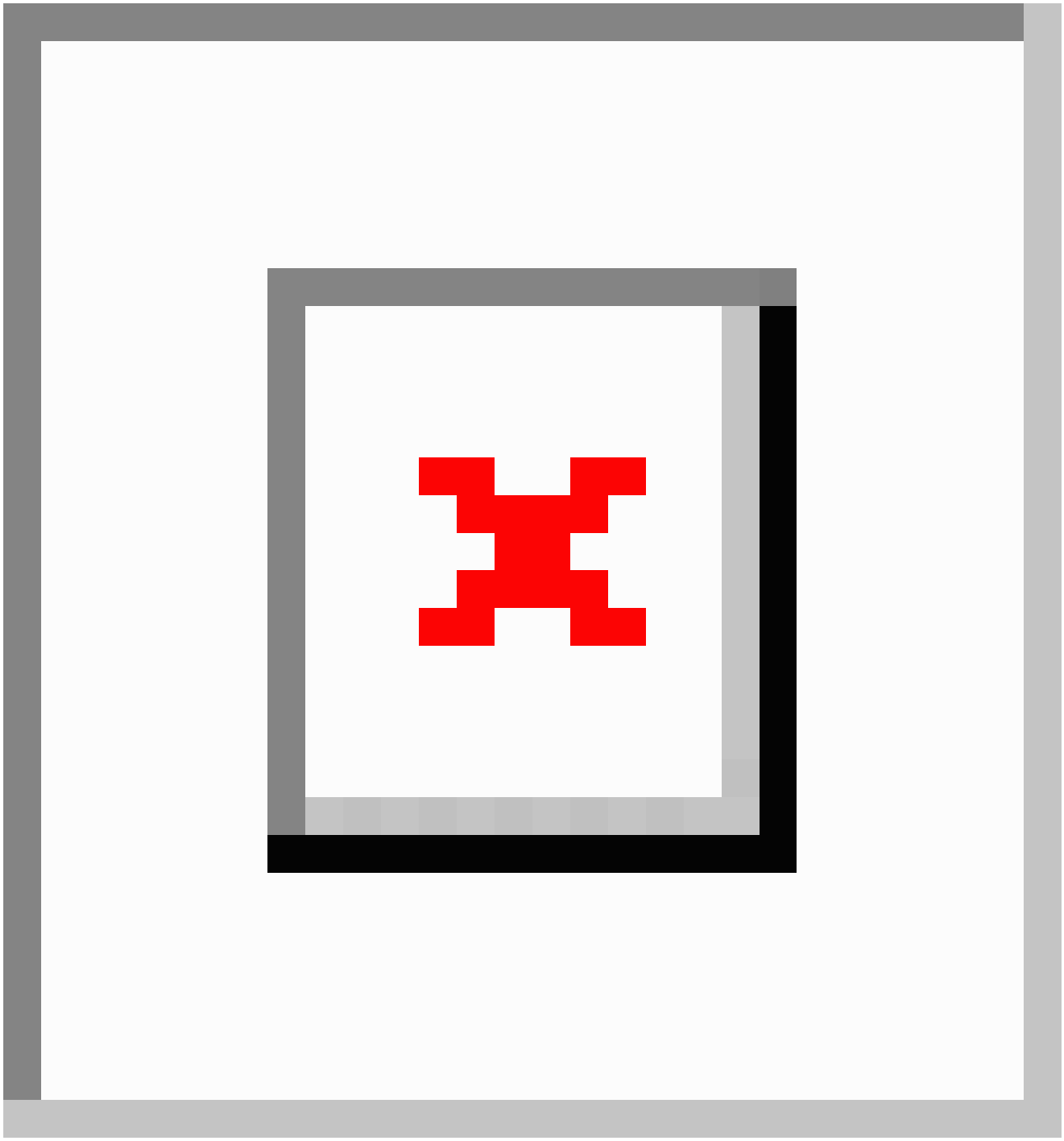
XSL•FO
RenderX

Figure 1. An Internet of Things (IoT) button. A United States quarter is pictured for scale.



XSL•FO
RenderX

Figure 2. Schematic of the process flow of an Internet of Things (IoT) button press. API: application programming interface.



XSL•FO
RenderX

Table 1. Advantages and challenges of existing hospital notification systems and Internet of Things (IoT) buttons.

Notification system	Current application	Advantages	Challenges
QR code[a] readers	Notification systems	• Universal code that can be accessed through mobile phones or dedicated barcode scanners	• Multistep process to activate • Requires a mobile phone with QR reading capability or barcode scanner
Patient call button	Patient-facing notification system to call nursing staff	• Recognizable device with simple user interface	• No context for notifications
EMR[b]-based notification rules	Signaling completed tasks based on EMR changes	• Improved process flows via EMR events • Can be applied quickly through a hospital	• Task must be based on EMR change • Each new application requires programming
Web-based paging system	Sending custom notifications to providers, hospital staff	• Web portal allows for access anywhere	• Requires accessing paging system to deliver each notification
IoT Buttons	Patient or staff facing	• Notification delivered with push of a button • Notifications can be simple or complex actions	• Requires installation of buttons • Security and privacy issues • Programming or configuration of buttons required

[a]QR code: quick response code.

[b]EMR: electronic medical record.

Methods

Potential Applications in Hospital Operations

Multiple potential applications for IoT buttons exist in hospital operations. We classified applications into two categories: patient and hospital staff applications, based on the primary intended user group (Textbox 1).

For patients, an IoT button may be a potential replacement for the traditional patient call button. A small, mobile wireless IoT button, or even an IoT button affixed to a rail of a hospital bed, that is given to a patient can be programmed to call for help. While a traditional hospital call button conveys a single piece of information (ie, the patient needs assistance), an IoT button can communicate a greater breadth of information. For example, a patient may use a single button press to call for help, but a responding nurse could then provide a long press to indicate that the call has been addressed or provide two presses to request additional help. The ease of button use may improve the reliability and validity of patient reported outcomes [5,6].

Hospital staff applications of IoT buttons can help focus not only on streamlining notifications but also on gathering real-time data that can be used for workflow quality improvement. For example, an IoT button on a linen cart can help hospital staff notify housekeeping when linens need to be restocked. Button presses for restock requests can be aggregated and analyzed to determine the patterns of usage. These data could be used to recommend and justify staffing changes to fulfill supply requests and subsequently to evaluate the response to such staffing changes. Similarly, IoT buttons can be used to analyze key operational chokepoints in patient flow. Button presses may flag inpatients who are ready for discharge on morning rounds, sending notifications to key individuals like case management, social work, and housekeeping that facilitate discharge and bed turnover. Similarly, IoT buttons may be leveraged to flag patients in the emergency department who are ready to be admitted, delivering notifications to responding clinicians, bed control specialists, and transport, thereby initiating multiple cascades of tasks required to admit a patient to the hospital.

XSL•FO
RenderX

Textbox 1. Potential applications of Internet of Things buttons to improve hospital operations classified by primary user group.

Patient applications

- Restroom cleaning alerts
- Use as a call button to:
  - Contact clinical teams
  - Report distress
  - Contact other hospital staff including research teams

Hospital staff applications

- Supply chain restock
- Notifications of critical orders or events, such as flagging patients for discharge
- Notifications to specific hospital services, such as respiratory, phlebotomy, and information technology support
- Optimizing hospital bed turnover
- Identifying potential research participants in the hospital
- Initiating a bed request for hospital admission

Results

We are planning several pilots to evaluate the use of IoT buttons in a hospital system and to help further identify best practices for deploying IoT buttons. Currently, we are piloting IoT buttons in hospital public restrooms to assess and improve their cleanliness. We are planning to use IoT buttons to understand demand and patterns of restocking for patient equipment like stretchers and wheelchairs in the emergency department. These interventions will also allow for real-time assessment of response time from staff. We anticipate initial results in the next year regarding the feasibility and acceptability of these pilots.

Discussion

Technological Considerations

Connectivity

Ensuring secure and reliable network connectivity is an important consideration in IoT button deployments. Unlike streaming applications, IoT buttons do not require continuous connectivity as they only need to transmit data upon button press. By connecting IoT buttons to the network on demand, battery life can be conserved. One vendor uses this strategy to achieve a battery lifetime of 2 years or 2000 clicks [7].

Deploying a large number of IoT buttons in a hospital environment has the potential to overwhelm network capacity if the proper engineering expertise is not consulted. Therefore, an initial, limited deployment should be considered to test the feasibility of IoT button use as well as the network bandwidth requirements. Additionally, the use of dedicated networks for IoT buttons may help ensure efficient IoT button performance and also minimize the chance of unintended consequences, such as network disruptions, to the primary hospital network.

Button Management Systems

Some vendors offer button management systems (BMS), Web-based administrator interfaces, or consoles for IoT button programming and monitoring. From a hospital perspective, important administrator features include audit trails, battery life monitoring, and the ability to group buttons by use case. Audit trails or log data are particularly important for quality assurance, whereas early warning systems, which alert an administrator to low battery life, ensure button functionality and availability. Similar to how website content management systems enable individuals to control content in specific website sections, the BMS should include controlled access to groups of buttons by characteristics such as department, physical location, or use case. BMS should also be flexible and intuitive so that users with limited technical proficiency can modify button configuration, whereas more sophisticated users can customize button functionality using programming languages. In addition, BMS support for batch programming will facilitate rapid deployment of a large number of buttons with identical functionality.

Privacy and Security

When considering the use of IoT buttons, hospital systems should understand the potential privacy and security risks. Privacy breaches can occur when unencrypted button messages containing protected health information (PHI) are intercepted. Network security can be compromised when IoT buttons are used as an entry point into hospital networks or used as a distributed denial of service (DDoS) attack. Understanding these risks and creating strategies to effectively mitigate them are central to safely deploying IoT buttons (Table 2).

Patient privacy may be compromised if button notifications contain PHI. For example, if an IoT button is configured to send the following message “John Smith in Room 300 needs help,” interception of these data could reveal not only the presence of a specific patient within a hospital but also pinpoint the patient’s location. To protect against breaches in patient privacy,

XSL•FO
RenderX

messages should be carefully constructed to avoid PHI. If PHI or other sensitive information needs to be transmitted, buttons should utilize modern encryption protocols.

Open firewall ports, used to deliver button notifications, may provide a portal to enter a hospital network and steal critical health information or conduct malicious attacks against hospital infrastructure. To minimize the chance of these attacks, IoT buttons can be programmed to only connect briefly to a hospital network, while a notification is being sent, minimizing the time a critical firewall port is open. Hospitals may also consider isolating IoT buttons on an independent network to protect hospital infrastructure from infiltration.

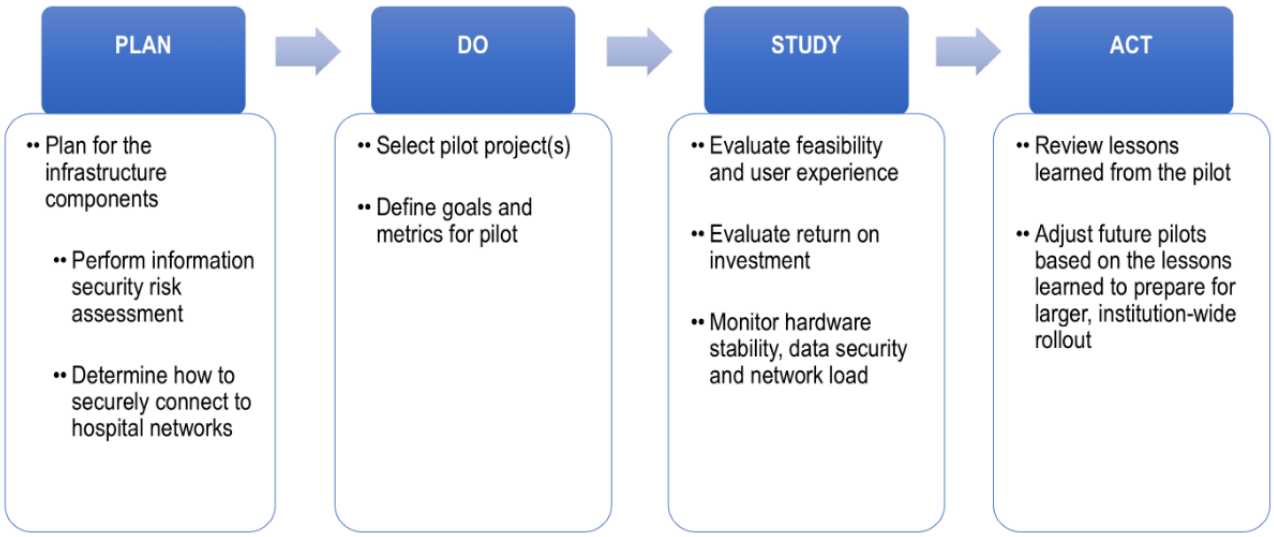
In order to transmit sensitive notifications containing PHI, IoT buttons should securely connect to an encrypted wireless network. Many IoT buttons support the Wi-Fi-Protected Access 2 (WPA2) mechanism, which requires a single, preshared password to connect to the wireless network. While the WPA2 mechanism conveys some protection, a hacker who learns a single preshared key can compromise the entire system, leading to reprogramming of IoT button functions, or disabling an IoT button network. IoT buttons should also support Wi-Fi-Protected Access-enterprise (WPA-enterprise) encryption system that requires a user to enter a unique username and password to log into the network, providing an additional layer of security necessary in networks that transmit confidential information. This way, even if the hackers learn the password of one IoT button, they cannot compromise the entire system. Another option that can reduce the chance of malicious activities is using IoT buttons with an integrated cellular network, bypassing the need to connect the IoT button to an institution's corporate network.

Malicious users could also conduct DDoS attacks by sending rapid, high-volume button presses from one or more IoT buttons in an attempt to overwhelm the hospital network [8]. As noted previously, buttons can be programmed to lock out for a period of time after each button press, which can help prevent DDoS attacks. In addition, button availability can be limited to appropriate users. For instance, buttons that communicate sensitive information should only be accessible to authorized hospital staff.

Table 2. Privacy and security considerations for Internet of Things (IoT) buttons.

Potential concerns	Potential solutions
Privacy and data breach	• Communicate deidentified data • Use encryption • Disallow continuous network connection and data transfer • Enable IoT buttons to communicate via cellular networks to avoid integration with hospital networks
Theft	• Strategic placement • Secure installation
Distributed denial of service (DDoS) attacks	• Lockout times on buttons to prevent DDoS based on number of button presses • IoT buttons placed on separate network
Failure	• Staggered adoption with careful testing of failure rates • Initial use in conjunction with existing notification methods

Figure 3. A proposed framework to deploy and evaluate the impact of Internet of Things (IoT) buttons in a hospital.



XSL•FO
RenderX

A Framework for Internet of Things Button Deployment

For health care organizations that seek to deploy IoT buttons across a large hospital, we recommend the following steps grounded in the Plan-Do-Study-Act (PDSA) method (Figure 3) [9,10]. PDSA is often used to accelerate quality improvement initiatives to rapidly test changes by planning them, implementing them, observing the results of the intervention, and iterating the changes based on what is learned [11]:

- Plan: The first step in an IoT button implementation is to plan for the infrastructure components. Since the buttons will require network connectivity, hospital information security officers should be engaged to perform a risk assessment of the technology and mitigate any high priority risks identified.
- Do: The team should select a limited, yet important, task that would benefit from a brief IoT button pilot.
- Study: The pilot should be evaluated to assess feasibility, user experience with and usage of the IoT buttons, and return on investment. In addition, hardware stability, data security and quality, and network load should be monitored during the pilot.
- Act: Lessons learned from the pilot, including technical, workflow, and other components of the sociotechnical model for health information technology, should be carefully reviewed [12]. To prepare for larger roll-outs, IoT button processes and protocols should be adjusted based on these lessons learned.

Limitations of Internet of Things Button Deployment

While there are many exciting applications for IoT buttons in the hospital setting, limitations also exist. First, not all hospital tasks are well suited for an IoT button intervention. Further, rapid, widespread deployments may lead to "button fatigue" as users are confronted with a bewildering array of buttons [13,14]. Therefore, a dedicated governance process, including the project team, information technology staff, and institutional leaders, is essential to triage new button requests for appropriateness and to help reduce risk of button fatigue. Second, it is important to understand button reliability and impact on technical infrastructure. The reliability of IoT buttons must be understood to ensure that buttons can be safely used for desired tasks. For example, a patient call button must be highly reliable (and may even require US Food and Drug Administration review). Plans must also be developed to ensure adequate network bandwidth to support the desired number of buttons. Third, the presence of "false presses" where a user inadvertently presses an IoT button may still overwhelm an IoT button system. Although timed lockout periods may mitigate the transmission of false presses, refinement of a protocol to detect and manage false IoT button presses is still needed. Fourth, the ethics and privacy implications of IoT buttons remain to be explored. Depending on the method in which IoT buttons are deployed, employers may be able to discover and better understand performance metrics of specific employees (for example, knowing how fast a nurse responds to a patient's IoT button press when used as a call button). Finally, IoT buttons are physical devices that may be lost, damaged, or stolen; however, these shortcomings may be minimized by changing how the buttons are secured or where the buttons are located or using completely software-based buttons [15].

Conclusion

IoT buttons may be a valuable tool to help optimize hospital operations and communication for a variety of use cases for both patients and staff. Key technical considerations for a successful deployment include ensuring appropriate network connectivity, selecting a product with a robust button management system, and carefully considering configuration to minimize privacy and security risks. The PDSA framework may guide hospitals starting with a small pilot, iteratively refining the process and, eventually, scaling to the entire organization. IoT buttons have significant promise, outweighing minor limitations, but need to be tested in real-world health care environments and rigorously evaluated to determine their impact.

Acknowledgments

This work was supported by the Brigham and Women's Hospital Digital Innovation Hub (iHub). EWB is supported by NIH K24DA037109; PRC is supported by NIH K23DA044874. We thank Visbyl (Germantown, MD) for providing IoT button hardware, software, and support for us to pilot at Brigham and Women's Hospital. Visybl was not involved in drafting or reviewing of this manuscript.

Conflicts of Interest

None declared.

References

1. Athanassopoulos A, Gounaris C. Assessing the technical and allocative efficiency of hospital operations in Greece and its resource allocation implications. Eur J Oper Res 2001 Jan 01;133(2):416-431. [doi: 10.1016/S0377-2217(00)00180-6]
2. Xia F, Yang LT, Wang L, Vinel A. Internet of Things. Int J Commun Syst 2012 Sep;25(9):1101-1102. [doi: 10.1002/dac.2417]
3. Amazon. AWS IoT Button Overview - Cloud Programmable Dash Button URL: https://aws.amazon.com/iotbutton/ [accessed 2018-07-23] [WebCite Cache ID 717zoiEFP]
4. Smith JW. The Uber-All Economy of the Future. The Independent Review Oakland 2016;20(3):390.

XSL•FO
RenderX

5. Lavallee D, Chenok K, Love R, Petersen C, Holve E, Segal C, et al. Incorporating Patient-Reported Outcomes Into Health Care To Engage Patients And Enhance Care. Health Aff (Millwood) 2016 Apr;35(4):575-582. [doi: 10.1377/hlthaff.2015.1362] [Medline: 27044954]
6. McCarthy MI. Putting the Patient in Patient Reported Outcomes: A Robust Methodology for Health Outcomes Assessment. Health Econ 2014 Oct 10;24(12):1588-1603. [doi: 10.1002/hec.3113] [Medline: 25303748]
7. Visybl. Visybl goButton. 2018. Visybl goButton URL: https://www.gobutton.io/ [accessed 2018-07-18] [WebCite Cache ID 710NgpVSr]
8. Sonar K, Upadhyay H. A Survey: DDOS Attack on Internet of Things. International Journal of Engineering Research and Development 2014 Nov 01;10(11):63. [doi: 10.1109/SIU.2015.7129976]
9. Bauchner H, Berwick D, Fontanarosa P. Innovations in Health Care Delivery and the Future of Medicine. JAMA 2016 Jan 05;315(1):30-31. [doi: 10.1001/jama.2015.17452] [Medline: 26746453]
10. Berwick DM. Developing and testing changes in delivery of care. Ann Intern Med 1998 Apr 15;128(8):651-656. [Medline: 9537939]
11. Agency for Healthcare Quality and Research. 2013 Jan 01. Plan-Do-Study-Act (PDSA) Cycle Internet URL: https://innovations.ahrq.gov/qualitytools/plan-do-study-act-pdsa-cycle [accessed 2018-07-17] [WebCite Cache ID 70yiVs3ip]
12. Sittig DF, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. Qual Saf Health Care 2010 Oct;19 Suppl 3:i68-i74 [FREE Full text] [doi: 10.1136/qshc.2010.042085] [Medline: 20959322]
13. van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. J Am Med Inform Assoc 2006 Mar;13(2):138-147 [FREE Full text] [doi: 10.1197/jamia.M1809] [Medline: 16357358]
14. Carspecken CW, Sharek PJ, Longhurst C, Pageler NM. A clinical case of electronic health record drug alert fatigue: consequences for patient outcome. Pediatrics 2013 Jun;131(6):e1970-e1973 [FREE Full text] [doi: 10.1542/peds.2012-3252] [Medline: 23713099]
15. Amazon Web Services. 2018. Your Dash Buttons: Shortcuts to shop your favorite products URL: https://www.amazon.com/ddb/learn-more [accessed 2018-07-18] [WebCite Cache ID 710Nll9gd]

Abbreviations

API: application programming interface
BMS: button management systems
DDoS: distributed denial of service
EMR: electronic medical record
IoT: Internet of Things
PDSA: Plan-Do-Study-Act
PHI: protected health information
QR code: quick response code
WPA2: Wi-Fi-Protected Access 2

Edited by G Eysenbach; submitted 08.02.18; peer-reviewed by S Carreiro, JH Lee, A Kotlo, B Arnoldussen; comments to author 24.03.18; revised version received 11.06.18; accepted 18.06.18; published 10.08.18.

Please cite as:
Chai PR, Zhang H, Baugh CW, Jambaulikar GD, McCabe JC, Gorman JM, Boyer EW, Landman A
Internet of Things Buttons for Real-Time Notifications in Hospital Operations: Proposal for Hospital Implementation
J Med Internet Res 2018;20(8):e251
URL: http://www.jmir.org/2018/8/e251/
doi:10.2196/jmir.9454
PMID:30097420

©Peter Ray Chai, Haipeng Zhang, Christopher W Baugh, Guruprasad D Jambaulikar, Jonathan C McCabe, Janet M Gorman, Edward W Boyer, Adam Landman. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 10.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Consumers Turning to the Internet Pharmacy Market: Cross-Sectional Study on the Frequency and Attitudes of Hungarian Patients Purchasing Medications Online

András Fittler[1], PharmD, PhD; Róbert György Vida[1], PharmD; Mátyás Káplár[2], PhD (Psych); Lajos Botz[1], PharmD, PhD

[1]Department of Pharmaceutics, Faculty of Pharmacy, University of Pécs, Pécs, Hungary

[2]Institute of Psychology, Faculty of Humanities, University of Pécs, Pécs, Hungary

Corresponding Author:
András Fittler, PharmD, PhD
Department of Pharmaceutics
Faculty of Pharmacy
University of Pécs
Honvéd str 3
Pécs, H7625
Hungary
Phone: 36 205566509
Fax: 36 72536285
Email: fittler.andras@pte.hu

Abstract

Background: During the past two decades, the internet has become an accepted way to purchase products and services. Buying medications online are no exception. Besides its benefits, several patient safety risks are linked to the purchase of medicines outside the traditional supply chain. Although thousands of internet pharmacies are accessible on the web, the actual size of the market is unknown. Currently, there is limited data available on the use of internet pharmacies, the number, and attitude of people obtaining medications and other health products from the internet.

Objective: This study aims to gather information on the frequency and attitudes of patients purchasing medications online in a nationally representative sample of outpatients. Attitudes towards main supply chain channels, perceived benefits, and disadvantages of influencing online medication purchase are evaluated.

Methods: A cross-sectional explorative study using a personally administered survey was conducted in a representative sample of Hungarian outpatients in 2018.

Results: A total of 1055 outpatients completed the survey (response rate 77.23%). The mean age was 45 years, and 456 (43.22%) reported having chronic health conditions. The majority (872/1055, 82.65%) of the respondents were aware that medications could be obtained online, but only 44 (4.17%) used the internet for previous medication purchases. Attitudes towards the different pharmaceutical supply chain retail channels showed significant differences (P<.001), respondents accepted retail pharmacy units as the most appropriate source of medications while rejected internet pharmacies. Respondents were asked to evaluate 9 statements regarding the potential benefits and disadvantages about the online medicine purchase, and based on the computed relative attitude rate there is a weak still significant tendency toward rejection (P<.001). Correspondence of demographic factors, internet usage behavior, and prospective online drug purchase attitude was evaluated. Respondents who use the internet more and purchase goods online will be more likely to buy medications online. Furthermore, youth and education will determine the medication purchase behavior.

Conclusions: Many patients will purchase medications on the internet in the future. Currently, there is an increased risk of patients buying products from illegal sites because these dominate the global online pharmacy market. Consequently, improved patient-provider communication and promotion campaigns are needed to inform the public about the safe use of internet pharmacies, as these initiatives can directly prevent patient safety threats.

(J Med Internet Res 2018;20(8):e11115) doi:10.2196/11115

XSL•FO
RenderX

KEYWORDS

survey; internet pharmacy; online medications; Hungary; attitude

Introduction

The Internet Market of Pharmaceuticals

The internet has revolutionized and changed our lives, communication, and procurement practices and strategies [1]. As access to the internet increases, its use to seek health information is also expanding. Estimates worldwide show that approximately 4.5% of all internet searches are linked to health-related questions or information [2]. Population-based surveys found that 72% of the online population in the US and 71% of internet users in Europe, searched for health information at least once in the previous twelve months [3-5]. These tendencies are further extended by mobile device uses [6]. However, consumers turn to the internet today not only for retrieving health information, but also to self-diagnose and obtain various health services or products [7,8].

According to an early definition by Fung et al [9], an online pharmacy is an internet-based vendor (legal or illegal), which sells medicine and may operate as an independent internet-only site, an online branch of "brick-and-mortar" pharmacy, or sites representing a partnership among pharmacies. Briefly, an online pharmacy is a website offering to deliver, distribute, or dispense medication on the internet directly to consumers [10,11]. The growing market of online pharmacies is facilitated by the rapid expansion of the internet, the ever-increasing digital health, the shift towards self-diagnosing from the direct doctor-patient relations, consumer experience in online purchases, the ease of mail-order trade, and distance selling [12,13].

The internet's supply of pharmaceuticals has developed in numerous ways and according to different models in each part of the world. This is due to diverse regulatory, economic, and cultural environments. In the US, the internet pharmacy market is mainly prescription based, while throughout Europe, this segment is forming according to a nonprescription based model [14]. Today, internet pharmacies can be accessed globally. Thus, the legislative and economic perspectives should be considered throughout every country, worldwide. Therefore, online pharmacies generate regulatory confusion as pharmaceuticals and health services "move" between jurisdictional boundaries. Hence, the countries of operation and delivery must be evaluated [15]. While the country of operation determines the licensing requirements and the quality assurance standards in support of the practice of internet selling of medications, mail-order must be performed in accordance with the latter. However, since many illegitimate websites are unwilling to indicate their actual location, one cannot be certain of the regulatory framework under which the internet pharmacy is operating [15]. It is further complicated by the fact that national authorities are typically powerless beyond their borders [16,17].

There are several patient safety risks linked to the online purchase of medicines outside the traditional supply chain, including counterfeit medications. The proportion of counterfeit medicine is estimated to be 10% worldwide [18] ranging from less than 1% in the developed countries [19] to over 30% in developing countries such as Africa, Asia, India, and Latin America [20,21].

Illegal actors primarily focus on the uncontrolled sale of prescription drugs outside the regulated drug supply system [22]. Their marketing strategy includes emphasizing the most commonly preferred benefits of online pharmacies (convenience, speed, discounts, privacy, not visiting the physician, bulk orders and discounts, bonus medicine as a gift) and retaining information regarding adverse effects, contraindications, and drug interactions [21]. Nearly every therapeutic category of drugs is available through the internet. Not only the performance and image-enhancing and "lifestyle drugs" [23,24], such as phosphodiesterase type 5 inhibitors [25-27] or anti-baldness products [28], but life-saving medicines (eg, from the World Health Organization Essential Medicines List), analgesics (nonsteroidal anti-inflammatory drugs, opioids) [29,30], psychiatric [31], obesity [32,33], and cardiologic drugs [18] can be purchased freely over the internet.

The primary characteristics of this illegal market segment consists in the trading of seemingly identical products in an uncontrolled environment, with no restrictions on the consumers (eg, people under 18 can also purchase medications via the internet) or on products (larger quantities can be purchased) from a large virtual supply [18,21,34,35]. During the past two decades, the internet has become an accepted way to purchase medications due to convenience, the potential to save money, and privacy. Early reports on the use of the internet for buying drugs indicate the practical reality of obtaining prescriptions or purchasing prescription drugs online is very small [36]. However, recent reports suggest that the use of internet pharmacies, the number of people obtaining medications, and other health products online is increasing [37].

Aims

This study aims to gather information on the frequency and attitudes of patients purchasing medications online using a nationally representative sample of outpatients in the Southern Transdanubian region of Hungary. Attitudes towards main supply chain channels, perceived benefits, and disadvantages influencing online medication purchases were evaluated.

Methods

In our cross-sectional explorative study, a personally administered survey was used. The characteristics and background of the respondents were measured through the following independent variables: (1) gender, (2) age, (3) level of education, (4) place of residence, (5) average income, (6) internet usage, (7) online purchase habits in general, and (8) self-reported health status.

A Hungarian language survey was developed by the authors (AF and RGV [pharmacists], and MK [psychologist]) in support of this study, based on previous research [38] and a prior pilot study. In the online pilot study, open questions were used, covering the topics of the study survey to map the general

XSL•FO
RenderX

attitudes of the prospective sample. In the main study, data were collected directly from Hungarian citizens using the outpatient health services for chronic or acute conditions, between January and March 2018. Participants were considered eligible if they were 16 years of age or older, and were excluded if they were unwilling to participate in the survey.

Trained research associates administered the 28-item survey. It consisted of an introductory paragraph on the aims of the survey, information on confidentiality, and anonymity followed by 5 main sections: (1) evaluation of channels available for procuring medications, (2) online medicine purchase experiences and attitude, (3) internet use, (4) health status and medication use, and (5) demographics. The study protocol was approved by the institutional review board (approval number 6835). In the survey, 5-point Likert-type scales, multiple-response, and multiple-choice questions were used. The English translated version of the survey (Multimedia Appendix 1) and the original Hungarian version (Multimedia Appendix 2) are provided as supplementary material. Statistical analyses were conducted using the SPSS software version 22. Descriptive statistics was used to describe respondent characteristics.

Results

Respondent Characteristics

Trained research associates approached a total of 1366 patients. This resulted in the completion of 1055 surveys indicating a response rate of 77.23% (see respondent characteristics in Table 1). The distribution of female (539/1055, 51.09%) and male (516/1055, 48.91%) respondents was nearly equal. Our sample consisted of people obtaining outpatient health service for chronic or acute conditions, thus it represents patients rather than the general population regarding age and number of medications used. The mean age was 45 (SD 17.36) years. Nearly half (456/1055, 43.22%) of the respondents reported to have chronic health conditions and a majority reportedly used at least one medication regularly.

Our survey sample represents the Hungarian society regarding the level of education [39]. According to recent statistics, 72% of individuals in the European Union member states accessed the internet daily, while 57% of Europeans (aged 16 to 74 years) ordered or bought goods or services over the internet for private use. Accordingly, our sample represents the European population for internet use and online purchases [40].

Table 1. Respondent demographic, health status, and internet use characteristics (N=1055).

Variable	Value
Age (years), mean (SD)[a]	45.08 (17.36)
Gender, n (%)	
Female	539 (51.09)
Male	516 (48.91)
Education, n (%)	
Completed primary school	68 (6.45)
Graduated high school	656 (61.18)
Graduated college or university	329 (31.18)
Advanced (PhD, Doctor of Liberal Arts)	2 (0.19)
Patients with chronic conditions, n (%)	456 (43.22)
Number of regular medications per patient, mean (SD)	1.55 (2.63)
Number of regular medications per patient, range	0-25
Frequency of internet use, n (%)	
Daily	737 (69.86)
Weekly	150 (14.22)
Never	168 (15.92)
Frequency of online shopping, n (%)	
Regularly	203 (19.24)
A few times	515 (48.82)
Never	337 (31.94)

[a]Median 45 years, range 16-89 years.

XSL•FO
RenderX

Evaluation of Supply Chain Retail Channels, Previous, and Prospective Purchases

Attitudes towards the 3 main supply chain retail participants: conventional community pharmacy units, nonpharmacy units, and internet pharmacies were evaluated. The respondents were asked to rate pharmacies, nonpharmacy units (eg, petrol stations) and internet sources on a 5-point Likert scale.

They were asked to express, according to their opinion, how appropriate were these sources regarding the purchase of medication. A score of 1 was given for "not appropriate at all" and 5 for "entirely appropriate." The results are shown in Table 2. A repeated measures analysis of variance (ANOVA) was conducted on the sample, and significant differences were found. The respondents accepted retail pharmacy units as the most appropriate source of medications, while they exhibited neutral attitudes toward nonpharmacy units and rejected internet pharmacies ($F_{1.95,2056.66}$=1776.78, P<.001)

According to our results, 872 (82.65%) of the respondents were aware that medications could be obtained online, while only 44 (4.17%) used the internet for the purchase of medication at least once. However, this number is likely to increase in the future as numerous patients were open to prospective online purchases: 100 (9.47%) stated they were very likely, and 146 (13.83%) noted that they were likely to purchase medications online in the future.

Perceived Benefits and Disadvantages

Based on the results of the pilot study, 9 statements regarding the potential benefits and disadvantages were measured to determine the factors influencing the attitudes toward online medication purchase. The respondents were asked to evaluate each statement (Table 3) on potential benefits and disadvantages regarding their own attitudes on a 5-point Likert scale. A score of 1 was given for "I don't agree" and 5 for "I agree."

Table 2. A summary of attitudes towards the 3 main supply chain retail channels using a 5-point Likert scale. A score of 1 was given for "not appropriate at all" while 5 was given for "entirely appropriate."

Retail channels	Mean (SD)
Pharmacy	4.79 (0.53)
Nonpharmacy units	2.94 (1.38)
Internet	2.25 (1.42)

Table 3. A comparative evaluation of potential benefits and disadvantages of online drug shopping.

Parameters	Evaluation, mean (SD)
Potential benefits	
Convenient	4.29 (1.07)
People who cannot get to a pharmacy can also purchase products	4.18 (1.11)
I can purchase medicines after opening hours	4.1 (1.19)
I can access products which are otherwise not available for me	3.34 (1.43)
Fast	3.71 (1.29)
Products can be compared faster and more easily than in the pharmacy	3.15 (1.34)
Inexpensive	2.87 (1.21)
I can get more information compared to the pharmacy	2.85 (1.43)
I can get products with better quality compared to the pharmacy	2.23 (1.19)
Potential disadvantages	
It is easier to abuse preparations	4.24 (1.13)
There is no control, so I can get products that I do not need or worsen my condition	4.22 (1.06)
I do not get proper information regarding the use of the products	3.86 (1.06)
Due to the delivery time, I'm getting the drug later compared to a pharmacy	3.80 (1.15)
The source of the product is not reliable	3.78 (1.27)
It is hard for me to choose between the great numbers of products	3.70 (1.25)
I do not get the right product	3.65 (1.31)
I receive counterfeit medicine	3.61 (1.25)
The quality of the product is lower compared than in local pharmacies	3.20 (1.29)

XSL•FO
RenderX

Table 4. The results of a correlation analysis between demographic factors, internet usage behavior, and prospective online drug purchase attitude.

Parameter	Prospective online medication purchase attitude
Age	–0.28
Average time spent on the internet	0.31
Internet purchase frequency in general	0.37
Settlement size	0.07
Level of education	0.20
Average income	0.06

The reliability of the answers on the benefits and disadvantages were calculated, and Cronbach's alpha was determined (benefits alpha=.76, disadvantages alpha=.84) suggesting the reliability values are satisfying. The reactionary attitude regarding the online purchase of medicine was weighted, and a relative attitude rate was computed with a mean of –0.37 (SD 1.25). There was a weak but still significant tendency toward rejection (t_{1054}=9.64, P<.001). Our results showed that there were several factors positively influencing the respondents' attitude toward online medication purchase, but they still tended to reject this source of drug acquisition. Linear regression analysis was conducted to measure the predictive power of the reported attitudes regarding the willingness to purchase medication online. These results show that attitudes have significant predictive power ($F_{2,1054}$=224.87, P<.001, R^2=.299).

Attitudes Towards Prospective Online Purchases

Our study could not substantiate a clear association with purchasing medications online, due to the small ratio of respondents with previous purchasing experience. However, we did evaluate the willingness to purchase medications on the internet. The correspondence of demographic factors, internet usage behavior, and prospective online drug purchase attitude was examined (see Table 4). A correlation analysis indicated that a significant correlation was found between age, average time spent on the internet, internet purchase in general, settlement size, and level of education.

A linear regression analysis was conducted using a stepwise method to measure the effect of correlating factors on willingness to purchase medications online. According to our results, the factor of internet purchase in general, level of education, and age together, will predict the attitude toward the online purchase of medications ($F_{3,1053}$=65.83, P<.001, R^2=.158). A further linear regression analysis was conducted to determine the predicting factors of online purchase behavior, in general. Our results showed the time spent on the internet, age, and education determines the general online purchase ($F_{3,1053}$=292.36, P<.001, R^2=.455). Also, the time spent on the internet is highly determined by age and the level of education ($F_{2,1053}$=445.13, P<.001, R^2=.459). Based on our results, an explanation model was developed. According to our results, the attitude toward online medication purchase can be explained by the factor of general online purchase behavior, which is determined by the time spent on the internet, which has a strong correspondence to age. The level of education will predict all the other factors, and through this, have a general impact on online medication purchase attitude.

Discussion

Consumers Purchasing Medicines Online

According to several surveys, the percentage of people purchasing medicines online varies as published data differs, due to type of product, sample population, education, and income or substance abuse status [10,12]. Thus, the authors aimed to summarize previously published data and key findings on patients procuring medications and health products on the web and perceptions regarding internet pharmacies.

Table 5 summarizes recent scientific data on online medication purchasing of the general population published between 2012-2017. Accordingly, it does not contain: data presented in the previous systematic review by Orizio et al [10], surveys focusing solely on online and mail-order pharmacy users, prescription drugs' customers, and patients participating in prescription refill programs. Questionnaires from a small sample size or specific patient groups (eg, drug abusers, people buying illegal drugs, men obtaining phosphodiesterase-5 inhibitors) were excluded. Furthermore, nonpeer reviewed publications and estimates were not included, mostly to eliminate bias. Studies not representing the current web used by the general population (eg, early reports and studies focusing on the dark web) were also excluded. It must be noted, the differentiation of actual product categories purchased online is somewhat difficult, as approved medicines, medicinal products, and dietary supplements are often measured as a single category in articles evaluating consumers purchasing from online pharmacies.

Several authors have studied the number, attitude, and characteristics of online pharmacy users. Although internet pharmacies have been in business for nearly two decades, there is only limited scientific evidence published regarding the prevalence of online pharmacy use by the general population [42]. In countries where the internet pharmacy market is dominated by retail pharmacy chains selling prescription-only medications and offering refill programs (eg, US, United Kingdom, Germany), patient characteristics differ from nations in which only nonprescription products can be marketed (eg, Central European Union) through the internet.

XSL•FO
RenderX

Table 5. Summary of recent studies on the prevalence of purchasing drugs and dietary supplements online.

Reference	Location (year of data collection)	N[a]	Sample population	Survey method	Respondents purchasing health products online (%)
Abanmy [41]	Saudi Arabia (2013-2014)	633	Random sample of internet users	Online	2.7
Alfahad et al [42]	Saudi Arabia (2014-2015)	346	Random sample of internet users	Online	1.4
Szekely et al [43]	Romania (2010-2011)	253	Community pharmacy patients	Personally administered	8.3
Desai et al [37]	USA (2007)	5074	Internet users	Data from HINTS[b] national dataset	14.5
Brown and Lee [44]	USA (2002-2010)	88,240	Noninstitutionalized individuals	Data from MEPS[c] national dataset	0.5[d]
Fittler et al [38]	Hungary (2010-2011)	422	Hospital patients	Personally administered	8.4
Mazer et al [45]	USA (2007)	1657	Emergency department patients	Personally administered	5.4[e]

[a]N refers to number of respondents.

[b]HINTS: Health Information National Trends Survey.

[c]MEPS: Medical Expenditure Panel Survey.

[d]Refers to prescription medication.

[e]Refers to medication in general.

As noted by numerous authors who have reviewed the published data, it is most difficult to measure the number of online pharmacies and their customers, especially illegitimate ones [10]. Existing research only provides estimates of the scale of the online purchase of medicine [46].

A US survey, in a cohort of 443 online pharmacy users, representing an average of 1.5 million individuals annually, found that, compared with nonusers, online users were older, are more likely to have private insurance, had more prescriptions, possessed a higher family income, and were more educated [44]. Atkinson et al [47] found that age and marital status was associated with online buying. In 2007, Desai and colleagues [37] documented that about 14.5% of the US population used the internet to obtain medications or vitamins, and these individuals were more often married, white, 50-64 years old with varying levels of college education and an income of US $75,000, or more. It should be noted, mail-order pharmacies are required by certain health plans in the US, which may result in a higher number of patients using online mail pharmacies for maintenance medications [48], compared to other countries. A Hungarian survey of hospital patients has shown in which 8.4% of the respondents ordered drugs or dietary supplements online and 3.7% of the respondents are considering this option in the future. Gender, age, and educational profile did not significantly affect the experience in ordering health-related products from the internet [38]. An Italian study involving more than 100 adult subjects investigated the use of the internet regarding the search for information on medicines, dietary supplements, and disease. Although 68.5% of the respondents were aware of the possibility to purchase medicines on the internet, only 9.2% expressed a favorable opinion. Interestingly, the number of participants with actual online medication purchase experience was not measured [49]. Mazer et al [45] found, among emergency department patients, 57% reported awareness of online pharmacies and 5.4% used the internet to order medications. Multiple medications and prescription plan significantly influenced online pharmacy use, while no difference in age or student status was observed between users and nonusers. A survey of Saudi citizens showed the online purchase of medicines is not yet widespread, as 23.1% of the respondents were aware of the existence of internet pharmacies and only 2.7% had purchased medicines online. However, the level of satisfaction was high among those who had such experience, and a suitable number of respondents (42.7%) indicated they are willing to try an online pharmacy in the future [41]. Similar results were reported by Alfahad et al [42] throughout the kingdom, as a clear majority of the Saudi respondents have not yet heard about online pharmacies (82.6%), and very few (1.4%) have purchased medicinal products online, however about two thirds (66.4%) were enthusiastic to utilize the online options in the purchasing of medicines. On the other hand, a Romanian survey found only a minority (3.2%) of the respondents have not heard of the possibility of purchasing medicines online, 8.3% have already purchased, moreover 7.1% intended to do so in the future [43].

A systematic review by Orizio et al [10] investigated the available evidence regarding online pharmacies, published between 2003-2010. The authors summarized population surveys on consumers perceptions and attitudes yet could not find consistent information regarding the number of consumers and their characteristics [10]. According to another review by Orsolini et al [50] a range of variables must be considered in profiling online pharmacy customers. Most online customers were reported to be young, Caucasian, and individuals without any health insurance. However, there are variations in gender and age depending on the type of medication purchased. Women and more educated individuals were associated with the online search of health-related information, and, conversely, subjects with a low literacy level are prone to purchase from illegal websites.

XSL•FO
RenderX

The above studies provide important clues into what genres of patients and consumers may be more likely to purchase products from online pharmacies. At the same time, we must admit, it is somewhat difficult to profile the "typical online pharmacy customer," because users are as diverse as the medications they are looking for [12]. This study has identified a knowledge deficit and lack of scientific evidence regarding studies on patients' attitudes towards internet pharmacies. Furthermore, we believe the additional scientific evidence is necessary to plan, implement, and evaluate prevention campaigns aiming to facilitate the safe use while navigating the online pharmaceutical market.

Principal Results

Patients' attitude towards online pharmacies and purchasing health products on the internet is a key element of maintaining the integrity of the medication supply chain and protecting patients from digital iatrogenesis [51]. Policy makers and authorities must be aware of how medications are different from most items purchased on the internet, as they directly have an impact on one's health [37].

Users of online pharmacies, whether legitimate sites or not, are purchasing medications used for both acute and chronic conditions, including medications of abuse. Without the appropriate advice and the supervision of the medical doctor or the pharmacist, all drugs (prescription only and over-the-counter) and even dietary-supplements can cause harm. The primary concerns are insufficient or incorrect information about the patient's health status and medications, inappropriate self-diagnosis, or incomplete management of drug-related issues, such as polypharmacy, therapeutic duplications, adverse events, and drug-to-drug or drug-herbal interactions [45]. Obviously, illegitimate online pharmacies pose a definite threat to patients through the marketing of counterfeit medications and substandard products. However, even legitimate actors possess issues associated with their use [37].

Given the variable quality of internet pharmacies, it is critical to identify patients' vulnerability and develop targeted campaigns to educate the public properly. Our results may support the development of patient-centered interventions by identifying consumer characteristics associated with willingness towards purchasing medications online.

A majority (872/1055, 82.65%) of the respondents were aware medications can be obtained online. In our study sample, 44 (4.17%) reportedly used the internet for purchasing medications previously, and this number is likely to increase in the future, as numerous patients were more likely to purchase medications online in the future (23.3%). Attitudes towards the three main supply chain retail channels showed respondents accepted community pharmacy units as the most appropriate source of medications, yet showed neutral attitudes toward nonpharmacy units and rejected internet pharmacies. The comparative evaluation of potential benefits and disadvantages of online drug shopping demonstrates, there are several factors positively influencing the respondents' attitude toward online medication purchase (eg, convenience, about individuals who cannot get to a pharmacy can purchase products, possibility to purchase medicines after opening hours). Interestingly, they still tend to reject this source of drug acquisition. Seemingly, Hungarian patients are not yet open towards online medication purchase, however, based on our findings, attitudes will likely change as more individuals gain experience in buying products or services online.

Based on our results, the main factors influencing the willingness towards online medication purchase was determined. We found, internet usage and online purchase behavior, in general, will predict the attitude toward online medication purchase. Essentially, this means the respondents who use the internet and purchase goods online will be more likely to purchase medications online. These 2 factors are highly influenced by age, meaning the younger generation is much more involved in the online market. In addition, the level of education and attitude will determine the medication purchase behavior. People possessing higher level college degrees and having a more positive attitude regarding online medication purchase, are more likely to do so. The attitude ratings are also related to age, mostly in aspects of benefits (r=.23). Based on these data, we summarize how online medication purchase is not an isolated phenomenon, but highly integrated with the behavioral tendency in which individuals often try to manage increasingly things online. Given this a general tendency throughout society, we believe efforts against online medication purchase, in general, is pointless. However, it is highly essential to develop a stronger control over online pharmacies and provide education regarding the responsible internet purchase behavior. Previous studies have noted that we must further emphasize the opportunity in support of health providers (general practitioners, pharmacists, nurses) to help patients navigate potential internet purchases, in the prevention of medication incidents generated using unapproved and illegitimate online pharmacies [37,38].

Strengths and Limitations

In this study, the potential misunderstanding of the survey was eliminated by personally administering the survey and the use of trained research associates. Compared to numerous previous studies, our method makes it possible to identify 2 seemingly similar, but different product categories (medications and dietary supplements) separately. In the beginning of the survey, the special attributes of medications were measured and discussed for future reference. In our sample, we gathered results from a balanced respondent population, representing the patient population regarding age, gender, and education. Previous survey results regarding online consumers may not reflect the prevalence and the attitude of the general population because internet use is strongly related to several demographic variables. Accordingly, by using personally administered surveys, our study eliminates such bias.

This study does have some limitations. Online pharmacy use was self-reported, therefore, subject to recall bias and untruthful reporting by the individual, which may underestimate actual prevalence. Legitimate and illegitimate actors are not differentiated in our study, mainly, since online sellers may mislead customers or not be able to differentiate between them [12,52]. Our study was performed in hospitals, general practitioners' offices, and community pharmacies throughout Southern Hungary. Consequently, it represents the national

XSL•FO
RenderX

patient population but not the entire Hungarian population. However, this can be a potential strength, as patients are more likely to purchase medications and they are also vulnerable to the potential dangers associated with internet pharmacies. Prevalence and attitudes of inpatients were measured and published previously by the authors [38].

Conclusions

Our results support our hypothesis: the use of the internet to purchase medications is present and national results are in correlation with international data. Despite a weak, but still significant tendency towards rejection to online pharmacies was identified, a reasonable number of patients were planning to purchase medications on the internet in the future.

We aimed to identify drivers in support of the purchase of online medication to develop targeted campaigns informing patients vulnerable to illegal websites. Based on the literature review and our study results, it is difficult to profile consumers turning to the internet pharmacy market because users are likely just as diverse as the treatments they are looking for. However, following the evaluation of prospective online drug purchase attitude, we arrived to the conclusion that respondents who use the internet more frequently to shop online will be more likely to purchase medications on the internet. Furthermore, age and education determine in the medication purchase behavior. Our results ideally will soon support educational interventions promoting safe online medication procuring practices and provide valuable data regarding patients' attitudes.

It is highly recommended that the participants of the health care system, both throughout the institutional and outpatient settings, begin documenting procurement information during the medical examination, interview, and anamnesis of their patients. This will ensure that more robust and informative data will be made available regarding the effective penetration of this relatively novel distribution channel of the drug supply system.

Future research should focus on exploring adverse effects resulting from medications purchased online. New emerging technologies, such as machine learning algorithms applied to "Big Data," is the basis for a new area of research referred to as, "digital" surveillance or "infoveillance" [53]. Accordingly, the actual patient safety risk in outpatients can be identified (1) within the health care system by collecting data obtained during medical examinations and anamnesis, (2) by the evaluation of patient records, or (3) performed online by novel data-science methods.

Improved patient-provider communication, promotion, and education campaigns are needed to inform and educate the public on the safe use of internet pharmacies. These initiatives can prevent threats to patient safety. Targeted interventions by pharmacists (medication review or reconciliation, professional advice regarding the evaluation of online distributors, and differentiate legal and illegal medication suppliers) are potential prevention strategies which must be emphasized, including the effective implementation in everyday practice [54].

Acknowledgments

AF conceived the review of recent literature regarding online pharmacy use and MK was responsible for analysis. AF, RGV, and MK jointly drafted the manuscript, and LB was involved in the critical revision. The authors presented preliminary data in the form of a research poster at the 23rd Congress of the European Association of Hospital Pharmacists, Gothenburg, Sweden. The study was supported by the UNKP-17-4-III-PTE-49 New National Excellence Program of The Ministry of Human Capacities.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey in English.

[PDF File (Adobe PDF File), 58KB - jmir_v20i8e11115_app1.pdf]

Multimedia Appendix 2

Survey in Hungarian.

[PDF File (Adobe PDF File), 61KB - jmir_v20i8e11115_app2.pdf]

References

1. von Rosen AJ, von Rosen FT, Tinnemann P, Müller-Riemenschneider F. Sexual Health and the Internet: Cross-Sectional Study of Online Preferences Among Adolescents. J Med Internet Res 2017 Nov 08;19(11):e379 [FREE Full text] [doi: 10.2196/jmir.7068] [Medline: 29117927]
2. Das A, Faxvaag A. What influences patient participation in an online forum for weight loss surgery? A qualitative case study. Interact J Med Res 2014 Feb 06;3(1):e4 [FREE Full text] [doi: 10.2196/ijmr.2847] [Medline: 24509408]
3. Fox S, Duggan M. Pew Research Center's Internet & American Life Project. 2013 Jan 15. Health Online 2013 - 35% of U.S. adults have gone online to figure out a medical condition; of these, half followed up with a visit to a medical professional

XSL•FO
RenderX

URL: http://www.pewinternet.org/files/old-media/Files/Reports/PIP_HealthOnline.pdf [accessed 2018-07-31] [WebCite Cache ID 71KBwln6G]
4. Andreassen H, Bujnowska-Fedak M, Chronaki C, Dumitru R, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007 Apr 10;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]
5. Takahashi Y, Ohura T, Ishizaki T, Okamoto S, Miki K, Naito M, et al. Internet use for health-related information via personal computers and cell phones in Japan: a cross-sectional population-based survey. J Med Internet Res 2011 Dec 14;13(4):e110 [FREE Full text] [doi: 10.2196/jmir.1796] [Medline: 22169526]
6. Gawron LM, Turok DK. Pills on the World Wide Web: reducing barriers through technology. Am J Obstet Gynecol 2015 Oct;213(4):500.e1-500.e4. [doi: 10.1016/j.ajog.2015.06.002] [Medline: 26071918]
7. Orizio G, Schulz P, Domenighini S, Caimi L, Rosati C, Rubinelli S, et al. Cyberdrugs: a cross-sectional study of online pharmacies characteristics. Eur J Public Health 2009 Aug;19(4):375-377. [doi: 10.1093/eurpub/ckn146] [Medline: 19151103]
8. Fox S, Duggan M. Health Online. 2013 Jan 15. Information Triage - For one-third of U.S. adults, the internet is a diagnostic tool URL: http://www.pewinternet.org/2013/01/15/information-triage/ [accessed 2018-05-22] [WebCite Cache ID 6zbZNFWhw]
9. Fung CH, Woo H, Asch S. Controversies and legal issues of prescribing and dispensing medications using the Internet. Mayo Clin Proc 2004 Feb;79(2):188-194. [doi: 10.4065/79.2.188] [Medline: 14959914]
10. Orizio G, Merla A, Schulz PJ, Gelatti U. Quality of online pharmacies and websites selling prescription drugs: a systematic review. J Med Internet Res 2011 Sep 30;13(3):e74 [FREE Full text] [doi: 10.2196/jmir.1795] [Medline: 21965220]
11. US Congress. 2008. Ryan Haight Online Pharmacy Consumer Protection Act of 2008 URL: https://www.govtrack.us/congress/bills/110/hr6353 [accessed 2018-05-22] [WebCite Cache ID 6zby11bQE]
12. Mackey TK, Nayyar G. Digital danger: a review of the global public health, patient safety and cybersecurity threats posed by illicit online pharmacies. Br Med Bull 2016 Jun;118(1):110-126 [FREE Full text] [doi: 10.1093/bmb/ldw016] [Medline: 27151957]
13. Gabay M. Regulation of Internet Pharmacies: A Continuing Challenge. Hosp Pharm 2015 Sep;50(8):681-682 [FREE Full text] [doi: 10.1310/hpj5008-681] [Medline: 26823617]
14. Dudley J. Entering the Digital Battleground. In: Mail Order and Internet Pharmacy in Europe. Kidderminster UK: James Dudley International Ltd; 2014.
15. Cohen JC. Public policy implications of cross-border Internet pharmacies. Manag Care 2004 Mar;13(3 Suppl):14-16 [FREE Full text] [Medline: 15074164]
16. Chilet-Rosell E, Martín LM, Ruiz CMT, Alonso-Coello P. Hormone replacement therapy advertising: sense and nonsense on the web pages of the best-selling pharmaceuticals in Spain. BMC Public Health 2010 Mar 16;10:134 [FREE Full text] [doi: 10.1186/1471-2458-10-134] [Medline: 20233393]
17. Mackey T, Kalyanam J, Klugman J, Kuzmenko E, Gupta R. Solution to Detect, Classify, and Report Illicit Online Marketing and Sales of Controlled Substances via Twitter: Using Machine Learning and Web Forensics to Combat Digital Opioid Access. J Med Internet Res 2018 Dec 27;20(4):e10029 [FREE Full text] [doi: 10.2196/10029] [Medline: 29613851]
18. Blackstone EA, Fuhr JJ, Pociask S. The health and economic effects of counterfeit drugs. Am Health Drug Benefits 2014 Jun;7(4):216-224 [FREE Full text] [Medline: 25126373]
19. Howard D. A silent epidemic: protecting the safety and security of drugs. Pharmaceutical Outsourcing 2010 Jul 01;11(4):16-18 [FREE Full text]
20. Dorsey PJ, Hellstrom WJ. Medscape. 2007. The Illicit Sale of Medications for the Treatment of Erectile Dysfunction URL: http://www.medscape.com/viewarticle/566897_3 [accessed 2018-05-23] [WebCite Cache ID 6zcZTp2x8]
21. Bate R. The Deadly World of Fake Drugs. Foreign Policy, Washington Post 2008;168:56-65.
22. Levaggi R, Orizio G, Domenighini S, Bressanelli M, Schulz PJ, Zani C, et al. Marketing and pricing strategies of online pharmacies. Health Policy 2009 Oct 01;92(2-3):187-196. [doi: 10.1016/j.healthpol.2009.03.010] [Medline: 19394104]
23. Corazza O, Bersani FS, Brunoro R, Valeriani G, Martinotti G, Schifano F. The diffusion of performance and image-enhancing drugs (PIEDs) on the internet: the abuse of the cognitive enhancer piracetam. Subst Use Misuse 2014 Dec;49(14):1849-1856. [doi: 10.3109/10826084.2014.912232] [Medline: 24827869]
24. Flower R. Lifestyle drugs: pharmacology and the social agenda. Trends Pharmacol Sci 2004 Apr 01;25(4):182-185. [doi: 10.1016/j.tips.2004.02.006] [Medline: 15063081]
25. Mendoza RL. Autonomy and paternalism in medical e-commerce. Med Health Care Philos 2015 Aug;18(3):379-391. [doi: 10.1007/s11019-014-9619-0] [Medline: 25547271]
26. Patterson R, Mabe P, Mitchell EN, Cory W. Lifestyle illicit drug seizures: a routine ESI-LC-MS method for the identification of sildenafil and vardenafil. Forensic Sci Int 2012 Oct 10;222(1-3):83-88. [doi: 10.1016/j.forsciint.2012.05.003] [Medline: 22673705]
27. Jackson G, Arver S, Banks I, Stecher VJ. Counterfeit phosphodiesterase type 5 inhibitors pose significant safety risks. Int J Clin Pract 2010 Mar;64(4):497-504 [FREE Full text] [doi: 10.1111/j.1742-1241.2009.02328.x] [Medline: 20088883]
28. Hellstrom WJG. The growing concerns regarding counterfeit medications. J Sex Med 2011 Jan;8(1):1-3. [doi: 10.1111/j.1743-6109.2010.02135.x] [Medline: 21199373]

XSL•FO
RenderX

29. Boyer EW, Wines JD. Impact of Internet pharmacy regulation on opioid analgesic availability. J Stud Alcohol Drugs 2008 Sep;69(5):703-708 [FREE Full text] [Medline: 18781245]
30. Raine C, Webb DJ, Maxwell SRJ. The availability of prescription-only analgesics purchased from the internet in the UK. Br J Clin Pharmacol 2009 Feb;67(2):250-254 [FREE Full text] [doi: 10.1111/j.1365-2125.2008.03343.x] [Medline: 19154446]
31. Trefi S, Gilard V, Balayssac S, Malet-Martino M, Martino R. Quality assessment of fluoxetine and fluvoxamine pharmaceutical formulations purchased in different countries or via the Internet by 19F and 2D DOSY 1H NMR. J Pharm Biomed Anal 2008 Mar 13;46(4):707-722. [doi: 10.1016/j.jpba.2007.11.038] [Medline: 18206329]
32. Hackett A, Krska J. Is it time to regulate over-the-counter weight-loss formulations? Int J Pharm Pract 2012 Jun;20(3):199-202. [doi: 10.1111/j.2042-7174.2011.00181.x] [Medline: 22554163]
33. Chambliss WG, Carroll WA, Kennedy D, Levine D, Moné MA, Ried LD, et al. Role of the pharmacist in preventing distribution of counterfeit medications. J Am Pharm Assoc (2003) 2012 Mar;52(2):195-199. [doi: 10.1331/JAPhA.2012.11085] [Medline: 22370383]
34. Ghodse H. Watching internet pharmacies. Br J Psychiatry 2010 Mar;196(3):169-170 [FREE Full text] [doi: 10.1192/bjp.bp.109.072413] [Medline: 20194534]
35. Schifano F, Deluca P, Baldacchino A. Online availability of dextropropoxyphene over time, 2003-2005. Cyberpsychol Behav 2006 Aug 01;9(4):404-409. [doi: 10.1089/cpb.2006.9.404] [Medline: 16901243]
36. Baker L, Wagner T, Singer S, Bundorf M. Use of the Internet and e-mail for health care information: results from a national survey. JAMA 2003 May 14;289(18):2400-2406. [doi: 10.1001/jama.289.18.2400] [Medline: 12746364]
37. Desai K, Chewning B, Mott D. Health care use amongst online buyers of medications and vitamins. Res Social Adm Pharm 2015 Nov;11(6):844-858. [doi: 10.1016/j.sapharm.2015.01.001] [Medline: 25769500]
38. Fittler A, Lankó E, Brachmann B, Botz L. Behaviour analysis of patients who purchase medicines on the internet: can hospital pharmacists facilitate online medication safety? Eur J Hosp Pharm 2012 Sep 07;20(1):8-12. [doi: 10.1136/ejhpharm-2012-000085]
39. Hungarian Central Statistical Office. 2011 Nov 01. Retrospective national data from population census 2011 - Number and characteristics of the population 2011 URL: http://www.ksh.hu/nepszamlalas/tables_regional_00 [accessed 2018-05-23] [WebCite Cache ID 6zcpw0azT]
40. Digital economy and society statistics - households and individuals. 2018 Mar 01. Eurostat URL: http://ec.europa.eu/eurostat/statistics-explained/index.php/Digital_economy_and_society_statistics_-_households_and_individuals [accessed 2018-05-23] [WebCite Cache ID 6zcr8BWoY]
41. Abanmy N. The extent of use of online pharmacies in Saudi Arabia. Saudi Pharm J 2017 Sep;25(6):891-899 [FREE Full text] [doi: 10.1016/j.jsps.2017.02.001] [Medline: 28951675]
42. Alfahad NF, Albelali MT, Khurshid F, Al-Arifi MN, Al-Dhawailie AA, Alsultam M. Perception and knowledge to online pharmacy services among consumers in Riyadh, Saudi Arabia: a pilot survey. Lat Am J Pharm 2015;34(6):1113-1118 [FREE Full text]
43. Szekely P, Kelemen L, Fittler A, Hancu G, Hajnal F, Ciurba A, et al. Attitude Of patients and customers regarding purchasing drugs online. Farmacia 2015;63(1):93-98 [FREE Full text]
44. Brown J, Li C. Characteristics of online pharmacy users in a nationally representative sample. J Am Pharm Assoc (2003) 2014 May;54(3):289-294. [doi: 10.1331/JAPhA.2014.13169] [Medline: 24816356]
45. Mazer M, DeRoos F, Shofer F, Hollander J, McCusker C, Peacock N, et al. Medications from the web: use of online pharmacies by emergency department patients. J Emerg Med 2012 Feb;42(2):227-232. [doi: 10.1016/j.jemermed.2010.05.035] [Medline: 20656438]
46. Sugiura L. Opportunities in Online Medicine Purchasing. In: Respectable Deviance and Purchasing Medicine Online: Opportunities and Risks for Consumers (Palgrave Studies in Cybercrime and Cybersecurity). Cham, Switzerland: Palgrave Macmillan; 2018:67-100.
47. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009 Feb;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]
48. Khandelwal N, Duncan I, Rubinstein E, Ahmed T, Pegus C. Community pharmacy and mail order cost and utilization for 90-day maintenance medication prescriptions. J Manag Care Pharm 2012 Apr;18(3):247-255 [FREE Full text] [doi: 10.18553/jmcp.2012.18.3.247] [Medline: 22468733]
49. Lombardo S, Cosentino M. Internet Use for Searching Information on Medicines and Disease: A Community Pharmacy-Based Survey Among Adult Pharmacy Customers. Interact J Med Res 2016 Jul 13;5(3):e22 [FREE Full text] [doi: 10.2196/ijmr.5231] [Medline: 27417304]
50. Orsolini L, Francesconi G, Papanti D, Giorgetti A, Schifano F. Profiling online recreational/prescription drugs' customers and overview of drug vending virtual marketplaces. Hum Psychopharmacol 2015 Jul;30(4):302-318. [doi: 10.1002/hup.2466] [Medline: 26216567]
51. Mackey TK, Liang B. Digital Iatrogenesis: Policy Advocacy to Prevent Patient Harm from Access to Dangerous Drugs Online. Harvard Health Policy Rev 2014;14(1):11-13 [FREE Full text]

XSL•FO
RenderX

52. Ivanitskaya L, Brookins-Fisher J, Vibbert D, Erofeev D, Fulton L. Dirt cheap and without prescription: how susceptible are young US consumers to purchasing drugs from rogue internet pharmacies? J Med Internet Res 2010 Apr 26;12(2):e11 [FREE Full text] [doi: 10.2196/jmir.1520] [Medline: 20439253]
53. Mackey TK, Nayyar G. A review of existing and emerging digital technologies to combat the global trade in fake medicines. Expert Opin Drug Saf 2017 May;16(5):587-602. [doi: 10.1080/14740338.2017.1313227] [Medline: 28349715]
54. Vida RG, Fittler A, Mikulka I, Ábrahám E, Sándor V, Kilár F, et al. Availability and quality of illegitimate somatropin products obtained from the Internet. Int J Clin Pharm 2017 Feb;39(1):78-87. [doi: 10.1007/s11096-016-0398-y] [Medline: 27888454]

Abbreviations

ANOVA: analysis of variance
HINTS: Health Information National Trends Survey
MEPS: Medical Expenditure Panel Survey

Edited by G Eysenbach; submitted 23.05.18; peer-reviewed by N Abanmy, R Alkoudmani, M Nounou; comments to author 13.06.18; revised version received 22.06.18; accepted 28.06.18; published 22.08.18.

Please cite as:
Fittler A, Vida RG, Káplár M, Botz L
Consumers Turning to the Internet Pharmacy Market: Cross-Sectional Study on the Frequency and Attitudes of Hungarian Patients Purchasing Medications Online
J Med Internet Res 2018;20(8):e11115
URL: http://www.jmir.org/2018/8/e11115/
doi:10.2196/11115
PMID:30135053

©András Fittler, Róbert György Vida, Mátyás Káplár, Lajos Botz. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 22.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Using Mobile Apps to Assess and Treat Depression in Hispanic and Latino Populations: Fully Remote Randomized Clinical Trial

Abhishek Pratap[1,2], MS; Brenna N Renn[3], PhD; Joshua Volponi[4,5], BS; Sean D Mooney[1], PhD; Adam Gazzaley[4,5], MD, PhD; Patricia A Arean[3], PhD; Joaquin A Anguera[4,5], PhD

[1]Department of Biomedical Informatics and Medical Education, School of Medicine, University of Washington, Seattle, WA, United States

[2]Sage Bionetworks, Seattle, WA, United States

[3]Department of Psychiatry & Behavioral Sciences, University of Washington, Seattle, WA, United States

[4]Department of Neurology, University of California San Francisco, San Francisco, WA, United States

[5]Department of Psychiatry, University of California San Francisco, San Francisco, CA, United States

Corresponding Author:
Abhishek Pratap, MS
Department of Biomedical Informatics and Medical Education
School of Medicine
University of Washington
UW Medicine South Lake Union, Building C, Box 358047
850 Republican Street
Seattle, WA, 98109
United States
Phone: 1 206 928 8263
Email: apratap@uw.edu

Abstract

Background: Most people with mental health disorders fail to receive timely access to adequate care. US Hispanic/Latino individuals are particularly underrepresented in mental health care and are historically a very difficult population to recruit into clinical trials; however, they have increasing access to mobile technology, with over 75% owning a smartphone. This technology has the potential to overcome known barriers to accessing and utilizing traditional assessment and treatment approaches.

Objective: This study aimed to compare recruitment and engagement in a fully remote trial of individuals with depression who either self-identify as Hispanic/Latino or not. A secondary aim was to assess treatment outcomes in these individuals using three different self-guided mobile apps: iPST (based on evidence-based therapeutic principles from problem-solving therapy, PST), Project Evolution (EVO; a cognitive training app based on cognitive neuroscience principles), and health tips (a health information app that served as an information control).

Methods: We recruited Spanish and English speaking participants through social media platforms, internet-based advertisements, and traditional fliers in select locations in each state across the United States. Assessment and self-guided treatment was conducted on each participant's smartphone or tablet. We enrolled 389 Hispanic/Latino and 637 non-Hispanic/Latino adults with mild to moderate depression as determined by Patient Health Questionnaire-9 (PHQ-9) score≥5 or related functional impairment. Participants were first asked about their preferences among the three apps and then randomized to their top two choices. Outcomes were depressive symptom severity (measured using PHQ-9) and functional impairment (assessed with Sheehan Disability Scale), collected over 3 months. Engagement in the study was assessed based on the number of times participants completed active surveys.

Results: We screened 4502 participants and enrolled 1040 participants from throughout the United States over 6 months, yielding a sample of 348 active users. Long-term engagement surfaced as a key issue among Hispanic/Latino participants, who dropped from the study 2 weeks earlier than their non-Hispanic/Latino counterparts (P<.02). No significant differences were observed for treatment outcomes between those identifying as Hispanic/Latino or not. Although depressive symptoms improved (beta=–2.66, P=.006) over the treatment course, outcomes did not vary by treatment app.

Conclusions: Fully remote mobile-based studies can attract a diverse participant pool including people from traditionally underserved communities in mental health care and research (here, Hispanic/Latino individuals). However, keeping participants engaged in this type of “low-touch” research study remains challenging. Hispanic/Latino populations may be less willing to use

XSL•FO
RenderX

mobile apps for assessing and managing depression. Future research endeavors should use a user-centered design to determine the role of mobile apps in the assessment and treatment of depression for this population, app features they would be interested in using, and strategies for long-term engagement.

Trial Registration: Clinicaltrials.gov NCT01808976; https://clinicaltrials.gov/ct2/show/NCT01808976 (Archived by WebCite at http://www.webcitation.org/70xI3ILkz)

(J Med Internet Res 2018;20(8):e10130) doi:10.2196/10130

KEYWORDS

mobile apps; smartphone; depression; Hispanics; Latinos; clinical trial; cognition; problem solving; mHealth; minority groups

Introduction

Technology is being leveraged as a way to perform large-scale clinical research targeting typically underrepresented populations. Given the extensive use of mobile devices across communities, remote research methods are becoming widely used. Additionally, technology is also seen as a potential method for bridging health disparities, which are typically driven by limited resources and stigma most apparent in minority communities. Of particular interest is the Hispanic/Latino community: Although they comprise one of the fastest-growing demographic segments in the United States [1], Hispanic/Latino populations are half as likely as their non-Hispanic white counterparts to receive mental health services [2]. This population is very difficult to recruit into research [3,4], and as a result, there is limited science to support treatment recommendations for this population. Recruitment of Hispanic/Latino samples into clinical research is particularly challenging in studies of mental health.

The widespread availability of digital technology has the potential to drive a sea change in access to psychosocial treatment for mental health problems in Hispanic/Latino communities [5]. Internet-based interventions have already demonstrated comparable treatment outcomes as traditional face-to-face psychotherapy [6], and given that 75% of Hispanic/Latino individuals own a smartphone [1], mobile-based mental health apps have the potential to increase treatment accessibility and engagement. Although there is potential for treating depression in Hispanic/Latino individuals using mobile devices, there is relatively little information about how this population interacts with apps, given their underrepresentation in mental health research. In particular, whether Hispanic/Latino smartphone owners (including both Spanish and English speakers) actually use mental health apps, and when they do, whether they follow the app protocols. We recently tested similar questions among a majority non-Hispanic white sample in a recent, fully remote trial (BRIGHTEN V1 [7,8]) and found that their interest in depression apps was high. It was far less challenging to recruit participants into our remote clinical trial compared with traditional in-person treatment trials. However, long-term engagement with the assigned apps trailed off significantly each week in the study, a finding that has been demonstrated in other studies [9]. However, Hispanic/Latino individuals, especially non-English speakers, do not typically have the same opportunity as majority groups to utilize mental health services and therefore may find mental health apps a useful alternative to traditional care. There is an immediate need for further research to develop and evaluate new solutions for mental health care for this population that are economically viable, scalable, and focused on engaging users to inform timely and evidence-based clinical interventions.

Therefore, the aim of this study was to determine the feasibility of conducting remote research with a Hispanic/Latino adult sample of smartphone users, how they interact with depression apps, and the potential clinical impact mHealth apps may have on treating depression in this population. We report recruitment, engagement, and cost in this 12-week, fully remote randomized controlled trial among Hispanic/Latino individuals with depression and a cohort of non-Hispanic/Latinos with depression to act as a direct comparator group (and extend our previous findings).

Methods

Approval

Ethical approval for the trial (NCT01808976) was granted by the Institutional Review Board of University of California, San Francisco. Specific research methods for this project replicated the BRIGHTEN V1 study and are described elsewhere [7,8], but are summarized here. Briefly, this was a fully remote treatment trial for depression, consisting of engagement with one of three treatment apps and periodic assessments detailed below.

Recruitment

Three different types of recruitment approaches, including traditional, social networking, and search-engine strategies, were used (Figure 1). Traditional methods consisted of Craigslist.org postings throughout the United States, specifically posting to the “Volunteer” and “Jobs etc” pages within Craigslist in at least one major city in every state. Social networking methods included regular postings on sites such as Facebook and Twitter and contextual-targeting methods to identify and directly push recruitment ads to potential participants, based on their Twitter and other social media comments. This approach was led entirely by Trialspark.com, which designed specific recruiting campaigns using machine learning approaches to create optimal advertising. Furthermore, we reached out to Hispanic/Latino Catholic Ministries in at least one city in every state to see if they would be willing to help with the recruitment for this study and post fliers in their communities. Each approach (described further in Multimedia Appendix 1 provided potentially interested participants a link to our custom study website, which was translated entirely for Spanish speakers and

XSL•FO
RenderX

included a welcome video featuring bilingual Hispanic/Latino researchers describing the goal of this study in Spanish. All translations involving text in the treatment apps were done by a combination of native Spanish speakers associated with this study and professionals at Babble-on.

Procedures

This study used an equipoise stratified clinical trial design [10], which factors participant preferences for treatment into randomization. Participants were randomly assigned one app among their two preferred intervention types and were asked to use it daily for 4 weeks. Participants completed primary outcome assessments, including the Patient Health Questionnaire-9 (PHQ-9) [11] and Sheehan Disability Scale (SDS) [12] once a week for 3 months, with other secondary measures (described below) completed at daily, weekly, or biweekly intervals. All treatments and assessments were delivered remotely via custom apps.

Screening

Interested participants completed a brief Web-based screening consisting of questions about their ability to speak Spanish ("Do you speak Spanish?; ¿Hablas Español?") and mobile device ownership ("Do you have an iPhone or Android smartphone?").

Consent

Participants were given the University of California, San Francisco consent form to read and were instructed to watch a video that highlighted the goals and procedures of the study, as well as risks and benefits of participation. After viewing the video, participants had to pass a quiz that confirmed their understanding that participation was voluntary, was not a substitute for treatment, and that they were to be randomized to treatment conditions. Each question had to be answered correctly before moving on to baseline assessment and randomization. Eligibility was established after consent was obtained. Upon being eligible, participants were sent a link to download their assessment app (Surveytory).

Participant Eligibility

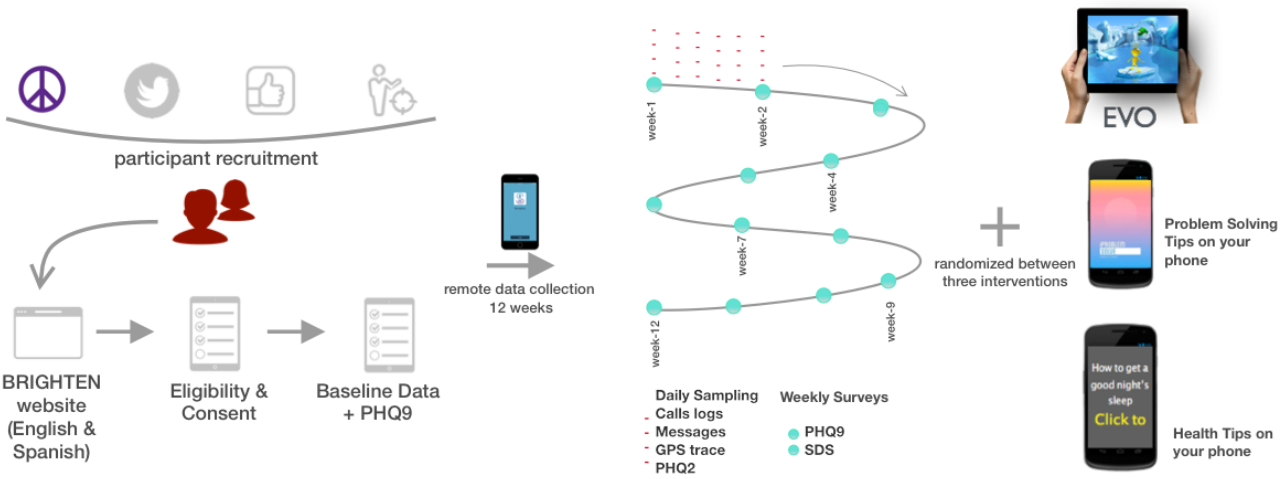
Participants had to speak English or Spanish, be 18 years old or older, and own either an iPhone with Wi-Fi or 3G/4G/LTE capabilities or an Android phone along with an Apple iPad version 2.0 or newer device. An iOS-based device was required as one of our intervention apps was only available on iOS devices at the time of the study. If a user had an Android phone, he or she was only eligible to participate if he or she also owned an Apple iPad version 2 or newer iOS tablet device. Participants had to endorse clinically significant symptoms of depression, as indicated by either a score of 5 or higher on PHQ-9 or a score of 2 or greater on PHQ item 10 (indicating feeling disabled in his or her life because of mood).

Assessment

Baseline

The baseline assessment included collecting demographic variables including age, race/ethnicity, marital and employment status, income, education, smartphone ownership, use of other health apps, and use of mental health services, including use of medications and psychotherapy. We collected information on mental health status using PHQ-9 [11] for depression and SDS [12] to assess self-reported disability. PHQ-9 rates the presence and severity of depressive symptoms across 9 items, with higher scores signifying more severe symptomatology (range 0-27). This is a reliable and well-validated screening instrument [13] that is responsive to depression treatment outcomes over time [11] and is included in the US Preventive Services Task Force recommendations for depression screening in adults [14]. PHQ-9 has been translated into several languages; we used both the original English language form and the validated Spanish translation [15]. The baseline PHQ-9 demonstrated good internal consistency in our sample (Cronbach alpha=.85, 95% CI 0.83-0.87).

Figure 1. Overall BRIGHTEN V2 study schematic showing participant recruitment, consent, enrollment, and randomization workflow along with weekly and daily data collection. EVO: Project Evolution; GPS: Global Positioning System; PHQ-2: 2-item Patient Health Questionnaire; PHQ-9: 9-item Patient Health Questionnaire; SDS: Sheehan Disability Scale.



XSL•FO
RenderX

SDS assesses perceived functional impairment across 3 domains (work/school, social life, and family/home responsibilities), yielding a sum score of 0-30, in which higher scores represent greater disability. SDS is popular in clinical trials given its sensitivity in detecting treatment effects [16]. As one of the official World Health Organization's measures of disability, this measure has also been translated into several languages; we used both the original English version and a validated Spanish translation of this scale [17]. SDS also demonstrated good internal consistency in our sample (Cronbach alpha=.89, 95% CI 0.87-0.91).

Follow-Up Assessments

Our custom mobile app, Surveytory, was used to collect all outcome and passive data. The assessments to measure changes in mood (PHQ-9) and disability (SDS) were administered weekly. Daily changes in mood were assessed using the PHQ-2 survey. Passive data collection included daily phone usage logs (call/text time, call duration, and text length) and mobility data (activity type and distance traveled using the phone's accelerometer and Global Positioning System). Participants were automatically notified every 8 hours for 24 hours if they had not completed a survey within 8 hours of its original delivery. A built-in reminder also prompted the participant to check for any surveys on a daily basis in case they missed a new survey notification. An assessment was considered missing if it was not completed within a 24-hour time frame.

Treatment

After confirming completion of baseline assessments (or 72 hours after the initiation of these assessments, whichever came first), participants were sent a Web-based survey that described each of the 3 treatment arms. Following this description, participants were asked to select which 2 apps they were most inclined to use in this study. Participants were then randomly assigned to one of these 2 preferred conditions and sent a link to download the intervention app, which included a brief video explaining how to download and use the assigned treatment app. This download also included a custom dashboard to monitor their study progress. Participants were asked to use their assigned app for 1 month.

The first app was a video game-inspired cognitive intervention (Project Evolution, EVO) designed to modulate cognitive control abilities, as declines in these abilities have been associated with depression [18]. This intervention has preliminary evidence for being an effective treatment for depression [18]. The second intervention was an app based on internet-based problem-solving therapy (iPST), an evidence-based treatment for depression, which has been shown to be both acceptable and efficacious for US-dwelling Hispanic/Latino populations. The final intervention app, an information control, provided daily health tips (HTips) for overcoming depressed mood such as self-care (eg, taking a shower) or physical activity (eg, taking a walk; see [8] for further descriptions of each).

Each of the 3 apps represented the most common type of self-guided depression apps available at the time of the study: apps based on psychotherapy principles, apps that claim to improve mood through therapeutic games, and apps that provide suggestions for mindfulness and behavioral exercises. Similar to the assessment notifications, each intervention app was equipped with built-in reminders asking the participant to use their app on a daily basis (reminders were sent once daily).

Incentives

Randomized participants were paid a total of US $75 in Amazon gift vouchers for completing all assessments over the 12 weeks. Participants received US $15 for completing the initial baseline assessment and an additional US $20 for each subsequent assessment at the 4-, 8-, and 12-week time points.

Procedures to Reduce Gaming

"Gaming" is a situation where a user enrolls in a study solely to acquire research payment or attempts to influence specific methodological aspects of the study. We utilized the following safeguards to prevent this: (1) locking the eligibility or treatment randomization survey if a participant tried to change a submitted answer so that only the initial answer was utilized, (2) using study links that are valid for one user/device, and (3) tracking internet protocol addresses to minimize duplicate enrollments.

Statistical Analyses

Participant self-reported race/ethnicity was used to create 2 groups of Hispanic/Latino and non-Hispanic/Latino adults (eg, all other races and ethnicities) to test our main study aims. Sample demographics and clinical characteristics were calculated using appropriate descriptive statistics. Comparisons between participant demographics were done using a chi-square test of independence for categorical variables and one-way analysis of variance to compare continuous variables across the groups. To assess the marginal effect (ie, association in the entire sample) between longitudinal weekly PHQ-9 and SDS scores and treatment arms, we used generalized estimating equations (GEEs) [19]. Briefly, GEE models extend generalized linear models to longitudinal or clustered data. GEEs use a working correlation structure that accounts for within-subject correlations of participant responses, thereby estimating robust and unbiased SEs compared with ordinary least squares regression [19,20]. We adjusted for age and gender to account for any potential confounding effects between outcome and main covariates of interest. Treatment response was further categorized into 3 groups based on a change of at least 5 points on PHQ-9 [11], the minimal clinically important difference [11], to comprise treatment responders (decrease PHQ-9≥5 points), nonresponders (change in PHQ-9<5 points), and those that deteriorated over treatment (increase in PHQ-9≥5 points). To assess participant engagement, we examined the proportion of participants who completed at least one activity in any given week. One-way analysis of variance was used to compare the daily, weekly, and overall participation differences between Hispanic/Latino and other participants. Univariate estimation of time to drop out from the study between Hispanic/Latino and non-Hispanic/Latino participants was computed using survival analysis. The distribution of survival days (total days active in the study) and nonparametric estimates of the survivor function was computed using the Kaplan-Meier method [21]. Log-rank test [22] was used to test for differences in survival between Hispanic/Latino and other participants. To compare dropout rates among the 3 interventions, a nonparametric Kruskal-Wallis test was used. Passive data was only used to compare user

XSL•FO
RenderX

engagement with active survey-based tasks. Given this study design is similar to that of our previous work [7], we used the same power analysis for this study. It indicated that 200 participants per intervention arm would provide 0.80 power to detect a medium treatment effect (eg, 2 points change on PHQ-9 scale, Cohen d=0.4) with an assumption of 50% participant dropout. However, this study was a feasibility trial of an understudied Hispanic/Latino population and was not sufficiently powered to detect a moderate effect size across the 3 interventions. All analyses were carried out using R (R Core Team, Vienna, Austria), statistical computing language version 3.4.2 [23].

Results

Recruitment and Enrollment

The BRIGHTEN V2 study started recruitment in August 2016 with screening and enrollment continuing for 7 months. A total of 4502 people were screened, and 23.10% (1040/4502) adults met the eligibility criteria and were enrolled in the study. Of these, 37.40% (389/1040) reported being Hispanic/Latinos. As in BRIGHTEN V1 study [7,8], the use of Craigslist.org was the most effective approach in recruiting, with more than 80% (843/1040) of our participants coming from this approach. An additional 8% (86/1040) were referred by friends or colleagues.

Enrolled participants lived throughout the United States, with all the metropolitan areas represented (Figure 2). Only 33.46% (348/1040) of the initially enrolled participants were active in the study (active cohort), as defined by completing at least one postenrollment weekly PHO-9 assessments or providing passive phone usage data within the first 12 weeks. The remaining 66.54% (692/1040) participants did not respond to any postenrollment surveys or provide passive data and were therefore considered to be study dropouts (Figure 3). Income, education, and race were significantly different between those who dropped and those who did not (P<.005; Multimedia Appendix 1). A large proportion of individuals who reported that they “can’t make ends meet” with regard to their income dropped out of the study (238/692, 34.4%) this effect was more pronounced for Hispanic/Latino individuals (135/283, 47.7%). Over half (171/283, 60.4%) of the Hispanic/Latino participants who dropped out of the study reported making US $20,000 or less annually compared to with 28.10% (112/398) of non-Hispanic/Latinos who dropped out. Of the 348 active individuals, 74 did not complete the treatment randomization survey and thus were not assigned an intervention. However, they continued to complete self-report surveys during the study period. For this reason, we categorized these participants as enrolled but not randomized (EnR) category. All further analyses were restricted to active individuals consisting of those in treatment (n=274) or EnR (n=74; total N=348) arms. See Figure 3 for the Consolidated Standards of Reporting Trials diagram illustrating participant flow through the study.

Of those who were randomized, 31.8% (87/274) attempted to change their assigned intervention by hitting the “back” button to return to the randomization page, while an additional 10.2% (28/274) participants returned to the survey a second time to change their preferences (9/274, 3.2%) of these individuals used both methods). Note that these attempts were unsuccessful because participant randomization was determined by the first answer given by a participant, and not any of the subsequent attempts made.

Sample Demographics

See Table 1 for participant characteristics, including comparisons across those identifying as Hispanic/Latino and not. The participants were predominantly young, with 69.81% (238/345) aged less than 40 years (mean 34.90, SD 10.92); female (205/266, 77.19%); and non-Hispanic white (98/184, 53.3%), with 30.7% (33/106) of our sample reporting Hispanic/Latino identity. The majority (168/241, 69.9%) reported some form of employment, and 87.8% (266/303) of all participants were iPhone users. There were significant differences between Hispanic/Latino and non-Hispanic/Latino participants; notably, a greater proportion (43/106, 40.6%) of Hispanic/Latino participants reported annual incomes of less than US $20,000, compared with only 24.7% (59/239) non-Hispanic/Latinos. Likewise, non-Hispanic/Latino participants were significantly more likely to be employed and more likely to have obtained a university education relative to Hispanic/Latino participants. Finally, Hispanic/Latino participants were slightly younger than their counterparts, although both groups were on average in their early-to-mid 30s.

Clinical Characteristics

Overall, the cohort reported moderate depressive symptomatology with a mean baseline PHQ-9 of 13.61 (SD 5.46). There was no difference in baseline depression between Hispanic/Latino and non-Hispanic/Latino participants (P=.07), and neither age nor gender showed a significant association with baseline PHQ-9 scores (age: =−0.09, P=.06; gender: F_{1336}=3.16, P=.07). Income satisfaction showed a moderate effect on baseline PHQ-9 scores (f^2=0.265, P<.001). Table 2 summarizes the associations and effect sizes of all baseline variables with baseline PHQ-9 scores. Participants who reported income satisfaction as “can’t make ends meet” showed significantly higher depression symptomatology (delta PHQ-9=+3.9, P<.001) than those who reported income level as “comfortable” (Figure 4). However, this discrepancy in depressive symptoms between income levels was not significantly different between Hispanic/Latinos and non-Hispanic/Latinos across other categories of income satisfaction.

XSL•FO
RenderX

Figure 2. US map showing the location of people who were screened (gray) and enrolled (red) in the BRIGHTEN V2 Study.

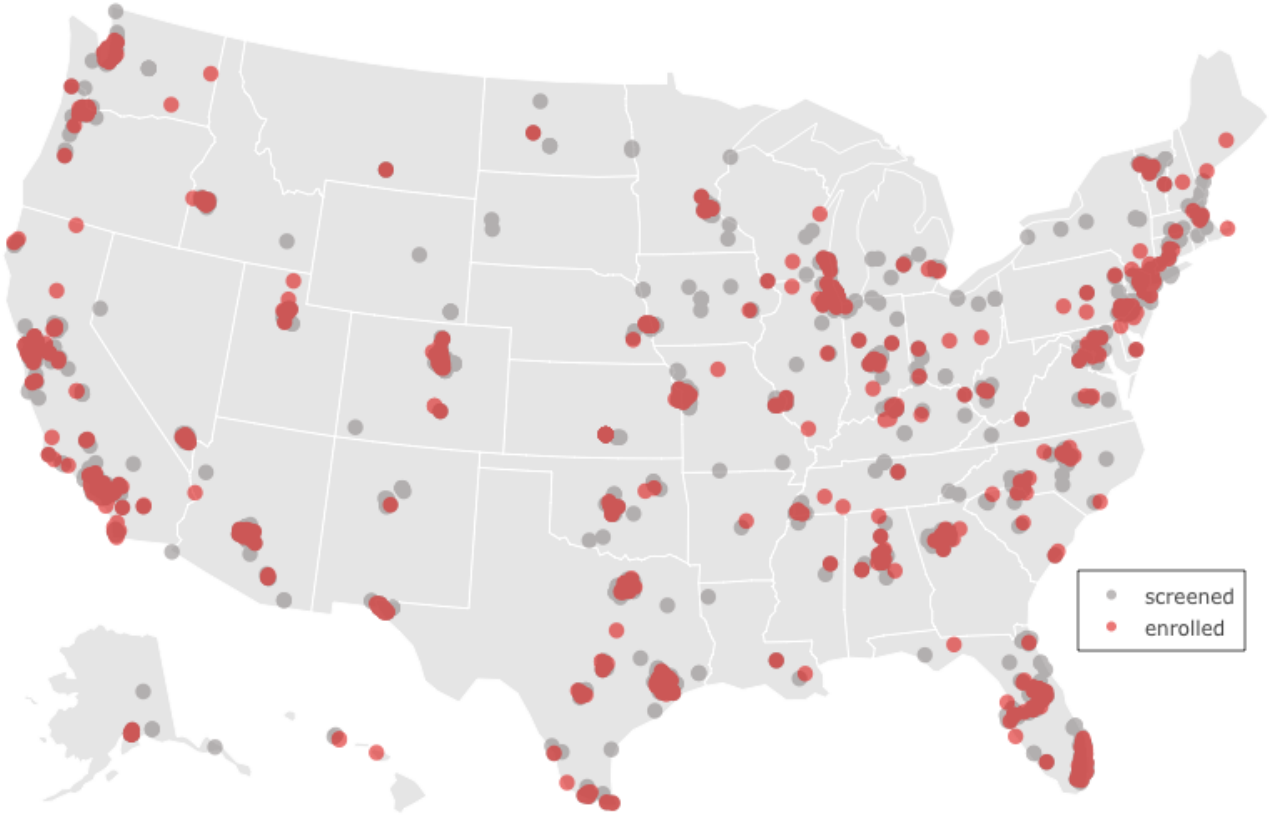
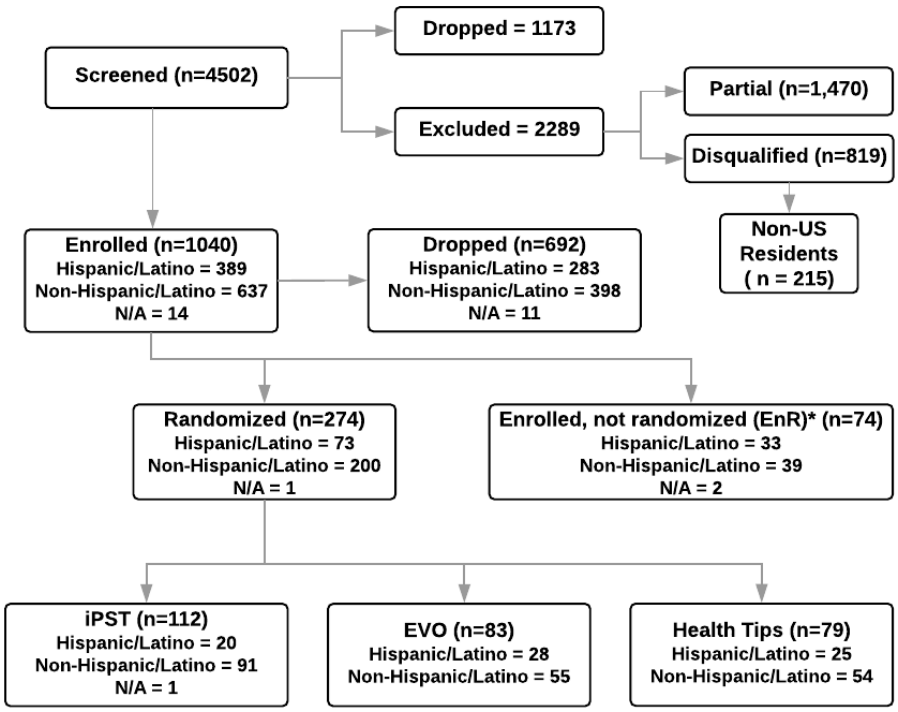


Figure 3. The Consolidated Standards of Reporting Trials flow diagram. iPST: internet-based problem-solving therapy; EVO: Project Evolution; N/A: not available.



XSL•FO
RenderX

Table 1. BRIGHTEN V2 participant characteristics.

Characteristics	Overall[a] (N=345)	Hispanic/Latino (n=106)	Non-Hispanic/Latino (n=239)	*P* value
Baseline Patient Health Questionnaire-9, mean (SD)	13.61 (5.46)	14.41 (5.69)	13.26 (5.34)	.08
Gender (female), n (%)	266 (77.1)	82 (77.4)	184 (77.0)	>.99
Age (years), mean (SD)	34.90 (10.92)	32.71 (10.10)	35.88 (11.15)	.02
Age (years), n (%)				.22
18-30	137 (40.2)	51 (48.6)	86 (36.4)	
31-40	101 (29.6)	27 (25.7)	74 (31.4)	
41-50	74 (21.7)	22 (21.0)	52 (22.0)	
51-60	23 (6.7)	5 (4.8)	18 (7.6)	
61-70	5 (1.5)	0 (0.0)	5 (2.1)	
>70	1 (0.3)	0 (0.0)	1 (0.4)	
Income last year (US $), n (%)				.005
20,000 or less	102 (29.6)	43 (40.6)	59 (24.7)	
20,000-40,000	90 (26.1)	31 (29.2)	59 (24.7)	
40,000-60,000	76 (22.0)	20 (18.9)	56 (23.4)	
60,000-80,000	32 (9.3)	5 (4.7)	27 (11.3)	
80,000-100,000	22 (6.4)	2 (1.9)	20 (8.4)	
100,000	23 (6.7)	5 (4.7)	18 (7.5)	
Education, n (%)				<.001
Community college	72 (20.9)	25 (23.6)	47 (19.7)	
Graduate degree	58 (16.8)	11 (10.4)	47 (19.7)	
High school	56 (16.2)	29 (27.4)	27 (11.3)	
University	159 (46.1)	41 (38.7)	118 (49.4)	
Device (iPhone), n (%)	303 (87.8)	89 (84.0)	214 (89.5)	.20
Working (Yes), n (%)	241 (69.9)	65 (61.3)	176 (73.6)	.03
Race, n (%)				<.001
Hispanic/Latinos	106 (30.7)	106 (100.0)	0 (0.0)	
Non-Hispanic white	184 (53.3)	0 (0.0)	184 (77.0)	
African-American/black	25 (7.2)	0 (0.0)	25 (10.5)	
American Indian/Alaskan Native	3 (0.9)	0 (0.0)	3 (1.3)	
Asian	24 (7.0)	0 (0.0)	24 (10.0)	
Other	3 (0.9)	0 (0.0)	3 (1.3)	
Speak Spanish (yes), n (%)	113 (32.8)	96 (90.6)	17 (7.1)	<.001
Income satisfaction, n (%)				.09
Comfortable	71 (20.6)	17 (16.0)	54 (22.6)	
Can't make ends meet	80 (23.2)	32 (30.2)	48 (20.1)	
Have enough to get along	194 (56.2)	57 (53.8)	137 (57.3)	
Marital status, n (%)				.28
Married/Partnered	135 (39.1)	35 (33.0)	100 (41.8)	
Separated/Widowed/Divorced	33 (9.6)	12 (11.3)	21 (8.8)	
Single	177 (51.3)	59 (55.7)	118 (49.4)	

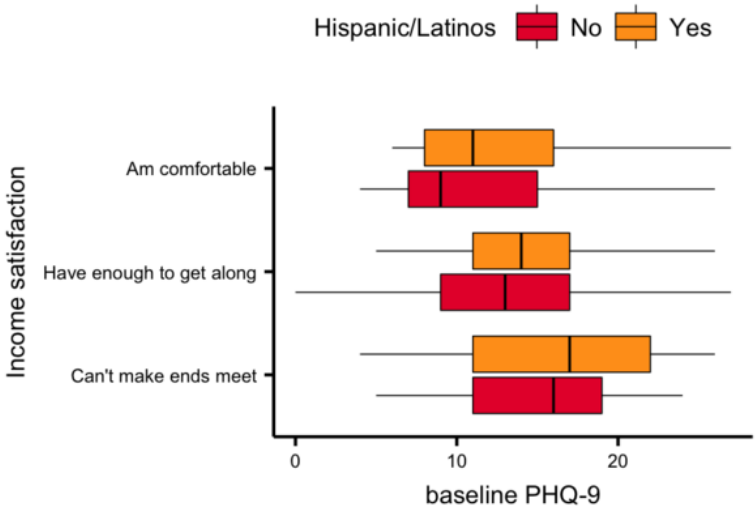
[a]Participants who did not self-report Hispanic/Latinos status (n=3) have not been compared.

XSL•FO
RenderX

Table 2. Association between baseline demographic variables and Patient Health Questionnaire-9 scores.

Baseline variables	Cohen f^2	False Discovery Rate
Income satisfaction	0.264	<0.001
Income	0.226	0.02
Spanish speaker	0.139	0.029
Education	0.160	0.076
Working	0.103	0.096
Hispanic/Latinos	0.098	0.101
Marital status	0.107	0.15
Race	0.161	0.15

Figure 4. Comparison of self-reported income satisfaction and baseline Patient Health Questionnaire-9 (PHQ-9) score between Hispanic/Latino and non-Hispanic/Latino participants.



Cost

Study costs beyond the initial infrastructure developed for BRIGHTEN V1 included participant payments (US $7540), website/enrollment portal/database development (US $4601), and total recruitment efforts (US $14,471; see Table 3). A bulk of recruitment spending was for 217 Spanish language ads placed on Craigslist throughout the country (US $5725), while only US $946 was spent on 33 English ads to obtain the reported enrollment. Furthermore, US $7800 was spent on targeted social media recruitment specifically for Spanish speakers via Trialspark.com; however, only 86 unique registrants came through this portal. Thus, participant acquisition costs differed dramatically between Spanish (US $31 per enrolled participant) and English speakers (US $1.49 per enrolled participant).

Engagement

Overall participation in the study (as measured by assessment completion, as opposed to intervention app use) decreased by approximately 50% from week 1 to week 4, with more than 4 out of 5 participants dropping (14%) out by the end of 12 weeks. At week 4, participants contributed twice as much passive data (ie, momentary Global Positioning System data) compared with that provided in survey assessments requiring active participation (Figure 5). Significant differences in participant engagement were observed between Hispanic/Latino and non-Hispanic/Latino participants (P=.02). Non-Hispanic/Latino individuals tended to participate in the study for 18.5 days longer than their Hispanic/Latino counterparts (median 53.5 days until dropout for non-Hispanic/Latinos and median 37 for Hispanic/Latino participants; see Figure 6). Finally, participants in the iPST and HTips arms were significantly more engaged than those in the EVO and EnR arms (P<.01), regardless of the race/ethnicity (Figure 7).

Depression Outcomes

Changes in weekly PHQ-9 scores were significantly associated with baseline severity of depressive symptoms (ie, mild, moderate, and severe; P<.001). Participants who reported severe depressive symptoms upon study entry evidenced the greatest decline in PHQ-9 scores during weeks 1-4 (beta=−4.19, P<.001) but no significant changes during weeks 5-12. Participants with moderate symptoms also showed an initial decline in PHQ-9 (beta=−1.96, P=.004) and a further decline of 0.70 points (beta=−2.66, P=.006) in weeks 5-12 (Table 4,Figure 8). With regard to treatment remission at the end of week 4, 34.42% participants responded to the interventions (a decrease in PHQ-9 score of ≥5 from baseline), 51.63% were nonresponders (change in PHQ-9 of <5 points), and a small proportion (11.48%) deteriorated (PHQ-9 worsened ≥5 points) during the course of

XSL•FO
RenderX

the study. However, there was no difference in depression outcomes among the 3 intervention arms. No differences in treatment remission were observed between Hispanic/Latino participants and non-Hispanic/Latinos.

Disability Outcomes

At the cohort level, disability based on SDS ratings decreased by an average 0.74 points (*P*=.03) in weeks 2-4 and further by 0.39 points (beta=−1.09, *P*=.02) in weeks 5-12. As with depression outcomes, there was no difference in disability outcomes across treatment arms. Hispanic/Latino and non-Hispanic/Latino participants did not differ in their disability outcomes (Table 5).

Table 3. Participant acquisition costs.

Recruitment approach	Amount spent (US $)	Participants reached, n	Cost per participant (US $)
Targeted Social Media (trialspark.com for Spanish Speakers)	7800	86	90.70
Craigslist.com (Spanish advertisements)	5275	303	17.41
Craigslist.com (English advertisements)	946	637	1.49

Figure 5. Comparison of participant attrition in the study across survey types and passive data stratified by Hispanic/Latinos and Non-Hispanic/Latinos. GPS: Global Positioning System; PHQ-2: 2-item Patient Health Questionnaire; PHQ-9: 9-item Patient Health Questionnaire; SDS: Sheehan Disability Scale.

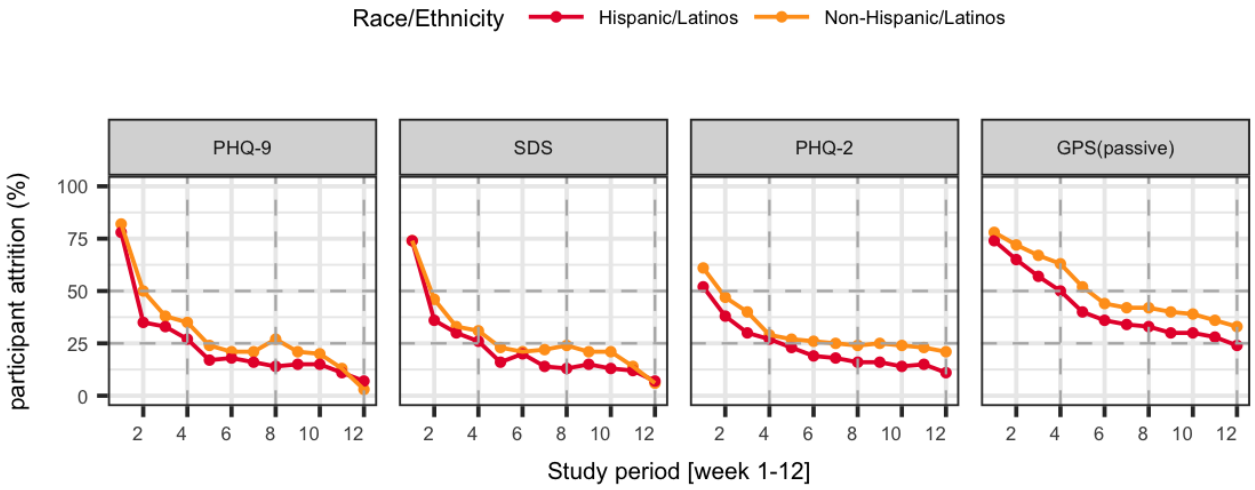
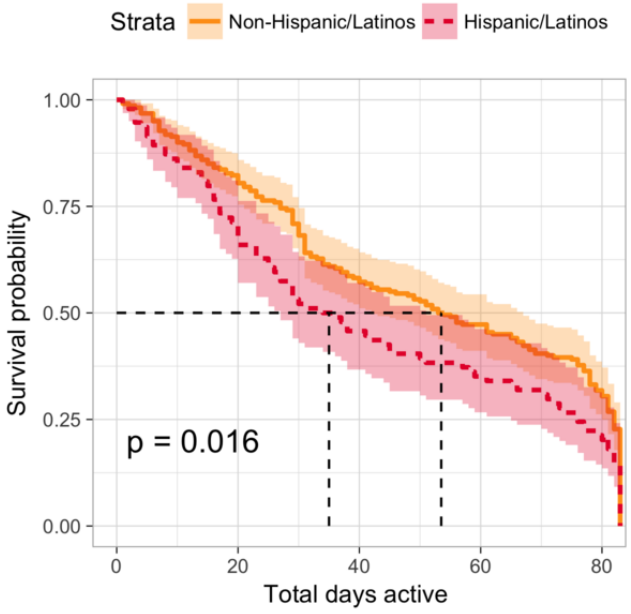


Figure 6. Comparison of Kaplan-Meier survival estimates for Hispanic/Latino and non-Hispanic/Latino participants during the course of the study (1-84) days.



XSL•FO
RenderX

Figure 7. Comparison of number of days participants were active across different treatment arms in the study. EnR: enrolled but not randomized; EVO: Project Evolution; HTips: health tips; iPST: internet-based problem-solving therapy.

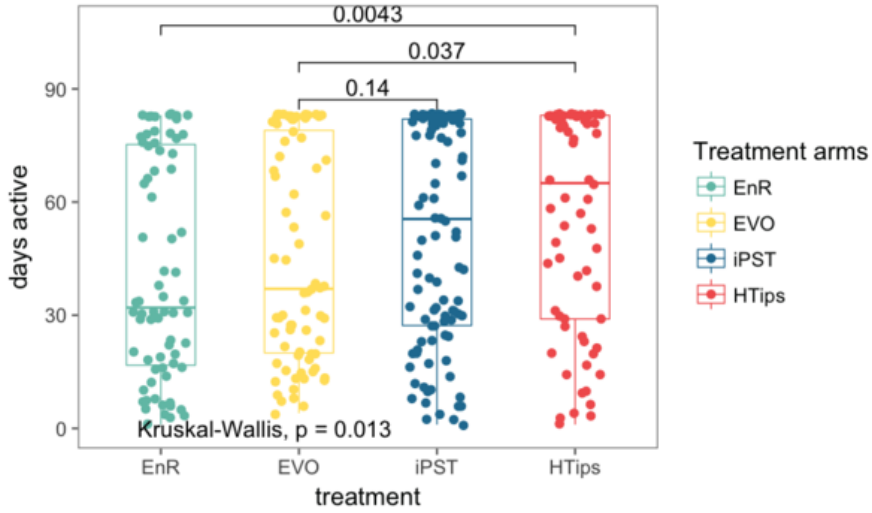


Table 4. Summary of estimates comparing weekly change in Patient Health Questionnaire-9 scores using a generalized estimating equations model.

Fixed effects	Effect size, beta (SE)	*P* value
Intercept	8.28 (0.77)	<.001
Gender (male)	.09 (0.50)	.85
Age	−.02 (0.02)	.23
Weeks 1-4	1.33 (0.55)	.02
Weeks 5-12	1.33 (0.72)	.06
Treatment (EVO[a])	.03 (0.57)	.96
Treatment (HTips[b])	−.93 (0.56)	.09
Treatment (iPST[c])	−.39 (0.53)	.45
Hispanic/Latinos (yes)	−0.15 (0.43)	.73
Baseline state (moderate)	5.35 (0.39)	<.001
Baseline state (severe)	12.26 (0.46)	<.001
Weeks 1-4: baseline state (moderate)	−1.96 (0.67)	.004
Weeks 5-12: baseline state (moderate)	−2.66 (0.96)	.006
Weeks 1-4: baseline state (severe)	−4.19 (0.77)	<.001
Weeks 5-12: baseline state (severe)	−4.31 (1.04)	<.001

[a]EVO: Project Evolution.

[b]HTips: health tips.

[c]iPST: internet-based problem-solving therapy.

XSL•FO
RenderX

Figure 8. Comparison of weekly mean Patient Health Questionnaire-9 (PHQ-9) scores with mean SEs stratified by baseline depression state.

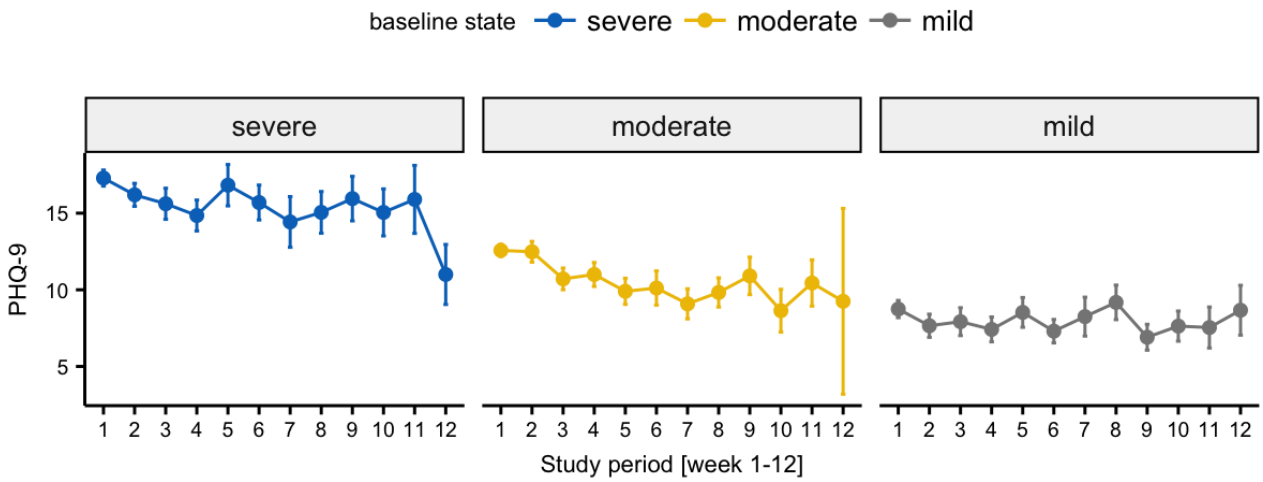


Table 5. Summary of estimates comparing weekly change in Sheehan Disability Scale score using a generalized estimating equations model.

Fixed effects	Effect size, beta (SE)	*P* value
Intercept	10.91 (1.61)	<.001
Gender (male)	.64 (0.85)	.46
Age	.00 (0.04)	.89
Treatment (EVO[a])	.32 (1.14)	.78
Treatment (HTips[b])	−.74 (1.07)	.49
Treatment (iPST[c])	−.12 (1.04)	.91
Weeks 2-4	−.70 (0.33)	.03
Weeks 5-12	−1.09 (0.47)	.02
Hispanic/Latinos (yes)	.12 (0.82)	.88

[a]EVO: Project Evolution.

[b]HTips: health tips.

[c]iPST: internet-based problem-solving therapy.

Discussion

Principal Findings

To our knowledge, BRIGHTEN V2 is the first large-scale effort to target the remote recruitment of Hispanic/Latino individuals with depression in the United States using digital health assessments and interventions that were translated into Spanish and administered solely on smartphones. We screened and enrolled one of the largest cohorts of Hispanic/Latino individuals with depression to date. Previous work has suggested that the lack of utilization of mental health care could be attributed to (1) cultural beliefs about mental health problems, (2) ineffective and inappropriate therapies, or (3) access problems or other barriers [24]. We attempted to address each of these issues by selectively targeting an underrepresented Hispanic/Latino population and using accessible, Spanish translated versions of the evidence-based intervention apps used in the initial study [8]. As has been found in other mobile-based mental health clinical trials [25,26], long-term engagement continues to be a significant challenge to these studies and is more pronounced among Hispanic/Latino participants. Although mobile devices are increasingly available in Hispanic/Latino communities [10], the availability of these devices as a means for conducting research and delivering care are not yet solutions that offset the widespread disparities seen in this population.

Feasibility and Acceptability

Similar to our previous work [7,27], this study has shown the feasibility of recruiting and enrolling a large and diverse sample of Hispanics/Latino adults. Previous research and observations from clinical practice suggest that Hispanics/Latino populations in the United States face barriers to research and treatment, including stigma and time constraints. This study was intended to overcome those very barriers by leveraging mobile apps that could be used at each participant's convenience. However, the engagement data showed that the Hispanics/Latino participants dropped out close to 2 weeks earlier than their non-Hispanics/Latinos counterparts, highlighting significant challenges in not only recruiting but also in keeping this population engaged. It was much more expensive and labor intensive to recruit Hispanics/Latino participants relative to the rest of the cohort. Attrition was particularly striking among the Hispanic/Latino subset, with only 18.7% (73/389) downloading

XSL•FO
RenderX

the treatment app. Highest dropout among the Hispanic/Latino sample were from participants reporting an annual income level of less than US $20,000.

Potential issues recruiting US Hispanic/Latino individuals for mental health research may hinge on (1) reluctance to be randomized, given the high number of the enrolled participants who tried to switch the initial randomly assigned intervention app and (2) privacy concerns such as the possibility that some of our lower income participants could be sharing the smartphones with other family members, potentially reducing the willingness to participate and causing high initial dropout [28]. Furthermore, the majority of participants were iPhone users, which may not be representative of the underlying population. While the ownership of an Android smartphone plus an iPad combination is relatively common as indicated by a 2014 survey [9], the ease of being able to participate in this study by only having to have a single device (iOS phone) likely spurred the bias toward iOS users in the sample.

Another potential issue in the study was the possible delay in receiving the intervention. The stratified equipoise randomization occurred after eligible participants attempted the assigned assessments (or after 72 hours, whichever came first); given that participants may have been waiting for their assigned intervention following their initial exposure to the assessment app, they may have lost interest in participating. Another consideration involves the appropriate incentive structure (eg, timing and amount of compensation) to maximize retention and engagement, as this factor is not well understood among such underrepresented samples such as ours. It is an empirical question to understand how the amount of payment affects one's participation in a given trial. Indeed, in the first version of this study (BRIGHTEN V1), we found that participants who received bonus payments remained in the study longer than those who did not receive bonuses [8]. In that study, the experimentation with two distinct incentive models to encourage retention revealed that participant payment was not enough to keep engagement from waning. Other work has shown that externalized benefits (eg, compensation) can dull motivation, whereas the creation of an internalized reward structure can enhance motivation and improve the aspects of adherence (eg, individualized presentation of study progress, personalized encouragements) [29,30]. This is a considerable hurdle to overcome for mental health researchers who are dependent upon trying to identify features that would align with greater engagement of a culturally unique population. Thus, these issues of acceptability and engagement must be dealt with not only for research but also for any scalable intervention to take hold in routine clinical practice.

Despite the poor engagement of the active components in this study, it is clear from the findings (and those from other mobile-based studies) that there is still a tremendous potential to capture passive data from smartphone use. This form of data capture is much less burdensome as it does not require the user to actively engage with an app. If one only considers the passive data compliance versus that of the active surveys in our study, passive data offers a viable opportunity to develop an individualized digital baseline (digital fingerprint) and investigate deviations from baseline phone usage to behavioral fluctuations. However, using cohort-level signals in passive data to predict depression states remains modest at best [31-33], suggesting that this approach will likely require larger studies and pairing with an active task-based component for the most effective solution.

Difference in Clinical Features and Outcomes

Similar to our earlier findings in the original study [7], participants on average reported improvement in both depression and disability measures over time, regardless of treatment arm. However, more than half of the participants, regardless of their race/ethnicity, did not evidence any clinically meaningful change (PHQ-9 change of less than 5 points from baseline) or actually deteriorated according to their PHQ-9 scores (worsening of more than 5 points from baseline on PHQ-9) during the course of the study. It is important to note that the participants in our trial did not have a clinical diagnosis of depression, rather they endorsed at least a mild level of depressive symptomatology at baseline screening on PHQ-9. Moreover, treatment outcomes were based on self-report using this screening measure. Perhaps unsurprisingly, treatment response was strongest in those with greater depressive symptomatology at baseline. Thus, we interpret our clinical findings with caution, as this is not a clinical sample or an effectiveness trial, but rather a feasibility trial in a sample of potential interest for future remote interventions. We also noted overall poor engagement in this sample with significant demographic differences between our Hispanic/Latino and non-Hispanic/Latino participants. Hispanic/Latinos reported lower income, less income satisfaction, and lower education; such factors have been previously reported to be associated with an increased incidence of depression [34].

Conclusions and Future Directions

mHealth platforms have the potential to deliver on-demand and as needed assessment and intervention alternatives despite known barriers of time constraints, cost, stigma, and cultural and language differences. Although mHealth holds great promise for closing the treatment gap for underserved communities, recruitment and retention remain problematic in such populations, and more research is needed to figure out better engagement strategies to best leverage mobile apps (eg, appropriate incentive levels, culturally responsive content and notifications along with user-centered design approaches [35]). Like other contactless programs (eg, self-help interventions), it is difficult to keep users engaged in active components without therapists or other in-person supports [36]. However, the ubiquity and relative unobtrusive nature of smartphones lend itself to acquiring passive sensing data, even in the absence of engagement with active components of the research or intervention protocol.

Our study offers preliminary lessons learned from doing such work in an understudied sample of Hispanic/Latino smartphone users. Scaling these types of remote assessments and interventions will hinge on the acceptance of such technology by both care teams and patients. This will be a problem for future research using remote technologies at scale to recruit and engage targeted communities (eg, Hispanic/Latino adults with depression) and will depend on understanding the population's

XSL•FO
RenderX

needs and addressing barriers to using mental health interventions via mobile apps.

Acknowledgments

Support for this research was provided by the National Institute of Mental Health (PAA R34MH100466, T32MH0182607, K24MH074717; BNR T32MH073553) and the National Institute on Aging (JAA P30AG15272). The authors thank Thomas Egan and Tojo Chemmachel for their help with data collection and data monitoring; Cecilia & Joaquin Anguera (author JAA's parents) for their help with culturally relevant translations within each app, website, video, and survey presented; Diana Albert for assistance in Web design; Diego Castaneda & Alinne Barrera for their willingness to speak in our promotional video; and Elias Chaibub Neto for helpful insights during the data analysis phase. The authors would also like to especially thank all the participants whose time and efforts made this work possible. We would also like to thank the entire Akili Interactive team as well as Wow Labz (especially R Omanakuttan**)** for helping with data collection and partnering with us on this project.

Conflicts of Interest

AG is cofounder, chief science advisor, and shareholder of Akili Interactive Labs, a company that develops cognitive training software. AG has a patent for a game-based cognitive training intervention, "Enhancing cognition in the presence of distraction and/or interruption," on which the cognitive training app (Project: EVO) that was used in this study was based. No other author has any conflict of interest to report.

Multimedia Appendix 1

Comparison of demographic variables.

[PDF File (Adobe PDF File), 459KB - jmir_v20i8e10130_app1.pdf]

Multimedia Appendix 2

CONSORT - EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 681KB - jmir_v20i8e10130_app2.pdf]

References

1. Pew Research Center. 2018. Mobile Fact Sheet URL: http://www.pewinternet.org/fact-sheet/mobile/ [WebCite Cache ID 6xFSdgCEV]
2. Olfson M, Blanco C, Marcus SC. Treatment of Adult Depression in the United States. JAMA Intern Med 2016 Oct 01;176(10):1482-1491. [doi: 10.1001/jamainternmed.2016.5057] [Medline: 27571438]
3. Arevalo M, Heredia N, Krasny S, Rangel M, Gatus L, McNeill L, et al. Mexican-American perspectives on participation in clinical trials: A qualitative study. Contemp Clin Trials Commun 2016 Dec 15;4:52-57 [FREE Full text] [doi: 10.1016/j.conctc.2016.06.009] [Medline: 27570845]
4. Miranda J, Nakamura R, Bernal G. Including Ethnic Minorities in Mental Health Intervention Research: A Practical Approach to a Long-Standing Problem. In: Cult Med Psychiatry. US: Springer; Dec 2003:467-486.
5. Fairburn C, Patel V. The impact of digital technology on psychological treatments and their dissemination. Behav Res Ther 2017 Dec;88:19-25 [FREE Full text] [doi: 10.1016/j.brat.2016.08.012] [Medline: 28110672]
6. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
7. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The Use and Effectiveness of Mobile Apps for Depression: Results From a Fully Remote Clinical Trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
8. Anguera J, Jordan J, Castaneda D, Gazzaley A, Areán PA. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. BMJ Innov 2016 Jan;2(1):14-21 [FREE Full text] [doi: 10.1136/bmjinnov-2015-000098] [Medline: 27019745]
9. Dorsey E, Yvonne CY, McConnell M, Shaw S, Trister A, Friend S. The Use of Smartphones for Health Research. Acad Med 2017 Dec;92(2):157-160. [doi: 10.1097/ACM.0000000000001205] [Medline: 27119325]
10. Lavori P, Rush A, Wisniewski S, Alpert J, Fava M, Kupfer D, et al. Strengthening clinical effectiveness trials: equipoise-stratified randomization. Biol Psychiatry 2001 Nov 15;50(10):792-801. [Medline: 11720698]
11. Löwe B, Unützer J, Callahan C, Perkins A, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. Med Care 2004 Dec;42(12):1194-1201. [Medline: 15550799]

XSL•FO
RenderX

12. Leon A, Olfson M, Portera L, Farber L, Sheehan D. Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. Int J Psychiatry Med 1997;27(2):93-105. [doi: 10.2190/T8EM-C8YH-373N-1UWD] [Medline: 9565717]
13. Kroenke K, Spitzer R, Williams J. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
14. Final Recommendation Statement: Depression in Adults: Screening - US Preventive Services Task Force Internet. 1 Jan 1AD URL: https://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/depression-in-adults-screening1 [accessed 2018-07-23] [WebCite Cache ID 718WyU5iB]
15. Wulsin L, Somoza E, Heck J. The Feasibility of Using the Spanish PHQ-9 to Screen for Depression in Primary Care in Honduras. Prim Care Companion J Clin Psychiatry 2002 Oct;4(5):191-195 [FREE Full text] [Medline: 15014707]
16. Sheehan K, Sheehan D. Assessing treatment effects in clinical trials with the discan metric of the Sheehan Disability Scale. Int Clin Psychopharmacol 2008 Mar;23(2):70-83. [doi: 10.1097/YIC.0b013e3282f2b4d6] [Medline: 18301121]
17. Bobes J, Badía X, Luque A, García M, González MP, Dal-Ré R. [Validation of the Spanish version of the Liebowitz social anxiety scale, social anxiety and distress scale and Sheehan disability inventory for the evaluation of social phobia]. Med Clin (Barc) 1999 Apr 24;112(14):530-538. [Medline: 10363239]
18. Anguera J, Gunning F, Areán PA. Improving late life depression and cognitive control through the use of therapeutic video game technology: A proof-of-concept randomized trial. Depress Anxiety 2017 Dec;34(6):508-517. [doi: 10.1002/da.22588] [Medline: 28052513]
19. Liang K, Zeger S. Longitudinal Data Analysis Using Generalized Linear Models. Biometrika 1986;73:13.
20. Ballinger G. Using Generalized Estimating Equations for Longitudinal Data Analysis. Organizational Research Methods 2004;7:150.
21. Rich J, Neely J, Paniello R, Voelker C, Nussenbaum B, Wang E. A practical guide to understanding Kaplan-Meier curves. Otolaryngol Head Neck Surg 2010 Sep;143(3):331-336 [FREE Full text] [doi: 10.1016/j.otohns.2010.05.007] [Medline: 20723767]
22. Bland J. The logrank test. BMJ 2004;328:1073.
23. The R Project for Statistical Computing Internet. URL: https://www.r-project.org/ [accessed 2018-07-23] [WebCite Cache ID 718XxyMa5]
24. Vega W, Kolody B, Aguilar-Gaxiola S, Catalano R. Gaps in service utilization by Mexican Americans with mental health problems. Am J Psychiatry 1999 Jun;156(6):928-934. [doi: 10.1176/ajp.156.6.928] [Medline: 10360134]
25. Miranda J, Azocar F, Organista K, Muñoz RF, Lieberman A. Recruiting and retaining low-income Latinos in psychotherapy research. J Consult Clin Psychol 1996 Oct;64(5):868-874. [Medline: 8916613]
26. Brown G, Marshall M, Bower P, Woodham A, Waheed W. Barriers to recruiting ethnic minorities to mental health research: a systematic review. Int J Methods Psychiatr Res 2014;23:48.
27. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The Use and Effectiveness of Mobile Apps for Depression: Results From a Fully Remote Clinical Trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
28. Karlson A, Brush A, Schechter S. Can i borrow your phone? 2009 Apr 09 Presented at: International conference on Human factors in computing systems - CHI 09; April 04-09, 2009; Boston p. 2009. [doi: 10.1145/1518701.1518953]
29. Cruz M, Pincus H, Harman J, Reynolds C, Post E. Barriers to care-seeking for depressed African Americans. Int J Psychiatry Med 2008;38(1):71-80. [doi: 10.2190/PM.38.1.g] [Medline: 18624019]
30. Van ED. Psychotherapy with older adults: benefits and barriers. J Psychosoc Nurs Ment Health Serv 2006 Nov;44(11):28-33. [Medline: 17128677]
31. Saeb S, Zhang M, Karr C, Schueller S, Corden M, Kording K, et al. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. J Med Internet Res 2015 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
32. Saeb S, Lattie EG, Schueller SM, Kording KP, Mohr DC. The relationship between mobile phone location sensor data and depressive symptom severity. PeerJ 2016;4:e2537 [FREE Full text] [doi: 10.7717/peerj.2537] [Medline: 28344895]
33. Pratap A, Anguera J, Renn B, Neto E, Volponi J, Mooney S. The feasibility of using smartphones to assess and remediate depression in Hispanic/Latino individuals nationally. 2017 Sep 12 Presented at: Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the ACM International Symposium on Wearable Computers on - UbiComp ?17; Sep 11-15, 2017; Maui. [doi: 10.1145/3123024.3127877]
34. Lorant V, Deliège D, Eaton W, Robert A, Philippot P, Ansseau M. Socioeconomic inequalities in depression: a meta-analysis. Am J Epidemiol 2003 Jan 15;157(2):98-112. [Medline: 12522017]
35. Vredenburg K, Mao J, Smith P, Carey T. A survey of user-centered design practice. 2002 Apr 20 Presented at: Proceedings of the SIGCHI conference on Human factors in computing systems Changing our world, changing ourselves - CHI ?02; 2002; Minneapolis. [doi: 10.1145/503457.503460]
36. Aguilera A. Digital Technology and Mental Health Interventions: Opportunities and Challenges. Arbor 2015;191:a210.

XSL•FO
RenderX

Abbreviations

GEE: generalized estimating equations
EnR: enrolled but not randomized
EVO: Project Evolution
HTips: health tips
iPST: internet-based problem-solving therapy
PHQ-9: Patient Health Questionnaire-9
SDS: Sheehan Disability Scale

Edited by G Wadley, R Calvo, M Czerwinski, J Torous; submitted 15.02.18; peer-reviewed by G Wadley, K Schneider; comments to author 14.03.18; revised version received 08.06.18; accepted 10.06.18; published 09.08.18.

Please cite as:
Pratap A, Renn BN, Volponi J, Mooney SD, Gazzaley A, Arean PA, Anguera JA
Using Mobile Apps to Assess and Treat Depression in Hispanic and Latino Populations: Fully Remote Randomized Clinical Trial
J Med Internet Res 2018;20(8):e10130
URL: http://www.jmir.org/2018/8/e10130/
doi:10.2196/10130
PMID:30093372

©Abhishek Pratap, Brenna N Renn, Joshua Volponi, Sean D Mooney, Adam Gazzaley, Patricia A Arean, Joaquin A Anguera. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 09.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Smartphone Cognitive Behavioral Therapy as an Adjunct to Pharmacotherapy for Refractory Depression: Randomized Controlled Trial

Akio Mantani[1], MD, PhD; Tadashi Kato[2], MD; Toshi A Furukawa[3], MD, PhD[‡]; Masaru Horikoshi[4], PhD; Hissei Imai[3], MD, PhD; Takahiro Hiroe[5], MD, PhD; Bun Chino[6], MD, PhD; Tadashi Funayama[7], MD, PhD; Naohiro Yonemoto[8], MPH; Qi Zhou[9], PhD; Nao Kawanishi[10], PhD

[1]Mantani Mental Clinic, Hiroshima, Japan
[2]Aratama Kokorono Clinic, Nagoya, Japan
[3]Department of Health Promotion and Human Behavior, School of Public Health, Graduate School of Medicine, Kyoto University, Kyoto, Japan
[4]National Center of Neurology and Psychiatry, Kodaira, Japan
[5]Waseda Clinic, Kani, Japan
[6]Ginza Taimei Clinic, Tokyo, Japan
[7]Funayama Mental Clinic, Nagoya, Japan
[8]Department of Biostatistics, School of Public Health, Graduate School of Medicine, Kyoto University, Kyoto, Japan
[9]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada
[10]Advanced Telecommunications Research Institute International, Kyoto, Japan
[‡]the FLATT Investigators

Corresponding Author:
Toshi A Furukawa, MD, PhD
Department of Health Promotion and Human Behavior, School of Public Health, Graduate School of Medicine, Kyoto University
Yoshida Konoe-cho,
Sakyo-ku,
Kyoto, 606-8501
Japan
Phone: 81 75 753 9491
Email: furukawa@kuhp.kyoto-u.ac.jp

Related Article:

Correction of: http://www.jmir.org/2017/11/e373/

(J Med Internet Res 2018;20(8):e11702) doi:10.2196/11702

The authors of “Smartphone Cognitive Behavioral Therapy as an Adjunct to Pharmacotherapy for Refractory Depression: Randomized Controlled Trial” (J Med Internet Res 2017;19(11):e373) incorrectly listed “September 2, 2015” as the starting date of patient eligibility assessment. However, “September 2, 2014” is the correct date.

The correction will appear in the online version of the paper on the JMIR website on August 31, 2018, together with the publication of this correction notice. Because this was made after submission to PubMed, Pubmed Central, and other full-text repositories, the corrected article also has been re-submitted to those repositories.

XSL•FO
RenderX

Edited by G Eysenbach; submitted 25.07.18; this is a non–peer-reviewed article;accepted 17.08.18; published 31.08.18.

Please cite as:
Mantani A, Kato T, Furukawa TA, Horikoshi M, Imai H, Hiroe T, Chino B, Funayama T, Yonemoto N, Zhou Q, Kawanishi N
Correction: Smartphone Cognitive Behavioral Therapy as an Adjunct to Pharmacotherapy for Refractory Depression: Randomized Controlled Trial
J Med Internet Res 2018;20(8):e11702
URL: http://www.jmir.org/2018/8/e11702/
doi:10.2196/11702
PMID:30168797

©Akio Mantani, Tadashi Kato, Toshi A Furukawa, Masaru Horikoshi, Hissei Imai, Takahiro Hiroe, Bun Chino, Tadashi Funayama, Naohiro Yonemoto, Qi Zhou, Nao Kawanishi. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 31.08.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

XSL•FO
RenderX