Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 14 (2012), Issue 5 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Reviews

Original Paper

Electronic Symptom Reporting Between Patient and Provider for Improved Health Care Service Quality: A Systematic Review of Randomized Controlled Trials. Part 1: State of the Art

Monika Alise Johansen[1], MSc; Eva Henriksen[1], MSc; Alexander Horsch[2,3], PhD; Tibor Schuster[2], PhD; Gro K Rosvold Berntsen[1], MD, PhD

[1]Norwegian Centre for Integrated Care and Telemedicine, University Hospital of North Norway, Tromsø, Norway

[2]Institute of Medical Statistics and Epidemiology, Technische Universität München, Munich, Germany

[3]Department of Telemedicine and e-Health, Institute of Clinical Medicine, University of Tromsø, Tromsø, Norway

Corresponding Author:
Monika Alise Johansen, MSc
Norwegian Centre for Integrated Care and Telemedicine
University Hospital of North Norway
Sykehusveien 23
pb35
Tromsø, 9038
Norway
Phone: 47 906 84 407
Fax: 47 777 54 099
Email: monika.alise.johansen@telemed.no

Abstract

Background: Over the last two decades, the number of studies on electronic symptom reporting has increased greatly. However, the field is very heterogeneous: the choices of patient groups, health service innovations, and research targets seem to involve a broad range of foci. To move the field forward, it is necessary to build on work that has been done and direct further research to the areas holding most promise. Therefore, we conducted a comprehensive review of randomized controlled trials (RCTs) focusing on electronic communication between patient and provider to improve health care service quality, presented in two parts. Part 2 investigates the methodological quality and effects of the RCTs, and demonstrates some promising benefits of electronic symptom reporting.

Objective: To give a comprehensive overview of the most mature part of this emerging field regarding (1) patient groups, (2) health service innovations, and (3) research targets relevant to electronic symptom reporting.

Methods: We searched Medline, EMBASE, PsycINFO, Cochrane Central Register of Controlled Trials, and IEEE Xplore for original studies presented in English-language articles published from 1990 to November 2011. Inclusion criteria were RCTs of interventions where patients or parents reported health information electronically to the health care system for health care purposes and were given feedback.

Results: Of 642 records identified, we included 32 articles representing 29 studies. The included articles were published from 2002, with 24 published during the last 5 years. The following five patient groups were represented: respiratory and lung diseases (12 studies), cancer (6), psychiatry (6), cardiovascular (3), and diabetes (1). In addition to these, 1 study had a mix of three groups. All included studies, except 1, focused on long-term conditions. We identified four categories of health service innovations: consultation support (7 studies), monitoring with clinician support (12), self-management with clinician support (9), and therapy (1). Most of the research (21/29, 72%) was conducted within four combinations: consultation support innovation in the cancer group (5/29, 17%), monitoring innovation in the respiratory and lung diseases group (8/29, 28%), and self-management innovations in psychiatry (4/29, 14%) and in the respiratory and lung diseases group (4/29, 14%). Research targets in the consultation support studies focused on increased patient centeredness, while monitoring and self-management mainly aimed at documenting health benefits. All except 1 study aiming for reduced health care costs were in the monitoring group.

Conclusion: RCT-based research on electronic symptom reporting has developed enormously since 2002. Research including additional patient groups or new combinations of patient groups with the four identified health service innovations can be expected in the near future. We suggest that developing a generic model (not diagnosis specific) for electronic patient symptom reporting for long-term conditions may benefit the field.

(J Med Internet Res 2012;14(5):e118) doi:10.2196/jmir.2214

KEYWORDS

Electronic symptom reporting; physician-patient relationship; patient participation; shared decision making; review; consultation; monitoring; self-management

Introduction

This paper presents the first part of a comprehensive review of randomized controlled trials (RCTs) focusing on electronic communication between patient and provider to improve health care service quality. Part 1 presents an overview of patient groups, health service innovations, and research targets relevant to electronic symptom reporting. Part 2 examines the methodological quality of the RCTs and summarizes effects and benefits of electronic symptom reporting of the methodologically best RCT studies from the reported data [1].

Patients today, including the elderly and less-educated [2], are quite motivated to use electronic services [3-5]. A new approach is being taken in countries with high e-readiness [6], focusing on the patient-provider partnership and information technology to promote patient-centered health care [7,8] and shared decision making [9,10]. In this approach, a new concept to improve patient centeredness is emerging, reflected in the rapidly rising number of studies during the past few years [11]: patients or parents reporting symptoms or health information electronically [11]. The patient reports to health care personnel, an institution, or a system, where the receiver processes and interprets the data and provides feedback to the patient [11]. The purpose, in general, is improved health care service quality, for example, by improving or avoiding consultation [11].

Patients support the idea of previsit reporting electronically [12-15] and believe it will improve the level of care and effectiveness [13,14]. Wald et al reported that when 2027 primary care patients, who already had an account to the secure electronic health record-connected Internet patient portal, were invited to provide health information electronically before consultation, 70% actually did so [16]. Patients felt more prepared for the visit and that their provider had more accurate information about them [16]. On the health system level, trials of electronic symptom reporting suggest that it may be possible to substitute about one-third or more of face-to-face consultations in primary care settings [17,18]. Further examples of the benefits that patients, health care personnel, and the health care system can gain from these tools are provided in part 2 of this study [1].

Patient Groups, Research Targets, and Health Service Innovations

A preliminary review conducted in 2010, based on abstracts, found that most studies in the field were small in terms of number of patients involved and are best described as feasibility studies [11]. This also called attention to the impression that electronic symptom reporting seemed to be more relevant for some patient groups or health conditions, such as complex conditions where it is challenging to cover all relevant issues during one short visit [11]. Examples are cancer [19], asthma [20,21], congestive heart failure [22,23], pain [24], neurological disorders [25], and mental health issues [3,26,27]. On the other hand, electronic symptom reporting was also used for less-severe problems such as atopic eczema [28], for follow-up after surgery [11,29,30], and in general primary care settings [17,18].

However, the health service innovations and research targets seem to involve a broad range of foci with regard to choices of patient groups, technology, organizational implementation, and outcome measures [11]. In such a heterogeneous field it is difficult to assess which patient groups are most likely to benefit, which types of interventions are the most promising, and which outcomes are likely to be improved by the interventions. This is not surprising, since telemedicine and eHealth are complex systems representing a blend of many disciplines [31]. To move the field forward it is necessary to create a map of what has been examined so far and to encourage more research into the areas holding most promise and the areas that are still unknowns on the map. No systematic review has yet addressed this theme, to the best of our knowledge.

Objective

The overall aim of the review was to systematically assemble the knowledge focusing on electronic communication between patient and provider to improve health care service quality. We wanted to limit our work to the most mature stage of a complex intervention before taking a service into ordinary use, the RCTs [32,33].

The objective for this first part of the review was to create a comprehensive overview of the most mature part of the field and to clarify what has been investigated so far with regard to different patient groups, health service innovations, and research targets relevant to electronic symptom reporting. Patient groups refers to either health conditions or to health services in cases where the trial did not focus on a specific diagnosis.

Methods

The review in general followed the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) recommendations [34]. To further improve the quality, we consulted the Cochrane handbook [35] for data extraction. The group conducting the review has a multidisciplinary background, including experience in medical and epidemiological research (GB, AH, TS), RCT methodology and statistics (TS, GB, AH), telemedicine and medical informatics (MAJ, EH, AH, TS), theoretical knowledge of electronic symptom reporting (MAJ, EH), and experience from earlier review work (AH, GB, TS).

Inclusion and Exclusion Criteria

Studies had to meet the following inclusion criteria: (1) it had to be an original study, (2) patients or parents in the intervention group had to report symptoms or health information electronically, either to clinical health care personnel or to a system, where the receiver processed and interpreted the data for health care purposes and provided feedback (we accepted that the feedback did not have to be given electronically; the focus was on asynchronous systems that can be established within the health care system, including e-diaries and personal health records accessible to health care providers), (3) the information reported had to be about the patient symptoms and health status at the time of reporting or during the preceding few days, and (4) it had to be an RCT comparing electronic symptom reporting versus a control group where symptom or health information was not received by the health care professionals or systems. This means that the control group may have varied from standard care or waiting lists to control groups where patients reported their symptoms or health information electronically but where this health information was not received by the health care professionals or the interpreting systems.

Studies fulfilling one or more of the following criteria were excluded: (1) retrospective questionnaires, prevalence surveys, general screening on the Internet, and tests of medications, (2) electronic communication requiring the patient and health care personnel to be present simultaneously, for instance in a video conference or through instant messaging, (3) automatic biometric measurements, since these are defined as reporting of signs, not symptoms, and 4) voice diary.

Search Methods for Identification of Studies

We searched the following electronic literature databases: Medline, EMBASE, PsycINFO, the Cochrane Central Register of Controlled Trials, and IEEE Xplore. The search was limited to publications from 1990 (due to no knowledge of older publication within this field), human medicine, English language, and RCTs (for PsycINFO: Treatment Outcome/Randomized Clinical Trial). We restricted EMBASE searches to exclude records imported from Medline. The first search was conducted in May 2011, and the search was last updated in October and November 2011.

We reviewed known eligible publications to identify possible indexing terms and relevant search words. It was necessary to establish a comprehensive search for two reasons. First, this is a new area without any established terms defining the field. Second, medical and medical informatics expressions evolve over time, where new terms appear and traditional terms are replaced by more specific ones [36]. Scope, indexing, and thesaurus terms are not equivalent in each database [37]. Thus, we had to adapt the initial Medline search to the search in other databases, keeping them as close as possible to the initial search.

We accessed Medline, EMBASE, and PsycINFO through the Ovid interface. Cochrane and the Ovid searches were built around four search files (What, Who, Why, and How), with a logical OR within the files, and an AND between the files. The Medline search was based on medical subject headings (MeSH) and the Text Words (TW) field to search titles and abstract information. The What file consisted of 22 search terms, including 3 MeSH terms, for symptoms and synonyms, such as “health data” or “health information*”. The Who file searched for “patient*” and “parent*” plus 16 relevant MeSH terms. The Why file included 51 search terms, of which only 3 were MeSH terms, for “self-report*”, “pre-report*”, and synonyms. Finally, the How file contained 38 search terms, including 11 MeSH terms, for the possible technologies involved. The search strategies were pilot tested and modified several times to ensure that they identified eligible publications. The Medline search strategy and search terms can be found in Multimedia Appendix 1.

The IEEE Xplore search had to be constructed in a different way because the limitation to a maximum of 10 search terms and 6 wildcards made it impossible to reuse the advanced Ovid searches. Since IEEE Xplore in general included few RCT metadata, we conducted a search for “RCT* OR (randomi* AND control* AND trial*)”.

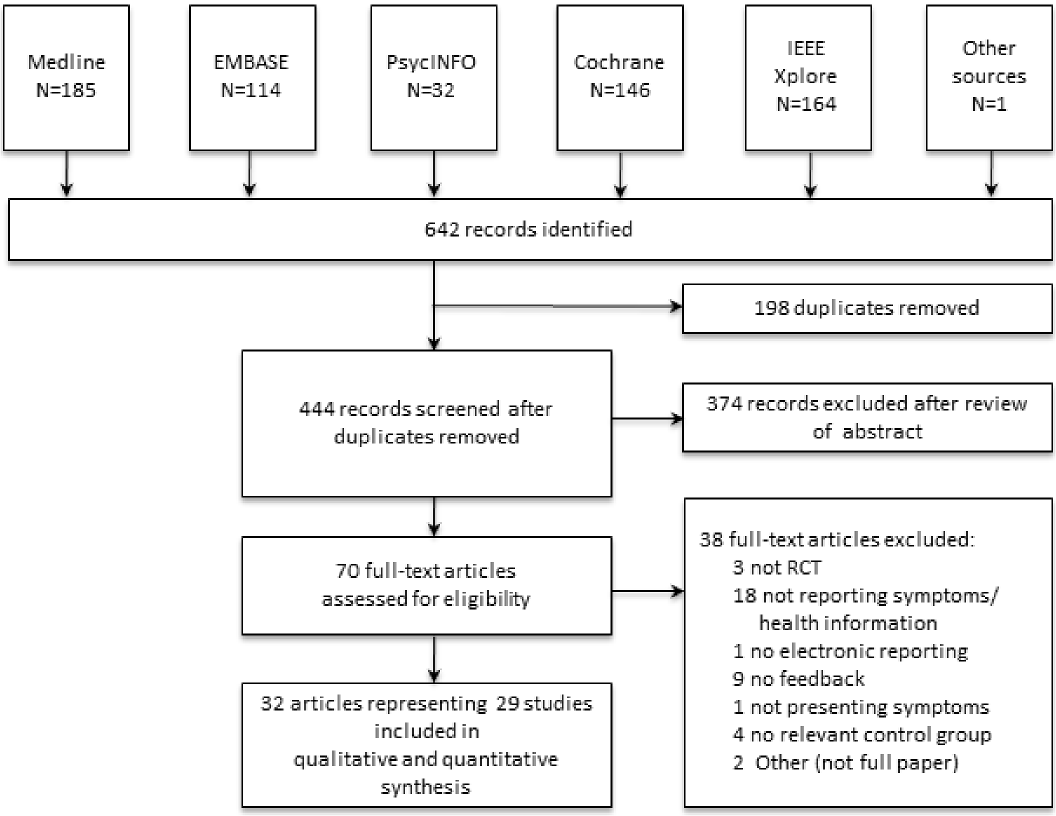
We did not include articles based on hand searches of reference lists, due to the Cochrane warning that “positive studies are more likely to be cited” and that “retrieving literature by scanning reference lists may thus produce a biased sample of studies” [35] (Cochrane 10.2.2.3, Citation bias). The only exception was if an article classified as relevant was a secondary analysis of an RCT, in which case we included the article presenting the primary analysis from the reference list.

Data Collection and Analysis

Selection of Studies

Search results were exported to EndNote X3 (Thomson Reuters, Carlsbad, CA, USA) for merging of databases, identification and deletion of duplicates, and review management. Abstract and full-text review were conducted independently, as presented in Figure 1, by two authors (MAJ and EH), who extracted data based on the inclusion and exclusion criteria into a structured spreadsheet. In the abstract review, we used only the information available in one specific abstract, and in the full-text review only the information available in one specific article, to determine eligibility for inclusion. All disagreements were resolved by consensus discussions. In a few cases, one author (GB) was consulted for full-text review and involved in the final conclusion.

Figure 1. Process for searching and selecting randomized controlled trials (RCTs) of electronic symptom reporting. The study flow diagram distinguishes between records and studies. A record is a source providing information about a study, presenting at a minimum an article title and abstract. Studies are the overall research projects themselves (here the RCTs), which may be represented by more than 1 article.



Combining Articles

Sometimes authors reported primary and secondary analyses from the same RCT in 2 separate articles. Other authors conducted a small RCT pilot before the main RCT. In these cases, we allowed both articles, if we judged both to be relevant, to be separately included and evaluated in the review. However, we counted and present them as 1 RCT study and 2 articles.

We linked articles deemed not to be relevant, but that published design, methodological, or theoretical information for an included study, to the included article when we extracted data and when presenting the characteristics of each study.

Data Extraction and Management

From each included article, authors MAJ or EH extracted variables, guided by the Cochrane data collection checklist [35] (Table 7.3.a in the Cochrane handbook), in addition to study-specific variables. These in total 84 variables represent (1) eligibility criteria, (2) study design and duration, (3) assessment of methodological quality including evaluation of the risk of bias in the results, (4) patient groups (either health conditions or health services in cases where the trial did not focus on specific diagnoses), (5) health services interventions and the corresponding control group, and (6) outcome measures and results relevant to electronic symptom reporting. A full presentation of the extracted variables and the citations can be found on the website of the Norwegian Centre for Integrated Care and Telemedicine [38].

Patient Groups, Health Service Innovations, and Research Targets

Due to the heterogeneity and complexity of the studies regarding patient groups, health service innovations, and research targets, and to provide a richer source of evidence, we chose an approach combining quantitative and qualitative narrative evaluation of the selected articles [39]. Qualitative methods are useful for exploring key domains in health service research [40]. The data were explored using content analysis to break them down into categories (or typologies) relevant to this review [41]. Under the qualitative analysis, we treated the text of individual studies more as a whole to identify major themes and categories, and then compared and contrasted them with those of the other studies [39].

The resulting patient groups depended on whether we found articles focusing on health services types without focusing on specific diagnoses. If we found only articles focusing on specific diagnoses, we defined the resulting patient groups by their health condition and categorized them as in our preliminary review [11], by the use of International Classification of Primary Care (ICPC) [42].

The research targets were classified according to the six areas of health service quality defined by the Institute of Medicine (IOM), which state that health care should be safe, effective in terms of health benefits (mortality, morbidity, and quality of life), patient centered, timely, efficient, and equitable [43]. Outcomes were in addition categorized according to who benefited: patients, health professionals, or the health care system. Table 1 shows the cross-link between who benefits, general clinical outcomes and the more specific outcomes variables extracted for this review, as well as the IOM outcomes.

Table 1. Research target typology: extracted outcomes grouped by who benefits from the intervention and Institute of Medicine (IOM) quality areas for health care [43]

Outcomes benefitting	Extracted outcome variables	IOM quality area
Patients	Clinical outcomes Improved health care service for patient Resource utilization for patient Satisfaction for patient Other benefits and results Unintended effects	Patient centeredness Health benefits Safety Timeliness for receiver
Health care professionals	Improved health care service for doctors and others Resource utilization for doctors and others Satisfaction for doctors and others Other benefits and results Unintended effects	Efficiency: resource utilization (for health professionals) Timeliness for health professionals
Health care system	Avoided consultations Other benefits and results Unintended effects	Efficiency: health care costs Efficiency: duration or time Equitability

Results

Selection of Studies

Of 642 records identified through the search and 444 abstracts reviewed, 32 articles and 29 studies were included (Figure 1) [44-75]. Three studies were reported in 2 articles, where 2 studies reported primary and secondary analyses in 2 separate articles ([59,60] and [73,74]), and 1 study had conducted a small RCT pilot before the main RCT [52,53].

The full-text review resulted in agreement on 49 articles, while we discussed 21 articles to reach consensus. Of these, 7 were finally included and 14 were excluded. The third author was involved in the discussion of 9 articles, where 3 were finally included and 6 excluded. Classification of abstracts from the database searches as not relevant or potentially relevant can be found in Multimedia Appendix 2.

Background Data

The 32 articles were published over 10 years, from 2002 to November 2011, with most (n = 24) published in the last 5 years. All 29 studies, except 2, were conducted in Western countries: 12 in the United States, 4 in the United Kingdom, 3 in the Netherlands, 2 in Sweden, 2 in Switzerland, and 1 each in Australia, Denmark, Norway, Singapore, Spain, and Taiwan. Four of the parallel studies had three arms [48,57,69,71]; all the others had two. All except 2 studies randomly allocated patients; the exceptions used cluster randomization: 1 randomized primary care practices [73,74], and 1 randomized clinics [49]. All studies included both genders with an average of 60% females (ranges 37.5%-93%). In nearly two-thirds (20/32) of the articles, the first author is female.

Patient Groups

We found no articles that did not focus on specific diagnoses. Thus, resulting patient groups were all defined by their health condition and categorized as in our preliminary review [11], mainly by the use of ICPC [42]. The exception is cancer, which is not a separate category in ICPC but is represented in a large and distinct body of the literature and is therefore presented separately.

The included articles resulted in five specific patient groups and one mixed group. Of the total of 29 studies, respiratory and lung diseases are clearly the largest group, with 12 studies (see Table 2).

All of the included studies, except that of Yardley et al [66], focused on long-term conditions or issues.

Health Service Innovations

In the content analysis of the included studies, we identified the following four categories of health service innovations: (1) consultation support, (2) monitoring with clinician support, (3) self-management with clinician support, and (4) therapy.

Consultation support includes patients or parents reporting symptoms or health information electronically prior to a consultation, where the main focus is how this affects the consultation. *Monitoring with clinician support* includes patients following a monitoring program. The patient reports measurements and health data, and a health care professional monitors the patient’s disease or treatment. *Self-management with clinician support* might include some monitoring elements, but most important is that patients in these studies have to follow a self-management program, with communication and supporting feedback from clinicians provided to the innovation group. Self-management programs focus on problem-solving skills to overcome barriers, making action plans and carrying them out, and education to increase patients’ confidence and their ability to manage their symptoms and illness [76,77]. The fourth category, *therapy*, is different from all the other categories in that it comprises innovations where the whole treatment, and all communication between therapists and patients, is conducted exclusively electronically. No specific self-management program or module is included in the therapy category.

We categorized studies that were difficult to categorize because they included elements from both monitoring and self-management according to our interpretation of the studies' main purpose. The following studies were categorized as monitoring but included some self-management elements: Chan et al [52,53], Jan et al [55], Rasmussen et al [57], Guendelman et al [54], and Nguyen et al 2009 [61]. On the other hand, Nguyen et al [68] and van der Meer et al [67] were categorized as self-management, but included some monitoring elements.

Table 2 presents the resulting health service innovations according to the resulting patient groups. The monitoring category is largest, including 12 studies, while self-management includes 9 studies, consultation 7, and therapy 1.

Most of the consultation support innovations were conducted in the cancer patient group (5/7), and most of the monitoring studies were in the respiratory and lung diseases group (8/12). In self-management, nearly half (4/9) of the studies were conducted in the field of psychiatry or in the respiratory and lung diseases patient group (4/9).

Table 2. Reviewed randomized controlled trials of electronic symptom reporting, by health service innovation category and patient group[a].

Patient group	Consultation support studies	Monitoring with clinician support studies	Self-management with clinician support studies	Therapy study	Total studies
Cancer	5 studies: Berry et al [44]; Boyes et al [45]; Ruland et al [46]; Ruland et al [47]; Velikova et al [48]	1 study: Kearney et al [51]	0	0	6
Respiratory and lung diseases: asthma	0	6 studies in 7 articles: Willems et al [58]; Chan et al [53] and Chan et al [52]; Guendelman et al [54]; Jan et al [55]; Prabhakaran et al [56]; Rasmussen et al [57]	1 study: van der Meer et al [67]	0	7
Respiratory and lung diseases: chronic obstructive pulmonary disease	0	2 studies in 3 articles: Lewis et al [59] (health care use) and Lewis et al [60] (quality of life); Nguyen et al [61]	1 study: Nguyen et al [68]	0	3
Respiratory and lung diseases: other	0	0	2 studies: DeVito Dabbs [65]; Yardley et al [66]	0	2
Cardiovascular diseases	0	3 studies: Carrasco et al [62]; Santamore et al [63]; Schwarz et al [64]	0	0	3
Psychiatry	1 study: Stevens et al [49]	0	4 studies: Berger et al [69]; Bergström et al [70]; Vernmark et al [71]; Oerlemans et al [72]	1 study: Wagner et al [75]	6
Diabetes	0	0	1 study in 2 articles: Williams et al [74] and Glasgow et al [73]	0	1
Mixed	1 study: Leveille et al [50]	0	0	0	1
Total studies	7	12	9	1	29

[a] Articles were identified in a comprehensive search in Medline, EMBASE, PsycINFO, Cochrane Central Register of Controlled Trials, and IEEE Xplore from 1990 to November 2011, and were published in the time period 2002–2011. References with *and* between them are articles that belong to the same study.

Table 3 presents the location of the patient at the time of communication and who the patient's main communication partner was [44-75,78-80]. Combining the results from Table 2 and Table 3 gives an overview of the main communication partner and the physical context of the patient's reporting situation within the different health innovation areas.

In the group of consultation support articles, symptom reporting was conducted at the clinic (one exception), and the physician, or both a physician and a nurse, was the main communication partner for the patient. In all the monitoring, self-management, and therapy articles, the patient was at home when reporting. The main communication partner was the nurse in monitoring studies (7 studies). In psychiatry self-management, psychologists were the main communication partners. A total of 3 self-management and 3 monitoring studies mainly used computer-tailored feedback to the patients, 5 with and 1 without nurse or physician support.

Table 3. Topic of reviewed randomized controlled trials of electronic symptom reporting, by patient's location at time of symptom reporting and main communication partner[a].

Main communication partner	Location of patient	
	Inside health care institution	Outside health care institution (at home)
Unclear	0	1 study: lung diseases [65][b]
Physician at hospital	3 studies: cancer [45] + [48]; psychiatry [49]	1 study: cardiovascular [63]
Physician and nurse	3 studies: cancer [44,78][c] + [46] + [47]	0
General practitioner or primary care physician	0	1 study: cardiovascular [62]
Psychologist	0	5 studies: psychiatry [75] + [69][d] + [70] + [71][d] + [72]
Nurse	0	9 studies: cancer [51,79][c]; asthma [58,80][c] + [52,53] + [56]; chronic obstructive pulmonary disease [59,60] + [61] + [68]; cardiovascular [64]; mixed [50]
CTF[e] only	0	1 study: lung diseases [66]
CTF and physician	0	2 studies: asthma [55] + [57]
CTF and nurse	0	3 studies: asthma [67] + [54]; diabetes [73,74]

[a] Articles were identified in a comprehensive search in Medline, EMBASE, PsycINFO, Cochrane Central Register of Controlled Trials, and IEEE Xplore from 1990 to November 2011, and were published in the time period 2002–2011.

[b] Patients communicated with a transplant team, a transplant provider, a coordinator, and a transplant coordinator at the hospital. The professions of these actors are not clearly defined [65].

[c] Articles 78, 79, and 80 were deemed not to be relevant, but included information necessary to understand the study in question.

[d] Conducted mainly by students [69,71] under the supervision of a more experienced or senior psychologist.

[e] Computer-tailored feedback.

Characteristics of Included Studies in Relation to Health Innovation Categories

The included studies are presented according to the resulting health service innovation categories (Table 4, Table 5, Table 6, and Table 7). The tables describe methods, participants and relevant considerations and components for replicating the intervention, according to Cochrane's minimum requirements [35] (11.2.2). In addition, the main findings column presents the results of individual studies, as recommended by PRISMA [34]. Since the studies are heterogeneous with respect to disease, interventions, and outcomes, the summary descriptions are not easily standardized. Thus, we produced a more detailed and comprehensive table than is common in most reviews.

The 7 consultation studies involved more patients per study than did the monitoring and self-management support studies: 2342 patients (range 52–878, median 241). The 12 monitoring studies included a total of 1824 patients (range 17–321, median 120). The 9 self-management studies included 2242 patients (range 50–886, median 88), and the therapy study included 55 patients; 10 studies included fewer than 100 patients.

Consultation studies generally followed patients through one consultation only, while the duration of other interventions varied from 1 to 12 months, where more than half lasted 4 months or less, 6 lasted between 6 and 8 months, and only 4 lasted as long as 12 months.

Table 4. Summary description of studies on consultation support in the reviewed randomized controlled trials of electronic symptom reporting[a].

Patient group	Trial and country[b]	Participant characteristics[c]	Study characteristics[d]	Health service innovation: consultation support	Main findings and research targets[e]
Cancer					
	Berry et al [44]; (Wolpin et al [78]); USA	262 clinicians from 2 clinics; 660 cancer patients, 18–86 (mean 54) years; female % not reported	Design: P + 2; inside clinic Duration: 2 visits (before treatment and 4–6 weeks later)	Enhancing patient–provider communication with electronic self-report assessment for cancer *Intervention: ESRA-C*[f]: a color graphical summary of the participant's self-reported symptoms and quality-of-life issues with predetermined thresholds flagged was printed and handed to the clinician immediately before the targeted clinic visit. No recommendations offered. *Control*: ESRA-C questionnaires were filled, but no summary was handed to the clinician.	*Berry et al [44]: Primary*: The likelihood of symptoms and quality-of-life issues being discussed between clinicians and patients differed by randomized group and depended on whether a symptoms and quality-of-life issue was first reported as problematic (*P* = .032). *Secondary:* Clinic visits were similar with regard to duration between groups, and clinicians reported the summary as useful. *Wolpin et al [78]*: The ESRA-C was easy for patients to use and acceptable across a range of user characteristics. *Research targets:* Patient centeredness, duration, resource utilization
	Boyes et al [45]; Australia	80 cancer patients, 20–85 years (mean not reported); female 59.5%	Design: P + 2; inside clinic Duration: before visit, 4 times	Effect of giving oncologist a summary of the cancer patient's self-reported psychosocial well-being *Intervention:* Touch-screen survey filled out before oncologist visit. Computer scored the answers and a printed summary report was placed in the patient's file for consideration during consultation. Suggested strategies for managing identified issues were included. *Control:* Touch-screen survey filled out, but no results made available to oncologist.	*Unclear primary outcome.* Intervention patients who reported a debilitating physical symptom at visit 2 were significantly less likely to report a debilitating physical symptom at visit 3 compared with control patients (odds ratio 2.8, *P* = .04). Reductions in levels of anxiety, depression, and perceived needs among intervention patients were not significantly different from those among control patients. *Research targets:* Health benefits, patient centeredness
	Ruland et al [46]; USA	14 physicians, 14 nurse practitioners; 52 cancer patients; 23–77 (mean 56.3) years; female 59%	Design: P + 2; inside clinic Duration: 1 consultation	Supporting shared decision making *All patients* scheduled for an outpatient visit used the system on a tablet computer to report their symptoms and preferences prior to consultation. The system highlighted for clinicians what symptoms patients were experiencing, including severity, degree of bother, and importance to patients. This information was printed and provided to the patient and clinician in the *experimental group* but not in the *control group*.	*Primary:* Significantly greater congruence between patients' reported symptoms and those addressed by their clinicians in the experimental group. *Secondary:* The system scored high on ease of use. There were no significant group differences in patient satisfaction. *Research target:* Patient centeredness

Patient group	Trial and country[b]	Participant characteristics[c]	Study characteristics[d]	Health service innovation: consultation support	Main findings and research targets[e]
	Ruland et al [47]; Norway	145 cancer patients (leukemia or lymphoma); ≥18 (mean in intervention: 50, in control: 49) years; female 38%	Design: P + 2; inside clinic Duration: up to 1 year (once per encounter during treatment, once per week during hospital stay, once per outpatient visit in up to 4 visits)	Effects of a computer-supported interactive tailored patient assessment tool Both groups used *Choice, an interactive tailored patient assessment*, touchpad tablet PC, for symptom assessments prior to inpatient and outpatient visits. The assessment summary, which displayed the patient's self-reported symptoms, problems, and distress in rank order of the patient's need for support, was provided to physicians and nurses in the *intervention group*. *Control group* patients used exactly the same tool, but the clinicians were not given any information from the patient's assessment.	*Primary:* Significantly more symptoms were addressed in the intervention group patient charts than in those of the control group. *Secondary:* Symptom distress in the intervention group decreased significantly over time in 11 (58%) of 19 symptom/problem categories versus 2 (10%) for the control group. Need for symptom management support over time also decreased significantly more for the intervention group than the control group in 13 (68%) symptom categories. *Research targets:* Patient centeredness, health benefits, resource utilization
	Velikova et al [48]; UK	28 physicians, 286 oncology patients; age range not reported; mean age 54.9 years; female 73%	Design: P + 3; inside clinic Duration: approximately 6 months	Improving communication and patient well-being *Intervention group* completed touch-screen HRQL[g] questionnaires in the waiting room before every encounter. A summary was presented to physicians. *Attention-control group* also completed HRQL questionnaires on touch-screen computer, but summary was not presented to physicians. *Control group* used no touch-screen measurement of HRQL before clinic encounters.	*Several primary outcomes:* Intervention and attention-control groups had better HRQL than the control group (*P* = .006, *P* = .01, respectively), but the intervention and attention-control groups were not significantly different (*P* = .80). A positive effect on emotional well-being was associated with data feedback (*P* = .008) but not with instrument completion (*P* = .12). A larger proportion of intervention patients showed clinically meaningful improvement in HRQL. More frequent discussion of chronic nonspecific symptoms (*P* = .03) was found in the intervention group, without prolonging encounters. The clinicians found the information useful. There was no detectable effect on patient management (*P* = .60). In the intervention patients, HRQL improvement was associated with explicit use of HRQL data (*P* = .016), discussion of pain, and role function (*P* = .046). *Research targets:* Health benefits, patient centeredness, duration, resource utilization

Patient group	Trial and country[b]	Participant characteristics[c]	Study characteristics[d]	Health service innovation: consultation support	Main findings and research targets[e]
Psychiatry	Stevens et al [49]; USA	878 potential behavioral concern patients from 9 clinics; 11–20 (mean 13.9) years; female 54%	Design: C + 2; inside clinic Duration: 1 consultation	Does screening increase clinicians' recognition of behavior concerns? The *Health eTouch system* collected self-report data from patients in the waiting rooms. At 5 sites, patients' screening results were printed and given to the primary care provider just before the face-to-face encounter (*immediate-results condition*). At 4 sites, the 1-page summary was mailed to the primary care provider 2–3 business days later (*delayed-results condition*). Providers had immediate access to screening results for youth reporting thoughts about suicide, regardless of group assignment.	*Primary:* In intent-to-treat analysis, difference approached but did not reach statistical significance (P = .058). However, if all youths who endorsed suicidal ideation, regardless of original condition assignment, were included in the immediate-results condition, then 68% of youths in the immediate-results condition who screened positive were identified as having a problem by their pediatrician, which was significantly higher than the recognition rate of 52% in the delayed-results condition (P = .001). *Research targets:* Health benefits, patient centeredness
Mixed	Leveille et al [50]; (Allen et al [81]); USA	34 physicians, 241 patients (depression, chronic pain, and mobility difficulty); 22–86 years (mean not reported); female 57%	Design: P+ 2; outside, probably at home Duration: from 4 weeks until index visit (start unclear).	Nurse coaching to promote patient–primary care physician discussion *Intervention: PatientSite* was used to enhance patient–provider communication regarding 3 common conditions (chronic pain, depression, and impaired mobility) during upcoming visits. Delivered online by nurse e-coaches, the intervention involved a standardized set of emails and worksheets targeting self-efficacy, patient education, and motivation to improve health. *Control:* Patients received a general message through PatientSite containing URL links to US government websites with general health information.	*Several primary outcomes:* Detection and treatment of the target conditions (1-week postvisit survey) and symptom burden related to these conditions. Similar high percentages of intervention (85%) and control (80%) participants reported discussing their condition during their primary care physician visit. More intervention than control patients reported their primary care physician gave them specific advice about their health (94% vs 84%; P = .03) and referred them to a specialist (51% vs 28%; P = .002). Intervention participants reported somewhat higher satisfaction than controls (P = .07). Results showed no significant differences in detection or management of screened conditions, symptom ratings, and quality of life between groups. *Research targets:* Health benefits, patient centeredness

[a] Articles were identified in a comprehensive search in Medline, EMBASE, PsycINFO, Cochrane Central Register of Controlled Trials, and IEEE Xplore from 1990 to November 2011, and were published in the time period 2002–2011.

[b] Main author, main reference. References in parentheses contributed relevant study information on the study in question.

[c] Number of clinicians, number of patients with diagnosis, age range (mean) of patients, percentage female patients.

[d] Design (P = parallel group design, CO = crossover, C = cluster, F = factorial, O = other, U = unclear; + number of study arms), where symptom reporting took place (outside or in the home; or inside a clinic), and duration of intervention.

[e] Main findings are in general presented as in the original article and refer to primary outcome if clearly defined and secondary outcomes considered relevant for the scope of the study. Research targets refers to the six areas of health service quality defined by the Institute of Medicine [43].

[f] Electronic Self-Report Assessment-Cancer.

[g] Health-related quality of life.

Table 5. Summary description of studies on monitoring in the reviewed randomized controlled trials of electronic symptom reporting[a].

Patient group	Trial and country[b]	Participant characteristics[c]	Study characteristics[d]	Health service innovation: monitoring	Main findings and research targets[e]
Cancer					
	Kearney et al [51]; (McCann et al [79]); UK	112 breast, lung, or colorectal cancer patients; >18 (mean 56) years; female 76.8%	Design: P + 2; outside/home Duration: 4 cycles of chemotherapy (12–16 weeks)	Management of chemotherapy-related toxicity *Intervention*: A mobile phone system (ASyMS) was used in the morning, evening, and at any time patients felt unwell on days 1–14 following their first 4 cycles of chemotherapy. Patients completed an electronic symptom questionnaire on their mobile, including reporting their temperature. Patients immediately received written feedback on the phone. Clinicians were advised to contact patients within 1 hour after receipt of a red alert. The system's alert to physician was based on a risk model. *Control:* Received standard care.	*Unclear primary outcome.* 2 of the 6 symptoms measured (fatigue and hand–foot syndrome) showed statistical significance between the 2 randomized groups: higher reports of fatigue in the control group and lower reports of hand–foot syndrome in the control group. *Research target:* Health benefits
Respiratory and lung disease					
	Chan et al [53]; and the small prestudy Chan et al [52]; USA	120 children with persistent asthma; 6-17 (mean in intervention: 10.2, in control: 9.0) years; female 37.5% (Chan et al [52]: 10 children)	Design: P + 2; outside/home Duration: 12 months	Internet-based monitoring and education of children with asthma *Intervention: The Asthma In-Home Monitoring* group received 3 in-person visits and Internet-based education. They reported asthma signs and symptoms daily. Peak flow videos were sent twice per week for 6 weeks and then once weekly. Case manager scored results based on standardized checklists. *Control:* Traditional in-person education and case management over 6 scheduled visits. *Both groups:* The case manager contacted patients by email (intervention) or telephone (control) twice per week for 6 weeks and once per week thereafter to review their information. The patients were able to contact the case manager by email (intervention) or telephone (control) whenever needed. Asthma education in both groups followed the same curriculum.	*Unclear primary outcome (both studies).* Virtual patients had higher metered-dose inhaler with valved holding chamber technique scores than did the office-based group at 52 weeks (94% vs 89%), had greater adherence to daily asthma symptom diary submission (35.4% vs 20.8%), had less participant time (636 vs 713 patient-months), and were older. Caregivers in both groups perceived an increase in quality of life and an increase in asthma knowledge scores from baseline. There were no other differences in therapeutic or disease control outcome measures. *Research targets:* Health benefits, patient centeredness, resource utilization

Patient group	Trial and country[b]	Participant characteristics[c]	Study characteristics[d]	Health service innovation: monitoring	Main findings and research targets[e]
	Guendelman et al [54]; USA	134 children with asthma; 8–16 (mean in intervention: 12, in control: 12.2) years; female in intervention: 60%; in control: 63%	Design: P + 2; outside/home Duration: 3 months	Asthma outcomes and self-management behaviors *Intervention: Health Buddy* enabled children to assess and monitor their asthma symptoms and quality of life daily and to transmit this information to a nurse. A protocol based on clinical practice guidelines consisting of 10 questions was designed. Patients answered daily queries from a nurse by pressing 1 of 4 buttons. Patients received immediate feedback from the Health Buddy. Asthma facts and trivia questions, which changed daily, were included to pique children's curiosity and enhance learning. *Control:* Participants used a standard asthma diary.	*Primary:* After adjusting for covariates, the odds of having any limitation in activity were significantly (P = .03) lower for Health Buddy children *Secondary:* The intervention group was also significantly (P = .01) less likely to report peak flow readings in the yellow or red zone or to make urgent calls to the hospital (P = .05). Self-care behaviors also improved far more for the intervention group. *Research targets:* Health benefits, patient centeredness
	Jan et al [55]; Taiwan	164 children with persistent asthma; 6–12 (mean in intervention: 10.9, in control: 9.9) years; female in intervention: 60.3%, in control: 63.2%	Design: P + 2; outside/home Duration: 3 months	Interactive asthma monitoring *Intervention:* With *Blue Angel for Asthma Kids*, children completed the electronic asthma diary and recorded symptoms, need for rescue medication, and PEF[f] values, preferably daily. The tool comprised a 3-color real-time warning system accompanied by a treatment plan. Patients were asked to follow instructions given by the computer and the physician; thereafter, the decision support system was used to check whether asthma had been brought under control. Physicians then instructed the patients by email or telephone to increase, decrease, or continue the usual treatment. (See Rasmussen et al [57] for a comparable intervention.) *Control:* Patients recorded the same PEF values and asthma symptoms on paper, and received the same optimal clinical care, education program (as part of usual care), and support from asthma management teams. Their written asthma diary was supplemented by instructions for self-management.	*Unclear primary outcome.* When the 2 groups were compared with regard to change from baseline, the children in the intervention group had a significant decrease of nighttime (P = .028) and daytime symptoms (P = .009) compared with the children in the control group. The adherence rates of therapeutic and diagnostic monitoring, global assessment of asthma control, knowledge of asthma self-management, and quality of life of caregivers were all significantly higher in the intervention group than in the conventional asthma care group. *Research targets:* Health benefits, patient centeredness

Patient group	Trial and country[b]	Participant characteristics[c]	Study characteristics[d]	Health service innovation: monitoring	Main findings and research targets[e]
	Prabhakaran et al [56]; Singapore	120 asthma patients; mean age in intervention: 37, in control: 40 years); female in intervention: 65%, in control: 53%	Design: P + 2; outside/home Duration: 3 months	Asthma monitoring *Intervention:* Patients received text messages to assist with asthma management, daily for 2 weeks, thereafter weekly for 10 weeks. New data were compared with previous results, and the frequency of reliever use was analyzed by the server receiving the data. If the value was too high (preset threshold) or the patient did not reply to 3 consecutive messages, an alert email was sent to the asthma nurse. All alerts were verified by the asthma nurse through telephone contact with the patients. *Control:* Patients had no text messaging support.	*Unclear primary outcome.* Asthma-control test scores improved for 36 participants in the intervention group compared with 28 in the control group (*P* = .113). Number of nebulizations decreased in 54 participants in the control group compared with 50 in the intervention group (*P* = .053). Emergency room visits decreased in 57 participants in the control group compared with 51 in the intervention group (*P* = .063). Admission rates did not decrease in either group (*P* = .5). The mean response rate to the messages was 82%, and 92% in the intervention group were satisfied with the text messaging service. *Research targets:* Health benefits, health care costs, patient centeredness
	Rasmussen et al [57]; Denmark	300 asthma patients; 18–45 (mean 29.5) years; female 69%	Design: P + 3; outside/home Duration: 6 months	Asthma monitoring *Intervention 1: Internet-based monitoring.* Patients completed an electronic diary and recorded symptoms, need for rescue medication, and PEF values, preferably daily. The Internet action plan calculated the level of asthma control and offered the patient advice on what to do next by using a 3-color warning system (green, yellow, and red). If the patient scored red, an email was sent to the physician. The physician used a decision support system to decide the level of treatment. Physician instructed patients by email or telephone. (See Jan et al [55] for a comparable intervention.) *Intervention 2: Specialist monitoring group* were taught how to use a peak flow meter and a written action plan daily (comprising a 3-color warning system based on the symptom score and PEF values) to adjust their medication. *Control:* In the *GP*[g] *group* the GP assessed the patient’s asthma symptoms and test results and from this decided the patient’s need for pharmaceutical treatment. The patients in the GP group did not receive any treatment or information about asthma from the study physician.	*Several primary outcomes:* Treatment and monitoring with the Internet-based management tool led to more significant improvement in the Internet group than in the other 2 groups regarding *asthma symptoms* (Internet vs specialist: odds ratio2.64, *P* = .002; Internet vs GP: odds ratio 3.26; *P* < .001), *quality of life* (Internet vs specialist: odds ratio 2.21, *P* = .03; Internet vs GP: odds ratio 2.10, *P* = .04), *lung function* (Internet vs specialist: odds ratio 3.26, *P* = .002; Internet vs GP: odds ratio 4.86, *P* < .001), and *airway responsiveness* (Internet) vs GP: odds ratio 3.06, *P* = .02). *Research target:* Health benefits

Patient group	Trial and country[b]	Participant characteristics[c]	Study characteristics[d]	Health service innovation: monitoring	Main findings and research targets[e]
	Willems et al [58]; (Willems et al [80]); the Netherlands	109 patients with mild to moderate asthma; 56 children 7–18 (mean 11) years, 53 adults ≥18 (mean 46) years; female 55.6%	Design: P + 2; outside/home Duration: 12 months	Nurse-led telemonitoring *Intervention:* Lung function values and symptoms registered at home twice daily on a portable handheld device (diaries) were transferred to the asthma nurse (main caregiver) monthly or when having asthma complaints. The nurse studied the data daily and classified the asthma following a stepwise intervention protocol. The nurse was allowed to decrease (after 3 months of stable asthma) or increase asthma medication by 1 step. Physician was consulted only if necessary. *Control:* Regular care.	*Willems et al [58]: Primary:* No significantly improved asthma-specific quality of life; *secondary*: no significant decrease of asthma symptoms or medical consumption (time and medication). *Willems et al [80]:* Higher mean health care costs per patient in the intervention group. A decrease in the price of the asthma monitor will substantially increase the probability of the program being cost effective. *Research targets:* Health benefits, resource utilization, health care costs
	Lewis et al [59] (Reduce health care use) and Lewis et al [60] (quality of life); UK	40 patients with moderate to severe COPD[h] who had completed at least 12 sessions of outpatient pulmonary rehabilitation; mean age in [59] in intervention: 67, in control: 70 years; mean age in [60] for intervention: 70, for control: 73 years; female 50%	Design: P + 2; outside/home Duration: 6 months + 6-month follow-up	Home telemonitors to reduce health care use [59], and improve quality of life for patients [60] *Intervention:* Patients received standard care plus *Docobo HealthHUB* handheld monitors at home for 26 weeks followed by 26 weeks of standard care (for observation and follow-up). During the monitoring period, patients recorded their symptoms and physical observations twice daily. Data were transmitted automatically at night via the home telephone line. Nurses could access the data through a website and receive alerting email messages if certain conditions were detected. *Control:* Standard care for 1 year.	*Reduced health care use: Primary:* No significant differences between the groups in hospital admissions [59]; *Secondary:* No significant differences between the groups in emergency room visits, days in hospital, or contacts with the specialist COPD community nurse team during the monitoring period, but fewer primary care contacts for chest problems ($P < .03$) in the intervention group [59]. After the monitors were removed, no significant differences were found between the groups for any of the health care contacts ($P > .20$ throughout) [59]. *Quality of life: Primary:* No significant differences in quality-of-life scores between the groups at any time, or consistently within either group over time [60]. *Research targets:* Health care costs, health benefits

Patient group	Trial and country[b]	Participant characteristics[c]	Study characteristics[d]	Health service innovation: monitoring	Main findings and research targets[e]
	Nguyen et al [61]; USA	17 patients with moderate to severe and stable COPD; mean 68 (SD 11) years; female 65%	Design: P + 2; outside/home Duration: 6 months	A cell phone-based exercise persistence intervention postrehabilitation for COPD *All participants* developed an individualized exercise plan, were issued a pedometer and exercise booklet, and were trained to log their daily exercise and symptoms. *Intervention: MOBILE-coached* patients submitted symptoms and exercise daily, and received immediate summary feedback from server and weekly reinforcement feedback by text message from nurse; reports of worsening symptoms were automatically flagged to the nurse for follow-up. *Control: MOBILE self-monitored:* Entered and submitted the same information on the cell phone, but no (information and) alarm to nurse and no coaching feedback from nurse.	*Unclear primary outcome.* Small feasibility study. Logging exercise and symptoms was easy, and keeping track of their exercise helped patients remain active. There were no significant differences between groups over time in maximal workload, 6-minute walk distance, or health-related quality of life ($P > .05$); however, MOBILE-self-monitored increased total steps per day, whereas MOBILE-coached logged fewer steps over 6 months ($P = .04$). *Research targets:* Health benefits, patient centeredness
Cardiovascular disease					
	Carrasco et al [62]; Spain	38 GPs, 285 hypertensive patients; (age range not reported), mean 62 years; female 40%	Design: P + 2; outside/home Duration: 6 months	Text message-based Patient–GP interaction on control of hypertension *Intervention:* Patients sent the mean results of blood pressure self-monitoring 4 times a week, and heart rate and body weight once a week. They could complete an optional questionnaire during each wireless application protocol session. GPs accessed the data sent via the Web and could send a text message regarding any related issue to the patient's phone. *Control:* Followed the same protocol, except that they recorded the data on paper and could only deliver it to their GP personally at the routine visits.	*Primary:* The influence of the interaction between patient and GP, in a nonspecialized setting, in the selected type of hypertensive patients, did not significantly improve the degree of hypertension control; *Secondary:* the course of hypertension during follow-up, adherence to the protocol, results of quality-of-life and anxiety questionnaires, or economic aspects such as the number of consultations or hospital admissions did not significantly improve. *Research targets:* Patient centeredness, health benefits, health care costs

Patient group	Trial and country[b]	Participant characteristics[c]	Study characteristics[d]	Health service innovation: monitoring	Main findings and research targets[e]
	Santamore et al [63]; USA	321 cardiovascular disease patients; 18–85 (mean in intervention: 62, in control: 63.2) years; female % not reported	Design: P + 2; outside/home Duration: 8 months	Telemedicine System to Decrease Cardiovascular Disease Risk *All patients* received a manometer with memory and a pedometer. *Intervention:* Patients exchanged data with their care provider via the Internet. Patient could enter data daily, or for several days at once. Patients reported weight, blood pressure and heart rate, physical activity (steps/day), and cigarette smoking, and received feedback on cardiovascular disease risk reduction. Data for 10–15 patients were presented simultaneously on the provider screen. Out-of-limits parameter (red) values were presented first. *Control group* received usual care plus manometer and pedometer.	*Unclear primary outcome.* Systolic and diastolic blood pressures decreased significantly in both groups. The decreases in systolic blood pressure were greater in the intervention group ($P < .05$). For both groups, low-density lipoprotein decreased and high-density lipoprotein remained unchanged. In diabetic patients, blood glucose and glycated hemoglobin decreased significantly ($P < .01$) only in the intervention group. In nondiabetic patients, the risk of diabetes and metabolic syndrome score decreased ($P < .01$) only in the intervention group. Rates of usage of the telemedicine system were very high (92%). This rate of self-monitoring greatly exceeded the self-monitoring rate in controls (48%). The telemedicine-entered blood pressure values were similar to the meter-recorded values and to the office values. *Research targets:* Health benefits, patient centeredness
	Schwarz et al [64]; USA	102 heart failure patients; 65–94 (mean 78.1) years; female 52%	Design: P + 2; outside/home Duration: 3 months	Telemonitoring of heart failure patients and their caregivers *Intervention: e-Cardiocom electronic home monitoring* system measured weight daily. The device asked the participants to answer yes or no to questions about symptoms. The heart failure care manager, an advanced practice nurse, was responsible for daily monitoring of parameters. Measurements outside of prescribed parameters were automatically displayed, resulting in the nurse calling the caregiver in the dyad to further assess the situation, provide education, and update the medication regimen. The nurse notified the primary physician or cardiologist about the patient's status as needed. *Control group:* Usual care.	*Several primary outcomes* (reducing subsequent hospital readmission, emergency department visits, and cost; and increasing the time between discharge and readmission). There were no significant differences due to telemonitoring for any outcomes. *Research targets:* Health care costs

---

[a] Articles were identified in a comprehensive search in Medline, EMBASE, PsycINFO, Cochrane Central Register of Controlled Trials, and IEEE Xplore from 1990 to November 2011, and were published in the time period 2002–2011.

[b] Main author, main reference. References with *and* between them are articles that belong to the same study. References in brackets contributed relevant study information on the study in question.

[c] Number of clinicians, number of patients with diagnosis, age range (mean) of patients, percentage female patients.

[d] Design (P = parallel group design, CO = crossover, C = cluster, F = factorial, O = other, U = unclear; + number of study arms), where symptom reporting took place (outside or in the home; or inside a clinic), and duration of intervention.

[e] Main findings are in general presented as in the original article and refer to primary outcome if clearly defined and secondary outcomes considered relevant for the scope of the study. Research targets refers to the six areas of health service quality defined by the Institute of Medicine [43].

[f] Peak expiratory flow rate.

[g] General practitioner.

[h] Chronic obstructive pulmonary disease.

Table 6. Summary description of studies on self-management in the reviewed randomized controlled trials of electronic symptom reporting[a].

Patient group	Trial and country[b]	Participant characteristics[c]	Study characteristics[d]	Health service innovation: self-management	Main findings and research targets[e]
Respiratory and lung disease					
	DeVito Dabbs et al [65]; USA	34 lung transplant recipients; >18 (mean 56) years; female 40%	Design: P + 2; outside/home Duration: the first 2 months after discharge	Early self-care behaviors and follow-up after lung transplant *Intervention: Pocket Personal Assistant for Tracking Health* (Pocket PATH): In addition to standard care, patients used a handheld device to record health data, review data trends by using the screens and graphs, and follow feedback instructions regarding reporting changes to their transplant coordinator. *Control:* Used standard paper-and-pencil logs to record data. *Both groups* were instructed to contact their transplant coordinator for any clinical questions or issues. Follow-up was identical.	*Several primary outcomes:* Patients in the Pocket PATH group showed significantly higher ratings of self-care agency, performed self-care behaviors at significantly higher rates, and reported significantly better health-related quality of life than standard-care controls. *Research targets:* Health benefits, patient centeredness
	Yardley et al [66]; UK	714 participants with minor respiratory symptoms; 18–79 years (62.1% were <25); female 72.3%	Design: P + 2; outside/home Duration: 1 access + follow-up after 48 hours (332) and 4 weeks (214)	Web-based intervention providing tailored advice for self-management of minor respiratory symptoms *Intervention:* Web-based *Internet Doctor*-provided tailored computer-generated advice for minor respiratory symptoms (cough, sore throat, fever, and runny or stuffy nose). Participants could access 3 main pages: (1) diagnostic pages asking a series of questions about symptoms, and a complex algorithm providing appropriate tailored advice on whether they needed to contact health services (+ options to possible diagnoses), (2) treatment pages providing information about medication for symptoms, (3) Common Questions section addressing common concerns and misconceptions about symptoms and treatment. *Control group* got access to a static webpage providing the best existing patient information.	*Several primary outcomes:* 1 month later the Internet Doctor resulted in higher levels of enablement (median 3 and 2, respectively; *P* = .03), and 11.6% (11) of participants consulted their doctor or used other health care services (mainly NHS Direct) for their symptoms, compared with a substantially greater proportion (21, 17.6%) in the control group (*P* = .22). *Secondary:* The Internet Doctor resulted in higher levels of satisfaction than the control information (mean 6.58 and 5.86, respectively; *P* = .002). Understanding of illness improved in the 48 hours following use of the Internet Doctor webpages, whereas it did not improve in the control group (mean change from baseline 0.21 and –0.06, respectively; *P* = .05). *Research targets:* Patient centeredness, resource utilization (for health professionals)
	van der Meer et al [67]; the Netherlands	200 asthma patients from 37 general practices and 1 academic outpatient department; 18–50 (mean in intervention: 36, in control: 37) years; females 69.5%	Design: P + 2; outside/home Duration: 12 months	Internet-based self-management plus education compared with usual care The *Internet-based self-management* program included weekly asthma-control monitoring and treatment advice, online and group education (face-to-face), and Web communication with a specialized asthma nurse, as an adjunct to usual care. Patients completed an electronic questionnaire on the website weekly and instantly received automated feedback on the state of their asthma control, along with advice on how to adjust their treatment (increasing or decreasing; contact asthma nurse). *Control group:* Usual physician-provided care according to Dutch general practice guidelines.	*Primary:* Asthma-related quality-of-life improvement of 0.5 point or more occurred in 54% and 27% of Internet and usual care patients, respectively (adjusted relative risk 2.00, confidence interval 1.38–3.04). Statistically significant, but not clinically significant. *Secondary:* Asthma control improved more in the Internet group than in the usual care group (adjusted difference –0.47, confidence interval –0.64 to –0.30). At 12 months, 63% of Internet patients and 52% of usual care patients reported symptom-free days in the previous 2 weeks (adjusted absolute difference 10.9%, confidence interval 0.05%–21.3%). *Research target:* Health benefits

Patient group	Trial and country[b]	Participant characteristics[c]	Study characteristics[d]	Health service innovation: self-management	Main findings and research targets[e]
	Nguyen et al [68]; USA	50 patients with moderate to severe chronic obstructive pulmonary disease; mean 69.5 years, range ± 8.5; female 44%	Design: P + 2; outside/home Duration: 6 months	Dyspnea self-management *Intervention*: Internet-based *(eDSMP)* dyspnea self-management *Control:* Face-to-face dyspnea self-management *(fDSMP)*. The content of the 2 programs was similar, focusing on education, skills training, and ongoing support for dyspnea self-management, including independent exercise. eDSMP participants submitted symptom and exercise information in real time via the PDA[f] or website. fDSMP paper diaries were mailed weekly. Nurse-provided feedback via email (eDSMP) or telephone (fDSMP), weekly for the first month and then biweekly for the next 5 months. Contacts were expected to be as similar as possible for the 2 groups, except that automated email alerts were sent to the study nurses if worsening of symptoms or reports of not performing exercise for at least 3 consecutive days.	*Primary:* Both groups showed similar clinically meaningful changes in dyspnea with activities of daily living after 3 months and sustained these improvements at 6 months. *Secondary:* Self-reported endurance exercise time (P = .001), physical functioning (P = .04), and self-efficacy for managing dyspnea (P = .02) also showed positive improvements over time in both groups with no significant differences with respect to program modality. *Research targets:* Health benefits, patient centeredness
Psychiatry					
	Berger et al [69]; Switzerland	81 patients with social phobia; 19–62 (mean 37.2) years; female 53.1%	Design: P + 3; outside/home Duration: 10-week treatment + 6-month follow-up	Internet-based treatments of social phobia *Intervention 1: Guided Internet-based self-help* program with weekly scheduled email feedback by a therapist and the possibility to ask questions via email (response time maximum 3 days). *Intervention 2: Step-up on demand* (same as control) but with the possibility to step up to guidance by email (intervention 1) or telephone. Both groups used an online diary to report anxiety-provoking situations, and related thoughts, feelings, and behaviors. *Control: Pure self-help* program (by Internet) without any therapist support.	Significant symptom reductions in all 3 treatment groups with large effect sizes for *primary measures* (self-reported measures of symptoms of social phobia) and for *secondary outcome* measures (symptoms of depression, interpersonal problems, and general symptomatology). No significant between-group effects were found. No significant difference between the 3 conditions regarding diagnosis-free status, clinical change, dropout rates, or adherence measures such as lessons or exercises completed. High level of patient satisfaction overall, with a significant difference favoring the guided Internet-based self-help group. *Research targets:* Health benefits, patient centeredness
	Bergstrom et al [70]; Sweden	113 patients with panic disorder; >18 (mean in intervention: 33.8, in control: 34.6) years; female 61.5%	Design: P + 2; outside/home Duration: 10 weeks + 6-month follow-up	Internet-based CBT[g] for patients with panic disorder *Intervention:* 10 Web-based self-help modules, 1 per week, with information, exercises, and homework assignments, based on established CBT principles. Psychologist provided feedback, gave access to next module, and replied to other messages within 24 hours on regular weekdays. Only email contact. *Control:* Regular psychiatric care setting. Psychologists presented the same self-help program as above during weekly 2-hour sessions, supported by handouts. Homework assignments were addressed during group sessions.	*Primary:* Internet CBT is as effective as the more widely used group CBT. No significant between-group effects were found. *Secondary:* Internet treatment had superior cost-effectiveness ratios in relation to group treatment at both posttreatment and follow-up. *Research targets:* Health benefits, health care costs

Patient group	Trial and country[b]	Participant characteristics[c]	Study characteristics[d]	Health service innovation: self-management	Main findings and research targets[e]
	Vernmark et al [71]; Sweden	88 patients with major depression; age range not reported (mean 37); female 68%	Design: P + 3; outside/home Duration: 8 weeks (we don't report from the 6-month follow-up, since the control group had received guided self-help by this time)	Internet-administered guided self-help versus individualized email therapy versus waiting lists *Intervention 1: Guided self-help* included weekly modules and homework assignments. Standard CBT components were presented. Therapists contributed with positive reinforcement on the progress made by participants. *Intervention 2: email therapy* did not use the self-help texts; all emails were written for the unique patient. The therapists had more or less the same role as in face-to-face psychological treatment. The contents of the email therapy overlapped with the self-help material but were tailored to each participant's needs. The treatment was based on a protocol manual developed by the team. The third group was a *waiting-list control group*. Each therapist was responsible for 5 participants in each group.	*Primary:* Both the email therapy and the self-help groups improved in symptom reduction compared with the waiting-list condition (P = .002 and P = .06). The 2 treatments did not differ (P = .41). At posttreatment 34.5% of the guided self-help group, 30% of the email therapy group, and 13.8% of the waiting-list group reached the criteria of high-end state functioning (P = .17), (Beck Depression Inventory score <9). *Research target:* Health benefits
	Oerlemans et al [72]; the Netherlands	76 irritable bowel syndrome patients; age range not reported; mean in intervention: 35.9, for control: 40.6 years; female in intervention: 91.9%, in control: 76.9%	Design: P + 2; outside/mobile Duration: 4 weeks + 3-month follow-up	Intervening on cognitions and behavior in irritable bowel syndrome *Intervention: Patients* received standard care supplemented with a 4-week CBT e-intervention on PDAs. Patients completed 3 diaries daily. The data were immediately accessible to the psychologist, who during weeks 2–4 sent situational feedback based on CBT via text message. Feedback was standardized through a developed protocol. *Control:* Received standard care consisting of reassurance, dietary advice, and education from their general practitioner.	*Several primary outcomes:* No significant differences between groups for dysfunctional cognitions. Between-group comparisons after 4 weeks showed more overall quality-of-life improvement, more improvement in catastrophizing thoughts, and more pain improvement in the intervention group. Only improvement in catastrophizing thoughts persisted in the long term. The eHealth intervention seems feasible, since all intervention group patients completed the diaries 3 times a day for 4 weeks. *Research target:* Health benefits

Patient group	Trial and country[b]	Participant characteristics[c]	Study characteristics[d]	Health service innovation: self-management	Main findings and research targets[e]
Diabetes	Glasgow et al [73] and Williams et al [74]; USA	52 primary care physicians, 886 type 2 diabetes patients; >25 (mean in intervention: 61.48, in control: 64.63) years; female in intervention: 52.3%, in control: 50% ([73])	Design: C + 2; inside clinic Duration: 12 months	Interactive computer technology to assist patients and clinicians in emphasizing patient-centered communication and improved quality of care *Computer-assisted intervention* patients were asked to come 30 minutes early to 2 diabetes-related visits scheduled 6 months apart to complete a touch-screen assessment procedure and set self-management goals, received computer- tailored feedback and individualized action plans, received a printout on general health risks, met with a care manager, and received follow-up phone calls from care manager (nurse or medical assistant). Physician and care manager received printout of patient's needs, self-management goals, and areas the patient wished to discuss. *Control patients* also completed touch-screen assessment procedure, received printout on general health risks, but did not set self-management goals, meet with a care manager, or receive follow-up phone calls.	*Glasgow et al [73]: Primary:* Significantly improved both the laboratory assays and patient-centered aspects of diabetes care that patients received compared with those in the control condition. *Secondary:* Overall improvement in lipids, glycated hemoglobin, quality of life, and depression scores; between-condition differences were not significant. *Williams et al [74]: Unclear primary outcome.* The intervention increased patient perception of autonomy support relative to a computer-based control condition ($P < .05$). Change in perceived competence partially mediated the effects of increased autonomy support on the change in lipids, diabetes distress, and depressive symptoms. The construct of autonomy support was separate from that of patient satisfaction. *Research targets:* Health benefits, patient centeredness

[a] Articles were identified in a comprehensive search in Medline, EMBASE, PsycINFO, Cochrane Central Register of Controlled Trials, and IEEE Xplore from 1990 to November 2011, and were published in the time period 2002–2011.

[b] Main author, main reference. References with *and* between them are articles that belong to the same study.

[c] Number of clinicians, number of patients with diagnosis, age range (mean) of patients, percentage female patients.

[d] Design (P = parallel group design, CO = crossover, C = cluster, F = factorial, O = other, U = unclear; + number of study arms), where symptom reporting took place (outside or in the home; or inside a clinic), and duration of intervention.

[e] Main findings are in general presented as in the original article and refer to primary outcome if clearly defined and secondary outcomes considered relevant for the scope of the study. Research targets refers to the six areas of health service quality defined by the Institute of Medicine [43].

[f] Personal digital assistant.

[g] Cognitive behavioral therapy.

Table 7. Summary description of study on therapy in the reviewed randomized controlled trials of electronic symptom reporting[a]

Trial and country[b]	Participant characteristics[c]	Study characteristics[d]	Health service innovation: therapy	Main findings and research targets[e]
Wagner et al [75]; Switzerland	55 people with complicated grief; 19–68 (mean 37) years; female 93% Patients lived in Germany, Austria, and Switzerland or were native German speakers living elsewhere	Design: P + 2; outside/home Duration: 5 weeks + 3-month follow-up	Internet-based cognitive behavioral therapy for complicated grief. In *Internet-based cognitive behavioral therapy*, patients were set 2 weekly 45-minute writing assignments over a period of approximately 5 weeks, with therapist and patient communicating exclusively by email. After every second essay, patients received feedback and further instructions from the therapist. Instructions were sent within 1 working day and based on a cognitive behavioral treatment protocol but tailored to the individual patient's needs. At the beginning of each phase of treatment, patients received psycho-education on the principles of the treatment module. *Control group:* Waiting list (received treatment 5 weeks after the treatment group termination for ethical reasons).	*Several primary outcomes:* Participants in the treatment group improved significantly relative to participants in the waiting condition on symptoms of intrusion, avoidance, maladaptive behavior, and general psychopathology, and showed a large treatment effect. Follow-up results show that this improvement was maintained after 3 months. *Research target:* Health benefits

[a] Articles were identified in a comprehensive search in Medline, EMBASE, PsycINFO, Cochrane Central Register of Controlled Trials, and IEEE Xplore from 1990 to November 2011, and were published in the time period 2002–2011.

[b] Main author, main reference.

[c] Number of clinicians, number of patients with diagnosis, age range (mean) of patients, percentage female patients.

[d] Design (P = parallel group design, CO = crossover, C = cluster, F = factorial, O = other, U = unclear; + number of study arms), where symptom reporting took place (outside or in the home; or inside a clinic), and duration of intervention.

[e] Main findings are in general presented as in the original article and refer to primary outcome if clearly defined and secondary outcomes considered relevant for the scope of the study. Research targets refers to the six areas of health service quality defined by the Institute of Medicine [43].

Research Targets Relevant to Electronic Symptom Reporting

According to the IOM categories of health service outcomes [43], the most common research target was disease-specific health benefits at the patient level; and, second, to provide patient-centered care (Table 4, Table 5, Table 6, and Table 7). Some of the studies also aimed at more efficient utilization of the health care system to reduce cost. The consultation support studies mainly aimed at providing patient-centered care, while monitoring and self-management studies mainly aimed for patient health benefits. The studies that aimed for reduced health care costs were all in the monitoring category, except for 1 in the self-management category.

The main research focus is presented according to the resulting health service innovations and patient groups in Multimedia Appendix 3.

Discussion

Principal Findings on Patient Groups, Health Service Innovations, and Research Targets

Of 642 records identified, 32 articles representing 29 studies were included. The articles were published from 2002, most (24/32, 75%) during the last 5 years. Nearly two-thirds of the articles had a female first author.

We categorized the studies into five patient groups: respiratory and lung diseases, cancer, psychiatry, cardiovascular diseases, diabetes, or a mix of these. All included studies, except 1 [66], focused on long-term conditions or issues that must take into account fluctuation in condition intensity and variations in how they influence the patient's life.

The content analysis identified four categories of health service innovations: consultation support, monitoring with clinician support, self-management with clinician support, and therapy.

The research targets in the group of articles on consultation support were mainly patient-centered outcomes, while the articles on monitoring and self-management mainly aimed to demonstrate health benefits. The studies aiming for reduced health care costs were all in the subgroup of monitoring articles, except for 1 study on self-management.

Interpretation of Results

We found that 75% of the articles were published during the last 5 years, which indicates that this is a growing field. The fact that all studies, except 2, were conducted in Western countries is not surprising, since these are the countries with the highest e-readiness [6].

It is no surprise, either, that nearly all the included studies focused on long-term conditions. In emergency or acute conditions, the time frame for decision making is short, and health professionals often need to make decisions on behalf of patients. In long-term conditions, however, the time frame is longer, and the decision process is often shared between the patient and the health care professional. Partnership or shared decision making is essential to improve the pathways for patients with long-term conditions who face complex and repeated decision making processes [9]. Supporting that decision process with self-management and patient education through technology makes sense.

Most of the consultation support innovation studies were conducted in the cancer patient group (5/7), most of the monitoring studies were in the respiratory and lung diseases (8/12), and the self-management studies were conducted mainly in psychiatry (4/9) or respiratory and lung diseases (4/9). Cancer patients who receive chemotherapy or radiation therapy (or both) for a period from 6 months to a year could theoretically benefit from both monitoring and self-management approaches, in the same way as patients with chronic obstructive pulmonary disease or asthma. Yet, electronic symptom reporting for this group of patients has mostly been studied in the context of consultation support. Likewise, electronic consultation support has not been studied in patients with chronic conditions such as chronic obstructive pulmonary disease and asthma. We can see no theoretical or practical reason to believe that these groups would not benefit in the same way as those with other long-term conditions. Thus, we are puzzled by the obvious blanks in our cross-tables of patient group by health service innovation, and that electronic symptom reporting systems seems to reflect the conventional approach in each health service field. We are concerned that health service innovations that may benefit all patients with long-term conditions are being introduced in a diagnosis-specific context. This makes it difficult for researchers and clinicians to glean more general lessons from the field. As we discuss further in the next paragraph, systems deviating from this conventional approach might benefit health care service quality. Based on this, we suggest that the field would benefit from the identification of general theoretical principles that are relevant to all electronic symptom reporting interventions, across diagnostic patient groups.

We identified four types of health service interventions, and we believe these four represent the full spectrum of services associated with electronic symptom reporting. One of the health service innovations groups, the consultation support group, was very different from the other three, while the monitoring, self-management support, and therapy groups partly overlapped. These three represent a continuum with increasing focus on treatment through electronic communication, and decreasing face-to-face or telephone contact. The ideal electronic symptom reporting service should provide both consultation support and elements of monitoring and self-management support, and when this is not enough, to support the therapeutic relationship whenever this is feasible.

The studies aiming at reduced health care costs were all in the monitoring category, except for 1 self-management study. The 4 studies defined as equivalence studies in part 2 of this review [1], which all belong to the self-management group, could be expected to focus on health care costs, but this is not the case. Equivalence here refers to the intervention hypothesis relative to the control, and not to equivalence with regard to cost. Only 1 of these studies focused on cost effectiveness, and then as its second aim [70], while the other 3 did not formally analyze cost effectiveness. However, Vernmark et al discussed cost effectiveness with regard to spent therapist time [71].

Limitations and Strengths of the Review

This is the first review to address this emerging field and to provide a systematic overview. One of the main strengths of this review is the comprehensive search. We searched all the databases recommended by Cochrane [35] (chapter 6), in addition to the IEEE Xplore and the PsycINFO databases. All searches, except those in IEEE Xplore, included 115 or more search terms, and we adapted the searches to the individual databases. Compared with other reviews, in this review we judged a quite high percentage of the identified records (444) in our searches as relevant articles (32).

Another main strength is that the selection and data extraction strategies were based on the Cochrane recommendations.

That the review was based on the most mature research in the field, the RCT trials [32,33], is also an important strength. We are familiar with the challenges related to the use of RCTs in medical informatics [82], which therefore are not always applied to test new complex interventions. Limiting our review to RCTs might have led us to miss interventions that could be relatively mature but not tested in an RCT. Even when considering guidelines for reporting medical informatics research [83], we did not see any other way of identifying mature interventions, without simultaneously including a large body of pilot studies and feasibility trials. Evidence-based medicine with RCTs as basic methodology is widely accepted as one important facet in improving clinical practice or patient outcome [84]. However, within the constraints of the review, we did not consider studies using other methods that might have contributed to knowledge about electronic symptom reporting.

Despite using a very comprehensive search strategy, we might have lost articles in the adaptation and translation of the search strategies for the different databases. In addition, we may have missed search words, resulting in overlooked articles. When designing this review, we decided not to include articles based on hand searching of reference lists, due to the Cochrane warning [35] (Cochrane 10.2.2.3, Citation bias). Nevertheless, we checked the reference lists to get a sense of the completeness of our search. We read abstracts for all references where the title included RCT and technology implying communication, and if the abstract seemed relevant, we read the full article. We repeated this process twice for new articles we identified. Our check showed that we did not capture some psychiatry articles—mainly those that focused on a fully electronic therapeutic relationship. In these articles, the electronic symptom reporting and responses to the specific symptoms were a part of the whole picture. We acknowledge this as a blank in our review, as only 1 article from this area came up with our applied search strategy. We propose that this area merits its own review, using psychiatry terminology. We suggest not focusing so much on symptom and its synonyms as search words, since they were lacking in many of these studies, but to include specific search strings such as Interapy [85,86], Internet-based therapy [86,87], Internet-based treatment [86,88-91], and online therapeutic relationship [86].

We included the 6 psychiatry studies, as we identified them in our originally designed search strategy. If they had not been included, the self-management category would have been reduced, and been less convincing, and the focus for self-management would mainly have been on respiratory and lung diseases. On the other hand, if we had conducted a search that had covered the psychiatry field better, we hypothesize, based on the studies from the reference lists, that the psychiatry studies would have been a mix of self-management and therapy studies. However, further research will have to confirm this hypothesis.

Future Research

The finding that nearly two-thirds of the articles had a female first author was surprising and actually something we did not look for, but was immediately obvious. This is much higher than is common in medicine, where only a few American journals have up to 30% female authors, while all others have less [92]. We do not have an explanation for this finding, but female researchers might be more engaged than their male colleagues in patient self-empowerment, defined as “a state in which an individual possesses a relatively high degree of actual power—that is, a genuine potential for making choices” [93] (p. 40). Further studies are, however, needed to investigate this observation.

As mentioned above, health service innovations in this area have so far mainly been developed and tested in the context of a given diagnosis. However, many possible ICPC-based patient groups are not represented in our findings. This is probably only a question of time, since the prereview, which was not limited to RCTs, identified 15 studies on musculoskeletal disease, 8 on gastrointestinal diseases, 8 on neurological diseases, 6 on human immunodeficiency syndrome/acquired immunodeficiency syndrome, and a large group that was not possible to categorize based on the information in the abstract [11]. These were in general pilot studies, where the next step probably is an RCT, if the innovation proves to be feasible.

What we have today is a highly heterogeneous field, where authors rarely seem to build on the experiences gained in other diagnostic settings. The fact that almost all of the patient groups had long-term conditions suggests that long-term conditions have commonalities that make this kind of intervention desirable. Future research should examine whether a common generic approach to electronic symptom reporting, regardless of diagnosis, could be useful.

A total of 3 self-management studies (4 articles) [66,67,73,74] and 3 monitoring studies [54,55,57] mainly used computer-tailored feedback to the patients. None of these studies focused on health care costs, even if these interventions may be the most promising ones to save both time and money for the health care system, compared with monitoring and self-management, where the physician or the nurse is the main communication partner. Therefore, health care costs should be an outcome in future computer-tailored feedback studies. This includes the 3 self-management studies [66,67,73,74] and the 2 monitoring studies [54,57] mentioned above, for which part 2 of this review confirms that they have acceptable methodological quality and that the hypothesis is confirmed [1].

In this regard, the Yardley et al study deserves some special attention [66]. This is the only study where no human recourses were involved on the provider side. The computer system’s tailored advice for patients with minor respiratory problems resulted in a higher level of enablement, higher satisfaction, better understanding of the illness, and a modest effect on reduced consultation rates [66] (see Part 2 of this review [1]). This concept is an excellent example of a Web-based decision support system for patients that seems to both help the patients and save time for patients, the health care system, and health care professionals. However, future studies of this concept and other Web-based decision support systems for patients, and investigation of their effect in routine care, are necessary.

In addition to the possible health cost benefits, giving the patient the opportunity to register symptoms continually and providing an interactive-feedback learning mechanism can provide the stimulus for the patient to build the necessary confidence to handle symptoms and self-management, and in this way support patient centeredness. We support Guendelman and colleagues’ suggestion that easy-to-use electronic devices “may be useful tools to empower children to provide their own care while reducing asthma symptoms and health care use in pediatric settings” [54], and might even be considered as a motivating and exciting tool for children with asthma. This idea is supported by the study of Jan et al, where several children reported that using the tool they received was fun and that it reminded them to take their medication [55]. Therefore, further research is needed to discover both the motivating (fun) and the self-management effects of technology “toys” with interactive-feedback learning mechanisms to handle symptoms.

Systematic use or reuse of electronically reported symptoms might also be useful in syndromic surveillance [94,95], to make clinicians aware of community trends and to enable them to issue the right tests and improve their diagnostic assessment [96]. There are several examples of patients reporting symptoms through the Internet for syndromic surveillance [97-100] and of relevant surveillance information being produced based on what people report on the Internet [101]. Whether symptoms reported before a consultation or reported to a decision support systems, such as that of Yardley et al [66], can be re-used for syndromic surveillance and thus result in a double effect should be investigated.

Investigating what the opportunity to easily contact care providers means to patients with long-term conditions, in terms of feeling secure, appears necessary, as nearly all of the patients who contacted the e-coach in the study of Leveille and colleagues were interested in further coaching [50,81], although the intervention had not been that promising regarding detection of screened conditions, symptom ratings, and quality of life [50].

With regard to research targets, the most interesting finding may be that none of the trials focused on safety, timeliness, and equity. Timeliness—that is, reducing delays for both providers and receivers of health care, for example, avoiding cancellation of surgery—is an area where we expect electronic symptom reporting to have a positive impact. As the mobile phone seems to narrow the digital divide [102], electronic symptom reporting might as well improve equity—that is, that health care does not vary in quality because of gender, ethnicity, geographic location, and socioeconomic status. Studies addressing these issues are needed to investigate the potential benefits.

Implications for Practice

The recent large increase in studies being conducted in electronic symptom reporting, as also shown in our preliminary review [11], reflects the establishment of a new concept of improved patient centeredness. In addition, 88% of doctors express that they want their patients to report health indicators via mobile devices, and 40% of doctors believe it can reduce the number of office visits, according to a company that both creates products for health care companies and conducts research [103]. Accordingly, some communities and countries have already taken serious steps to achieve maximum benefits from these types of innovations. Plans to include patient-reported information as part of the electronic health record [104] have been developed. Sweden is now establishing an electronic health record and personal health record to make it possible for the patient to read health record information written by their providers, and for the health care professional to read specific parts of the patient's personal health record describing symptoms and health information relevant to the patient's problems and future consultations [105,106]. England's Department of Health has just announced that they want their general practitioners to prescribe apps rather than doctors' visits whenever possible [107-110]. They want patients to use mobile devices to monitor and track their health status, and to identify—and if possible solve—the problem before they request a visit [107-110]. They suggest these initiatives will improve quality, save money, and give patients more control over their own health. In addition, these initiatives will probably also inspire other counties to establish similar strategies. We welcome these initiatives, but recommend basing design and implementation plans on research with regard to how the technology can be used to provide safe, effective, patient-centered, timely, efficient, and equitable health care.

Conclusion

The RCT-based research on electronic symptom reporting has developed enormously since 2002, with 75% of the articles published during the last 5 years. This indicates that a new concept to improve patient centeredness is being established. So far, the research has focused on five specific patient groups and health conditions: cancer, respiratory and lung diseases, cardiovascular disease, psychiatry, and diabetes. The evidence from RCTs can be structured into four health service innovation categories: consultation support, monitoring with clinician support, self-management with clinician support, and therapy. Most of the research (72%) has been conducted within the following four combinations: consultation support innovation in the cancer patient group (5/29, 17%), monitoring innovation in the respiratory and lung diseases patient group (8/29, 28%), and self-management innovation in the psychiatry patient group (4/29, 14%) and in the respiratory and lung diseases patient group (4/29, 14%). New patient groups, and combinations of patient groups and the four identified health service innovations, are expected in the near future. We suggest that the development of a generic (not diagnosis-specific) model for electronic patient symptom reporting for long-term conditions may benefit the development of this field.

Acknowledgments

This study was partly funded by grants from the Health North Authorities (TFP-661-07), Norwegian Research Council grants (1749349), and The Norwegian Centre for Integrated Care and Telemedicine (NST). We want to express our thanks to colleagues at NST for useful advice, support, and feedback.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Medline search strategy and search terms.

[PDF File (Adobe PDF File), 130KB - jmir_v14i5e118_app1.pdf]

Multimedia Appendix 2

The classification process based on abstracts from the database searches.

[PDF File (Adobe PDF File), 9KB - jmir_v14i5e118_app2.pdf]

Multimedia Appendix 3

The main focus of the research targets presented according to health service innovations and patient groups.

[PDF File (Adobe PDF File), 82KB - jmir_v14i5e118_app3.pdf]

References

1. Johansen MA, Berntsen GKR, Schuster T, Henriksen E, Horsch A. Electronic Symptom Reporting Between Patient and Provider for Improved Health Care Service Quality: A Systematic Review of Randomized Controlled Trials. Part 2: Methodological Quality and Effects. J Med Internet Res 2012 Oct 03;14(5):e126 [FREE Full text] [doi: 10.2196/jmir.2216]

2. Nijland N, van Gemert-Pijnen JE, Boer H, Steehouder MF, Seydel ER. Increasing the use of e-consultation in primary care: results of an online survey among non-users of e-consultation. Int J Med Inform 2009 Oct;78(10):688-703. [doi: 10.1016/j.ijmedinf.2009.06.002] [Medline: 19625210]
3. Kummervold PE, Gammon D, Bergvik S, Johnsen JA, Hasvold T, Rosenvinge JH. Social support in a wired world: use of online mental health forums in Norway. Nord J Psychiatry 2002;56(1):59-65. [doi: 10.1080/08039480252803945] [Medline: 11869468]
4. Webb PM, Zimet GD, Fortenberry JD, Blythe MJ. Comparability of a computer-assisted versus written method for collecting health behavior information from adolescent patients. J Adolesc Health 1999 Jun;24(6):383-388. [Medline: 10401965]
5. Millsopp L, Frackleton S, Lowe D, Rogers SN. A feasibility study of computer-assisted health-related quality of life data collection in patients with oral and oropharyngeal cancer. Int J Oral Maxillofac Surg 2006 Aug;35(8):761-764. [doi: 10.1016/j.ijom.2006.03.011] [Medline: 16697148]
6. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]
7. Kaplan B, Brennan PF. Consumer informatics supporting patients as co-producers of quality. J Am Med Inform Assoc 2001;8(4):309-316 [FREE Full text] [Medline: 11418537]
8. Randeree E, Whetstone M. Personal health records: patients in control. In: Wilson EV , editor. Patient-Centered E-Health. Hershey PA: IGI Global; 2009:47-59.
9. Drake RE, Deegan PE, Woltmann E, Haslett W, Drake T, Rapp CA. Comprehensive electronic decision support systems. Psychiatr Serv 2010 Jul;61(7):714-717. [doi: 10.1176/appi.ps.61.7.714] [Medline: 20592007]
10. Drake RE, Cimpean D, Torrey WC. Shared decision making in mental health: prospects for personalized medicine. Dialogues Clin Neurosci 2009;11(4):455-463. [Medline: 20135903]
11. Johansen MA, Henriksen E, Berntsen G, Horsch A. Electronic symptom reporting by patients: a literature review. Stud Health Technol Inform 2011;169:13-17. [Medline: 21893705]
12. Sciamanna CN, Diaz J, Myne P. Patient attitudes toward using computers to improve health services delivery. BMC Health Serv Res 2002 Sep 11;2(1):19 [FREE Full text] [Medline: 12225617]
13. Benoit A, Dykes P, Chang F, Gertman P, Vandever W, Li Q, et al. Using electronic questionnaires to collect patient reported history. AMIA Annu Symp Proc 2007:871. [Medline: 18693972]
14. Johansen MA, Berntsen G, Shrestha N, Bellika JG, Johnsen JA. An exploratory study of patient attitudes towards symptom reporting in a primary care setting. Benefits for medical consultation and syndromic surveillance? Methods Inf Med 2011;50(5):479-486. [doi: 10.3414/ME11-02-0005] [Medline: 21897995]
15. Buzaglo JS, Millard JL, Ridgway CG, Ross EA, Antaramian SP, Miller SM, et al. An Internet method to assess cancer patient information needs and enhance doctor-patient communication: a pilot study. J Cancer Educ 2007;22(4):233-240. [doi: 10.1080/08858190701634870] [Medline: 18067435]
16. Wald JS, Businger A, Gandhi TK, Grant RW, Poon EG, Schnipper JL, et al. Implementing practice-linked pre-visit electronic journals in primary care: patient and physician use and satisfaction. J Am Med Inform Assoc 2010;17(5):502-506 [FREE Full text] [doi: 10.1136/jamia.2009.001362] [Medline: 20819852]
17. Adamson SC, Bachman JW. Pilot study of providing online care in a primary care setting. Mayo Clin Proc 2010 Aug;85(8):704-710. [doi: 10.4065/mcp.2010.0145] [Medline: 20516427]
18. Kummervold PE, Trondsen M, Andreassen H, Gammon D, Hjortdahl P. [Patient-physician interaction over the internet]. Tidsskr Nor Laegeforen 2004 Oct 21;124(20):2633-2636 [FREE Full text] [Medline: 15534640]
19. Head BA, Keeney C, Studts JL, Khayat M, Bumpous J, Pfeifer M. Feasibility and Acceptance of a Telehealth Intervention to Promote Symptom Management during Treatment for Head and Neck Cancer. J Support Oncol 2011 Jan 1;9(1):e1-e11. [doi: 10.1016/j.suponc.2010.12.006] [Medline: 21499540]
20. Porter SC, Kohane IS, Goldmann DA. Parents as partners in obtaining the medication history. J Am Med Inform Assoc 2005;12(3):299-305 [FREE Full text] [doi: 10.1197/jamia.M1713] [Medline: 15684127]
21. de Jongste JC, Carraro S, Hop WC, CHARISM Study Group, Baraldi E. Daily telemonitoring of exhaled nitric oxide and symptoms in the treatment of childhood asthma. Am J Respir Crit Care Med 2009 Jan 15;179(2):93-97 [FREE Full text] [doi: 10.1164/rccm.200807-1010OC] [Medline: 18931330]
22. Artinian NT, Harden JK, Kronenberg MW, Vander Wal JS, Daher E, Stephens Q, et al. Pilot study of a Web-based compliance monitoring device for patients with congestive heart failure. Heart Lung 2003;32(4):226-233. [Medline: 12891162]
23. Ross SE, Moore LA, Earnest MA, Wittevrongel L, Lin CT. Providing a web-based online medical record with electronic communication capabilities to patients with congestive heart failure: randomized trial. J Med Internet Res 2004 May 14;6(2):e12 [FREE Full text] [doi: 10.2196/jmir.6.2.e12] [Medline: 15249261]
24. Gaertner J, Elsner F, Pollmann-Dahmen K, Radbruch L, Sabatowski R. Electronic pain diary: a randomized crossover study. J Pain Symptom Manage 2004 Sep;28(3):259-267. [doi: 10.1016/j.jpainsymman.2003.12.017] [Medline: 15336338]
25. Miller DM, Moore SM, Fox RJ, Atreja A, Fu AZ, Lee JC, et al. Web-based self-management for patients with multiple sclerosis: a practical, randomized trial. Telemed J E Health 2011;17(1):5-13. [doi: 10.1089/tmj.2010.0133] [Medline: 21214498]

26. Lieberman DZ, Kelly TF, Douglas L, Goodwin FK. A randomized comparison of online and paper mood charts for people with bipolar disorder. J Affect Disord 2010 Jul;124(1-2):85-89. [doi: 10.1016/j.jad.2009.10.019] [Medline: 19896202]
27. Ljótsson B, Hedman E, Andersson E, Hesser H, Lindfors P, Hursti T, et al. Internet-delivered exposure-based treatment vs. stress management for irritable bowel syndrome: a randomized trial. Am J Gastroenterol 2011 Aug;106(8):1481-1491. [doi: 10.1038/ajg.2011.139] [Medline: 21537360]
28. Eysenbach G, Diepgen TL. Epidemiological data can be gathered with world wide web. BMJ 1998 Jan 3;316(7124):72. [Medline: 9451290]
29. Barnason S, Zimmerman L, Nieveen J, Schulz P, Miller C, Hertzog M, et al. Influence of a symptom management telehealth intervention on older adults' early recovery outcomes after coronary artery bypass surgery. Heart Lung 2009;38(5):364-376. [doi: 10.1016/j.hrtlng.2009.01.005] [Medline: 19755186]
30. Zimmerman L, Barnason S, Hertzog M, Young L, Nieveen J, Schulz P, et al. Gender differences in recovery outcomes after an early recovery symptom management intervention. Heart Lung 2011 Oct;40(5):429-439. [doi: 10.1016/j.hrtlng.2010.07.018] [Medline: 21501872]
31. Lorenzi NM, Riley RT, Blyth AJ, Southon G, Dixon BJ. Antecedents of the people and organizational aspects of medical informatics: review of the literature. J Am Med Inform Assoc 1997;4(2):79-93 [FREE Full text] [Medline: 9067874]
32. Campbell NC, Murray E, Darbyshire J, Emery J, Farmer A, Griffiths F, et al. Designing and evaluating complex interventions to improve health care. BMJ 2007 Mar 3;334(7591):455-459. [doi: 10.1136/bmj.39108.379965.BE] [Medline: 17332585]
33. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. BMJ 2000 Sep 16;321(7262):694-696. [Medline: 10987780]
34. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. Open Med 2009;3(3):e123-e130. [Medline: 21603045]
35. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0 [updated March 2011]. 2011. URL: http://www.cochrane-handbook.org/ [accessed 2012-08-23] [WebCite Cache ID 6A8elFwgQ]
36. Eklund AM. Medical knowledge evolution query constraining aspects. Stud Health Technol Inform 2011;169:549-553. [Medline: 21893809]
37. Wong SS, Wilczynski NL, Haynes RB. Comparison of top-performing search strategies for detecting clinically sound treatment studies and systematic reviews in MEDLINE and EMBASE. J Med Libr Assoc 2006 Oct;94(4):451-455. [Medline: 17082841]
38. Johansen MA. The Norwegian Centre for Integrated Care and Telemedicine. 2012. List of Variables URL: http://telemed.custompublish.com/getfile.php/1936271.357.dvtfsvvepq/Variabelliste.pdf [accessed 2012-05-08] [WebCite Cache ID 67VZUbsga]
39. Evans D, Pearson A. Systematic reviews of qualitative research. Clin Eff Nurs 2001;5:111-119. [doi: 10.1054/cein.2001.0219]
40. Devers KJ. Qualitative methods in health services and management research: pockets of excellence and progress, but still a long way to go. Med Care Res Rev 2011 Feb;68(1):41-48. [doi: 10.1177/1077558710384269] [Medline: 21252376]
41. Pope C, Ziebland S, Mays N. Qualitative research in health care. Analysing qualitative data. BMJ 2000 Jan 8;320(7227):114-116. [Medline: 10625273]
42. KITH. Norwegian Centre for Informatics in Health and Social Care in cooperation with "Norsk selskap for allemennmedisin". 2004. ICPC-2 Kodekort International Classification of Primary Care, Norwegian version. Translated and adapted to Norwegian conditions by KITH URL: http://www.kith.no/upload/1895/ICPC-2-kodekort_fastsatt-04.2004.pdf [accessed 2012-05-29] [WebCite Cache ID 6815fPfhJ]
43. Committee on Quality of Health Care in America, Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press; 2001.
44. Berry DL, Blumenstein BA, Halpenny B, Wolpin S, Fann JR, Austin-Seymour M, et al. Enhancing patient-provider communication with the electronic self-report assessment for cancer: a randomized trial. J Clin Oncol 2011 Mar 10;29(8):1029-1035. [doi: 10.1200/JCO.2010.30.3909] [Medline: 21282548]
45. Boyes A, Newell S, Girgis A, McElduff P, Sanson-Fisher R. Does routine assessment and real-time feedback improve cancer patients' psychosocial well-being? Eur J Cancer Care (Engl) 2006 May;15(2):163-171. [doi: 10.1111/j.1365-2354.2005.00633.x] [Medline: 16643264]
46. Ruland CM, White T, Stevens M, Fanciullo G, Khilani SM. Effects of a computerized system to support shared decision making in symptom management of cancer patients: preliminary results. J Am Med Inform Assoc 2003;10(6):573-579 [FREE Full text] [doi: 10.1197/jamia.M1365] [Medline: 12925545]
47. Ruland CM, Holte HH, Røislien J, Heaven C, Hamilton GA, Kristiansen J, et al. Effects of a computer-supported interactive tailored patient assessment tool on patient care, symptom distress, and patients' need for symptom management support: a randomized clinical trial. J Am Med Inform Assoc 2010;17(4):403-410 [FREE Full text] [doi: 10.1136/jamia.2010.005660] [Medline: 20595307]
48. Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. J Clin Oncol 2004 Feb 15;22(4):714-724. [doi: 10.1200/JCO.2004.06.078] [Medline: 14966096]

49. Stevens J, Kelleher KJ, Gardner W, Chisolm D, McGeehan J, Pajer K, et al. Trial of computerized screening for adolescent behavioral concerns. Pediatrics 2008 Jun;121(6):1099-1105 [FREE Full text] [doi: 10.1542/peds.2007-1878] [Medline: 18519478]
50. Leveille SG, Huang A, Tsai SB, Allen M, Weingart SN, Iezzoni LI. Health coaching via an internet portal for primary care patients with chronic conditions: a randomized controlled trial. Med Care 2009 Jan;47(1):41-47. [doi: 10.1097/MLR.0b013e3181844dd0] [Medline: 19106729]
51. Kearney N, McCann L, Norrie J, Taylor L, Gray P, McGee-Lennon M, et al. Evaluation of a mobile phone-based, advanced symptom management system (ASyMS) in the management of chemotherapy-related toxicity. Support Care Cancer 2009 Apr;17(4):437-444. [doi: 10.1007/s00520-008-0515-0] [Medline: 18953579]
52. Chan DS, Callahan CW, Sheets SJ, Moreno CN, Malone FJ. An Internet-based store-and-forward video home telehealth system for improving asthma outcomes in children. Am J Health Syst Pharm 2003 Oct 1;60(19):1976-1981. [Medline: 14531243]
53. Chan DS, Callahan CW, Hatch-Pigott VB, Lawless A, Proffitt HL, Manning NE, et al. Internet-based home monitoring and education of children with asthma is comparable to ideal office-based care: results of a 1-year asthma in-home monitoring trial. Pediatrics 2007 Mar;119(3):569-578 [FREE Full text] [doi: 10.1542/peds.2006-1884] [Medline: 17332210]
54. Guendelman S, Meade K, Benson M, Chen YQ, Samuels S. Improving asthma outcomes and self-management behaviors of inner-city children: a randomized trial of the Health Buddy interactive device and an asthma diary. Arch Pediatr Adolesc Med 2002 Feb;156(2):114-120. [Medline: 11814370]
55. Jan RL, Wang JY, Huang MC, Tseng SM, Su HJ, Liu LF. An internet-based interactive telemonitoring system for improving childhood asthma outcomes in Taiwan. Telemed J E Health 2007 Jun;13(3):257-268. [doi: 10.1089/tmj.2006.0053] [Medline: 17603828]
56. Prabhakaran L, Chee WY, Chua KC, Abisheganaden J, Wong WM. The use of text messaging to improve asthma control: a pilot study using the mobile phone short messaging service (SMS). J Telemed Telecare 2010 Jul;16(5):286-290. [doi: 10.1258/jtt.2010.090809] [Medline: 20576744]
57. Rasmussen LM, Phanareth K, Nolte H, Backer V. Internet-based monitoring of asthma: a long-term, randomized clinical study of 300 asthmatic subjects. J Allergy Clin Immunol 2005 Jun;115(6):1137-1142. [doi: 10.1016/j.jaci.2005.03.030] [Medline: 15940125]
58. Willems DC, Joore MA, Hendriks JJ, Nieman FH, Severens JL, Wouters EF. The effectiveness of nurse-led telemonitoring of asthma: results of a randomized controlled trial. J Eval Clin Pract 2008 Aug;14(4):600-609. [doi: 10.1111/j.1365-2753.2007.00936.x] [Medline: 19126178]
59. Lewis KE, Annandale JA, Warm DL, Rees SE, Hurlin C, Blyth H, et al. Does home telemonitoring after pulmonary rehabilitation reduce healthcare use in optimized COPD? A pilot randomized trial. COPD 2010 Feb;7(1):44-50. [doi: 10.3109/15412550903499555] [Medline: 20214462]
60. Lewis KE, Annandale JA, Warm DL, Hurlin C, Lewis MJ, Lewis L. Home telemonitoring and quality of life in stable, optimised chronic obstructive pulmonary disease. J Telemed Telecare 2010 Jul;16(5):253-259. [doi: 10.1258/jtt.2009.090907] [Medline: 20483881]
61. Nguyen HQ, Gill DP, Wolpin S, Steele BG, Benditt JO. Pilot study of a cell phone-based exercise persistence intervention post-rehabilitation for COPD. Int J Chron Obstruct Pulmon Dis 2009;4:301-313 [FREE Full text] [Medline: 19750190]
62. Carrasco MP, Salvador CH, Sagredo PG, Márquez-Montes J, González de Mingo MA, Fragua JA, et al. Impact of patient-general practitioner short-messages-based interaction on the control of hypertension in a follow-up service for low-to-medium risk hypertensive patients: a randomized controlled trial. IEEE Trans Inf Technol Biomed 2008 Nov;12(6):780-791. [doi: 10.1109/TITB.2008.926429] [Medline: 19000959]
63. Santamore WP, Homko CJ, Kashem A, McConnell TR, Bove AA. Using a telemedicine system to decrease cardiovascular disease risk in an underserved population: design, use, and interim results. Conf Proc IEEE Eng Med Biol Soc 2007;2007:3701-3704. [doi: 10.1109/IEMBS.2007.4353135] [Medline: 18002801]
64. Schwarz KA, Mion LC, Hudock D, Litman G. Telemonitoring of heart failure patients and their caregivers: a pilot randomized controlled trial. Prog Cardiovasc Nurs 2008;23(1):18-26. [Medline: 18326990]
65. DeVito Dabbs A, Dew MA, Myers B, Begey A, Hawkins R, Ren D, et al. Evaluation of a hand-held, computer-based intervention to promote early self-care behaviors after lung transplant. Clin Transplant 2009;23(4):537-545. [doi: 10.1111/j.1399-0012.2009.00992.x] [Medline: 19473201]
66. Yardley L, Joseph J, Michie S, Weal M, Wills G, Little P. Evaluation of a Web-based intervention providing tailored advice for self-management of minor respiratory symptoms: exploratory randomized controlled trial. J Med Internet Res 2010;12(4):e66 [FREE Full text] [doi: 10.2196/jmir.1599] [Medline: 21159599]
67. van der Meer V, Bakker MJ, van den Hout WB, Rabe KF, Sterk PJ, Kievit J, SMASHING (Self-Management in Asthma Supported by Hospitals‚ ICT‚ Nurses and General Practitioners) Study Group. Internet-based self-management plus education compared with usual care in asthma: a randomized trial. Ann Intern Med 2009 Jul 21;151(2):110-120. [Medline: 19620163]

68. Nguyen HQ, Donesky-Cuenco D, Wolpin S, Reinke LF, Benditt JO, Paul SM, et al. Randomized controlled trial of an internet-based versus face-to-face dyspnea self-management program for patients with chronic obstructive pulmonary disease: pilot study. J Med Internet Res 2008;10(2):e9 [FREE Full text] [doi: 10.2196/jmir.990] [Medline: 18417444]
69. Berger T, Caspar F, Richardson R, Kneubühler B, Sutter D, Andersson G. Internet-based treatment of social phobia: a randomized controlled trial comparing unguided with two types of guided self-help. Behav Res Ther 2011 Mar;49(3):158-169. [doi: 10.1016/j.brat.2010.12.007] [Medline: 21255767]
70. Bergström J, Andersson G, Ljótsson B, Rück C, Andréewitch S, Karlsson A, et al. Internet-versus group-administered cognitive behaviour therapy for panic disorder in a psychiatric setting: a randomised trial. BMC Psychiatry 2010;10:54 [FREE Full text] [doi: 10.1186/1471-244X-10-54] [Medline: 20598127]
71. Vernmark K, Lenndin J, Bjärehed J, Carlsson M, Karlsson J, Oberg J, et al. Internet administered guided self-help versus individualized e-mail therapy: A randomized trial of two versions of CBT for major depression. Behav Res Ther 2010 May;48(5):368-376. [doi: 10.1016/j.brat.2010.01.005] [Medline: 20152960]
72. Oerlemans S, van Cranenburgh O, Herremans PJ, Spreeuwenberg P, van Dulmen S. Intervening on cognitions and behavior in irritable bowel syndrome: A feasibility trial using PDAs. J Psychosom Res 2011 Mar;70(3):267-277. [doi: 10.1016/j.jpsychores.2010.09.018] [Medline: 21334498]
73. Glasgow RE, Nutting PA, King DK, Nelson CC, Cutter G, Gaglio B, et al. Randomized effectiveness trial of a computer-assisted intervention to improve diabetes care. Diabetes Care 2005 Jan;28(1):33-39. [Medline: 15616230]
74. Williams GC, Lynch M, Glasgow RE. Computer-assisted intervention improves patient-centered diabetes care by increasing autonomy support. Health Psychol 2007 Nov;26(6):728-734. [doi: 10.1037/0278-6133.26.6.728] [Medline: 18020845]
75. Wagner B, Knaevelsrud C, Maercker A. Internet-based cognitive-behavioral therapy for complicated grief: a randomized controlled trial. Death Stud 2006 Jun;30(5):429-453. [doi: 10.1080/07481180600614385] [Medline: 16610157]
76. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [Medline: 12867348]
77. 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 [FREE Full text] [Medline: 11816692]
78. Wolpin S, Berry D, Austin-Seymour M, Bush N, Fann JR, Halpenny B, et al. Acceptability of an Electronic Self-Report Assessment Program for patients with cancer. Comput Inform Nurs 2008;26(6):332-338. [doi: 10.1097/01.NCN.0000336464.79692.6a] [Medline: 19047882]
79. McCann L, Maguire R, Miller M, Kearney N. Patients' perceptions and experiences of using a mobile phone-based advanced symptom management system (ASyMS) to monitor and manage chemotherapy related toxicity. Eur J Cancer Care (Engl) 2009 Mar;18(2):156-164. [doi: 10.1111/j.1365-2354.2008.00938.x] [Medline: 19267731]
80. Willems DC, Joore MA, Hendriks JJ, Wouters EF, Severens JL. Cost-effectiveness of a nurse-led telemonitoring intervention based on peak expiratory flow measurements in asthmatics: results of a randomised controlled trial. Cost Eff Resour Alloc 2007 Jul;5:10 [FREE Full text] [doi: 10.1186/1478-7547-5-10] [Medline: 17662113]
81. Allen M, Iezzoni LI, Huang A, Huang L, Leveille SG. Improving patient-clinician communication about chronic conditions: description of an internet-based nurse E-coach intervention. Nurs Res 2008;57(2):107-112. [doi: 10.1097/01.NNR.0000313478.47379.98] [Medline: 18347482]
82. Liu JL, Wyatt JC. The case for randomized controlled trials to assess the impact of clinical information systems. J Am Med Inform Assoc 2011;18(2):173-180 [FREE Full text] [doi: 10.1136/jamia.2010.010306] [Medline: 21270132]
83. Talmon J, Ammenwerth E, Brender J, de Keizer N, Nykänen P, Rigby M. STARE-HI--Statement on reporting of evaluation studies in Health Informatics. Int J Med Inform 2009 Jan;78(1):1-9. [doi: 10.1016/j.ijmedinf.2008.09.002] [Medline: 18930696]
84. Sackett DL, Haynes RB, Tugwell P. Clinical Epidemiology: A Basic Science for Clinical Medicine. 1st edition. Boston, MA: Little, Brown; 1985.
85. Lange A, Rietdijk D, Hudcovicova M, van de Ven JP, Schrieken B, Emmelkamp PM. Interapy: a controlled randomized trial of the standardized treatment of posttraumatic stress through the internet. J Consult Clin Psychol 2003 Oct;71(5):901-909. [doi: 10.1037/0022-006X.71.5.901] [Medline: 14516238]
86. Knaevelsrud C, Maercker A. Internet-based treatment for PTSD reduces distress and facilitates the development of a strong therapeutic alliance: a randomized controlled clinical trial. BMC Psychiatry 2007;7:13 [FREE Full text] [doi: 10.1186/1471-244X-7-13] [Medline: 17442125]
87. Andersson G, Bergström J, Holländare F, Carlbring P, Kaldo V, Ekselius L. Internet-based self-help for depression: randomised controlled trial. Br J Psychiatry 2005 Nov;187:456-461 [FREE Full text] [doi: 10.1192/bjp.187.5.456] [Medline: 16260822]
88. Berger T, Hohl E, Caspar F. Internet-based treatment for social phobia: a randomized controlled trial. J Clin Psychol 2009 Oct;65(10):1021-1035. [doi: 10.1002/jclp.20603] [Medline: 19437505]
89. Perini S, Titov N, Andrews G. Clinician-assisted Internet-based treatment is effective for depression: randomized controlled trial. Aust N Z J Psychiatry 2009 Jun;43(6):571-578. [doi: 10.1080/00048670902873722] [Medline: 19440890]

90. Titov N, Andrews G, Choi I, Schwencke G, Mahoney A. Shyness 3: randomized controlled trial of guided versus unguided Internet-based CBT for social phobia. Aust N Z J Psychiatry 2008 Dec;42(12):1030-1040. [doi: 10.1080/00048670802512107] [Medline: 19016091]
91. Wims E, Titov N, Andrews G, Choi I. Clinician-assisted Internet-based treatment is effective for panic: A randomized controlled trial. Aust N Z J Psychiatry 2010 Jul;44(7):599-607. [doi: 10.3109/00048671003614171] [Medline: 20560847]
92. Baethge C. First authors in Deutsches Arzteblatt: women are catching up. The number of female authors in medical literature is increasing, but is still considerably lower than that of male authors and corresponds to the proportion of women working in academic medicine. Dtsch Arztebl Int 2008 Jul;105(28-29):507-509 [FREE Full text] [doi: 10.3238/arztebl.2008.0507] [Medline: 19626204]
93. Tones K, Tilford S. Health Promotion: Effectiveness, Efficiency and Equity. 3rd edition. Cheltenham: Nelson Thornes; 2001.
94. Mandl KD, Overhage JM, Wagner MM, Lober WB, Sebastiani P, Mostashari F, et al. Implementing syndromic surveillance: a practical guide informed by the early experience. J Am Med Inform Assoc 2004;11(2):141-150 [FREE Full text] [doi: 10.1197/jamia.M1356] [Medline: 14633933]
95. Reingold A. If syndromic surveillance is the answer, what is the question? Biosecur Bioterror 2003;1(2):77-81. [doi: 10.1089/153871303766275745] [Medline: 15040185]
96. Foldy SL. Linking better surveillance to better outcomes. MMWR Morb Mortal Wkly Rep 2004 Sep 24;53 Suppl:12-17 [FREE Full text] [Medline: 15717406]
97. Friesema IH, Koppeschaar CE, Donker GA, Dijkstra F, van Noort SP, Smallenburg R, et al. Internet-based monitoring of influenza-like illness in the general population: experience of five influenza seasons in The Netherlands. Vaccine 2009 Oct 23;27(45):6353-6357. [doi: 10.1016/j.vaccine.2009.05.042] [Medline: 19840672]
98. Marquet RL, Bartelds AI, van Noort SP, Koppeschaar CE, Paget J, Schellevis FG, et al. Internet-based monitoring of influenza-like illness (ILI) in the general population of the Netherlands during the 2003-2004 influenza season. BMC Public Health 2006;6:242 [FREE Full text] [doi: 10.1186/1471-2458-6-242] [Medline: 17018161]
99. van Noort SP, Muehlen M, Rebelo de Andrade H, Koppeschaar C, Lima Lourenço JM, Gomes MG. Gripenet: an internet-based system to monitor influenza-like illness uniformly across Europe. Euro Surveill 2007 Jul;12(7):E5-E6. [Medline: 17991409]
100. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
101. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
102. Gallaga OL. The Statesman. Can Mobile Phones Narrow the Digital Divide? 2012 May 8 URL: http://www.statesman.com/business/technology/can-mobile-phones-narrow-the-digital-divide-784691.html [accessed 2012-05-08] [WebCite Cache ID 67VQkZJ4Q]
103. Murphy S. Mashable, Inc. 2012 Mar 12. Doctors Believe Using Health Apps Will Cut Down on Visits URL: http://mashable.com/2012/03/12/mobile-health/ [accessed 2012-05-08] [WebCite Cache ID 67VRome4l]
104. Glasgow RE, Kaplan RM, Ockene JK, Fisher EB, Emmons KM. Patient-reported measures of psychosocial issues and health behavior should be added to electronic health records. Health Aff (Millwood) 2012 Mar;31(3):497-504. [doi: 10.1377/hlthaff.2010.1295] [Medline: 22392660]
105. Agerberg M. Läkartidningen. 2012. Journal på Nätet ska bli del i hälsopaket [109(6):266-7] URL: http://www.lakartidningen.se/store/articlepdf/1/17783/LKT1206s266_271.pdf [accessed 2012-05-29] [WebCite Cache ID 681TDrGs8]
106. Ekendahl M. Inera AB. 2012 Feb 8. En modell för att hantera journaluppgifter på internet för patienter: slutrapport från förstudien Din journal på nätet URL: http://www.inera.se/Invanartjanster/Din-journal-pa-natet/ [accessed 2012-05-08] [WebCite Cache ID 67VS9xf8h]
107. Hammond N. Marie Claire. 2012 Feb 22. Patients Told to Use Apps to Check Health URL: http://www.marieclaire.co.uk/news/health/535004/patients-told-to-use-apps-to-check-health.html [accessed 2012-05-08] [WebCite Cache ID 67VSetbli]
108. Kee E. ÜberGizmo. 2012 Feb 22. Sick? Doctor Will Prescribe You Smartphone Apps URL: http://www.ubergizmo.com/2012/02/doctor-prescribes-smartphone-apps/ [accessed 2012-05-08] [WebCite Cache ID 67VSptaaL]
109. Roberts M. BBC News Health. 2012 Feb 22. GPs 'to Prescribe Health Apps' URL: http://www.bbc.co.uk/news/health-17111092 [accessed 2012-05-08] [WebCite Cache ID 67VSzgtYJ]
110. Husain I. iMedicalApps. Dept of Health to Start Encouraging Physicians to Prescribe Health Apps: Good or bad idea? 2012 Feb 22 URL: http://www.imedicalapps.com/2012/02/department-health-physicians-prescribe-apps/ [accessed 2012-05-08] [WebCite Cache ID 67VTAnWSO]

Abbreviations

ICPC: International Classification of Primary Care
IOM: Institute of Medicine

MeSH: Medical Subject Headings
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses
RCT: Randomized Controlled Trial

Edited by G Eysenbach; submitted 09.06.12; peer-reviewed by E Ammenwerth, K Antypas; comments to author 27.06.12; revised version received 11.07.12; accepted 12.07.12; published 03.10.12.

Please cite as:
Johansen MA, Henriksen E, Horsch A, Schuster T, Berntsen GKR
Electronic Symptom Reporting Between Patient and Provider for Improved Health Care Service Quality: A Systematic Review of Randomized Controlled Trials. Part 1: State of the Art
J Med Internet Res 2012;14(5):e118
URL: http://www.jmir.org/2012/5/e118/
doi:10.2196/jmir.2214
PMID:23032300

Review

Using Noninferiority Tests to Evaluate Telemedicine and E-Health Services: Systematic Review

Per Egil Kummervold[1,2], PhD; Jan-Are K Johnsen[2,3], PhD; Stein Olav Skrøvseth[4], PhD; Rolf Wynn[2,3], MD, PhD

[1]Northern Research Institute, Tromsø, Norway
[2]University of Tromsø, Faculty of Health Sciences, Department of Clinical Medicine, Tromsø, Norway
[3]University Hospital of North Norway, Division of Addictions and Specialized Psychiatric Services, Tromsø, Norway
[4]University Hospital of North Norway, Norwegian Centre for Integrated Care and Telemedicine, Tromsø, Norway

Corresponding Author:
Per Egil Kummervold, PhD
Northern Research Institute
Postboks 6434 Forskningsparken
Tromsø, 9294
Norway
Phone: 47 41435795
Fax: 47 77629401
Email: peregil@norut.no

Abstract

Background: An increasing number of studies within the field of telemedicine and e-health are designed as noninferiority studies, aiming to show that the telemedicine/e-health solution is not inferior to the traditional way of treating patients.

Objective: The objective is to review and sum up the status of noninferiority studies within this field, describing advantages and pitfalls of this approach.

Methods: PubMed was searched according to defined criteria, and 16 relevant articles were identified from the period 2008-June 2011.

Results: Most of the studies were related to the fields of psychiatry and emergency medicine, and most were published in journals relating to these fields or in general scientific or general medicine journals. All the studies claimed to be noninferiority studies, but 7 out of 16 tested for statistical differences as a proxy of noninferiority.

Conclusions: The methodological quality of the studies varied. We discuss optimal procedures for future noninferiority studies within the field of telemedicine and e-health and situations in which this approach is most appropriate.

(J Med Internet Res 2012;14(5):e132) doi:10.2196/jmir.2169

KEYWORDS

noninferiority; non-inferiority; e-health; telemedicine

Introduction

In the field of telemedicine and e-health, there is often a need to demonstrate that a new solution/application is equal in quality or efficacy of treatment to the traditional or established way of treating patients. Demonstrating superiority of the new solution in terms of quality or efficacy of treatment is not always necessary, as the telemedicine/e-health solution/application may have other types of advantages, including saved travel time or saved costs. Testing that the new solution is not inferior to a traditional counterpart may therefore seem to be sufficient in many cases. As would be expected from this line of reasoning, there has been an increase in published studies within the field of telemedicine and e-health, using a noninferiority design, ie, studies that aim to show that the new telemedical solution is not of a lower quality than the established way of treating patients.

In the present study, we performed a systematic review of the published literature and found 16 studies [1-16] within the field of telemedicine and e-health as commonly defined: “E-health is an emerging field in the intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the Internet and related technologies” [17] and claiming to use noninferiority tests. We assessed the current status of the field and the strengths and weaknesses of the published studies.

The review aims to follow the criteria outlined in the PRISMA statement [18], but not all points are relevant since this is not a meta-analysis.

Why is a Failed Test of Superiority not the Same as Noninferiority?

A good starting point for understanding what an insignificant result really means is by considering the famous quote by astronomer Carl Sagan: “Absence of evidence is not evidence of absence” [19].

Consider an experiment where we evaluate a video-based telemedicine service called T. We have decided to test whether this service is superior to a traditional clinical treatment called C. For simplicity we are looking at one single aspect, the patient’s blood sugar levels.

We do a single sided *t* test of the mean blood sugar levels to check if T is superior to C, but we end up with a *P* value higher than .05. In other words, we have an insignificant result. Unfortunately, from a statistical point of view, this is nothing more than a failed test of superiority. It is not evidence that superiority does not exist. The only thing we are certain about is that our test was unable to prove any superiority.

The easiest way to understand this is that by reducing the number of participants, we are much more likely to get an insignificant result. It should be fairly obvious that a reduction in the number of participants is not making the groups more equal. It will result only in a study of lower quality and that is less able to detect if the new service is superior.

Including more persons in the trial will increase the chance of detecting superiority (if it exists). However, whenever we end up with an insignificant result, we are still facing Sagan’s observation that the absence of evidence is not evidence of absence.

If the ultimate goal is to prove that service T is not inferior to service C, the only way of approaching this is to first define what we mean by “inferior”. Note that “inferiority” is an empirical definition. When comparing two groups in medical trials, we never end up with exactly the same results, and what margins we define should be based on clinical considerations of what are meaningful margins, not upon our ability to measure them.

In noninferiority trials, we therefore first define that a margin (M) below C is to be considered as noninferior. How to set this margin is discussed in “Methods”. We then go on to test if T really is superior to this margin.

Methods

Statistical Considerations

Testing for equivalence has become an essential statistical tool in the process of securing approval for new generic drugs [20]. Equivalence testing makes it possible to show that the generic drug is no different from the drug it is going to replace, without having to compare the new drug directly to placebo. Technically, a noninferiority test is nothing other than a one-sided equivalence test, requiring fewer participants to obtain the same power.

As described in the Introduction, there are multiple reasons a failed test of superiority is insufficient for concluding noninferiority, among them is sample that’s too small *(ie, lack of power*) or that the study is not able to detect a real world difference *(ie, lack of assay sensitivity*).

In order to demonstrate *noninferiority*, we need to define a margin for when the test group is worse than the control group. We call this the *noninferiority margin* and let M represent this value. If we let T represent the efficacy of the new test service and C represent the efficacy of the control, noninferiority can be expressed as: $C - T < M$. This is the alternate hypothesis in a noninferiority trial. The corresponding null hypothesis will be $H_{0}: C - T \geq M$ [21]. According to the CONSORT statement, a recommended way of performing a noninferiority test is constructing a two-sided 90% confidence interval (since α in principle may be different from 5%, the precise definition of the CI is $1\text{-}2\alpha$), and if the upper limit of the interval is less than M, the null hypothesis is rejected, ie, noninferiority is considered proven [22].

Setting the margin (M) must be done at the start of the trial, and in a clinical trial it should be related to what experts find clinically relevant. Wellek [20] stresses that the setting of M must be done after careful consideration in every project but mentions that everyday experience indicates that most people would consider a difference between C and T of 10% (strict) and 20% (liberal) to be of the same magnitude. This is also similar to what the FDA suggests as the threshold for establishing bioequivalence [23].

However, not only the difference between C and T is relevant for setting M. The margin must also be set in a way that a certain amount of the real effect of the active control over nontreatment/placebo (C-P) is conserved. Within biomedicine, it is discussed how small M could be in relation to C-P, and values ranging from 50-80% have been mentioned [21,24]. Setting M too small could lead to proving that the trial (T) is noninferior to the control (C), while at the same time not being clinically superior to nontreatment (P).

In an ordinary trial, a significant result does automatically prove the ability to detect a difference—typically called the trial’s *assay sensitivity*. A noninferiority trial does not have built-in assay sensitivity. Even if we get a significant result proving $C - T < M$, it is not proven that the two treatments have an effect. In fact, in a situation where our tools did not detect anything, we would also end up with $C - T < M$. In cases where it is impossible to include a placebo, assay sensitivity must be established drawing on historical data.

Summing up, the following factors are essential in noninferiority trials:

1. Finding a clinical relevant definition of M. M should be independent of factors like variance and sample size. While some have suggested that M could be in the range of 10-20% of C, this needs to be set individually for each project and must

be done before the trial. It is not an error to clinically decide that M should be lower.

2. Making sure that M conserves the main effect between the active control and nontreatment. Values of M should be at least 50% of C-P.

3. Assuring assay sensitivity, either by including a placebo or by drawing on historical data.

Whether it is possible to find a formal determination of M and whether it is possible to prove assay sensitivity using historical data are both questions that are still discussed vigorously among statisticians [25].

Search Strategy and Selection

The inclusion criteria are English-language articles that apply accepted definitions of telemedicine or e-health [17] and that use noninferiority tests as part of their methodology. The search terms are meant to reflect these criteria. A search in PubMed found 36 articles meeting the search criteria, which are given in Table 1.

Specific technological channels were included (eg, videoconference, Internet) in order to include articles within an intersection of fields that is not clearly defined as telemedicine or e-health in the article’s title or abstract. After the search, articles were manually scanned to exclude articles not fullfilling the inclusion criterias. Eighteen articles were excluded because they were clearly unrelated to telemedicine or e-health (in most cases this was caused by abstracts with the words “video” or “Internet”). One additional article was excluded because the main article was available only in Japanese, and another article was excluded since it referred to other noninferiority trials only in the abstract. This left 16 articles for further analysis (Table 2). The strategy is outlined in Figure 1.

Of the included articles, three were from 2008, three from 2009, five from 2010, and five from 2011 (until June 2011). No articles meeting the inclusion criteria were published prior to 2008.

Figure 1. Strategy Flowchart.

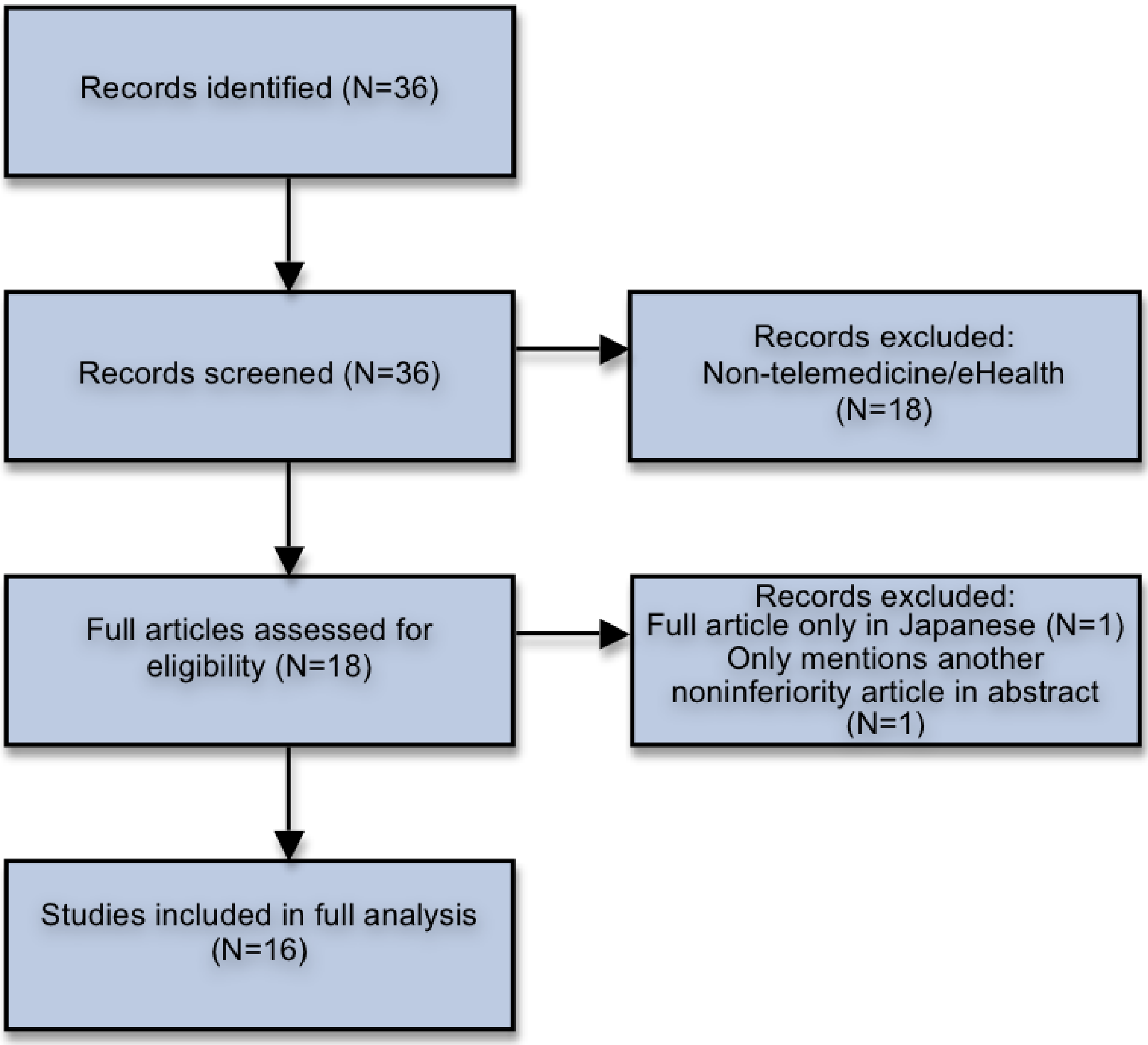


Table 1. Description of search criteria.

Search Criteria
(noninferior OR noninferiority OR non-inferiority OR ("non inferior") OR ("not inferior")) AND (telemedicine[Title/Abstract] OR videoconference[Title/Abstract] OR video[Title/Abstract] OR videoconferencing[Title/Abstract] OR online[Title/Abstract] OR Internet[Title/Abstract] OR ehealth[Title/Abstract] OR e-health[Title/Abstract])

Review Process

In the review of the articles, two reviewers (Authors 1 and 2) identified how the noninferiority margin was set and the reason that was provided for setting it. They also noted whether an actual noninferiority test was performed or if it was a test for difference. Finally, they registered how assay sensitivity was assured.

Results

Six of the included articles dealt with matters related to psychiatric treatment (post-traumatic stress disorder, generalized anxiety disorder, depression), four of the articles dealt with medical procedures particularly relevant to emergency medicine (vascular access, defibrillation, advanced life support), one was within the field of urology, one within rehabilitation after surgery, one within endocrinology, one within hematology, and two within medical communication studies. With regard to where the papers were published, only one was published in a telemedicine journal, five were published in emergency medicine journals, two in a psychiatric journal, one in an orthopedic surgery journal, one in an endocrinology journal, and six in general scientific or general medical journals.

The Setting the Noninferiority Margin

Various ways of defining the inferiority margin were used in the 16 articles reviewed (Table 2). In two articles [2,6], the inferiority margin was set to of 10% deviance from the main effect. In four articles [5,8,9,14], it was related to absolute values on validated questionnaires. In one of these articles [5], Cohen's $d = 0.5$ was used for setting the inferiority margin for some of the measures. Cohen's d was also used in another article [3], but here the margin was set to 0.2.

In one article [1], the margin was defined as 0.15SD. Another article [10] defined the noninferiority margin as a 10 percentage points difference between the proportions in the two groups.

Two studies referred to the lower bound of the confidence interval for the scores of the reference group. One of them [16] used the 90% confidence interval; the other [4] used a 95% confidence interval. In one article [11], the inferiority margin was set to RR = 0.95.

In the four remaining studies [7,12,13,15], the authors made no attempt to set an inferiority margin.

Reasons Given for Setting a Specific Margin

Five articles [2,5,8,9,14] referred to expert consultations or clinical relevance as the main source for setting the margin, while two of them [2,9] also stated that this value was similar to the value set in prior studies.

In one article [6], it was argued that 10% is a typical value in medical trials. One article [1] stated that setting of the noninferiority margin was guided by Cohen's [26] conventional criterion for small, medium, and large effect sizes.

One article [3] simply stated that the margin was defined as being relevant, while four articles [4,10,11,16] did not provide a reason.

In the remaining articles [7,12,13,15], the authors did not set a specific margin.

Testing for Inferiority

Another question is whether a noninferiority test was actually performed, ie, that it was tested that the target effect was larger than the noninferiority margin. This could be accomplished either by checking whether the entire confidence interval for the means difference was above the noninferiority margin or by calculating a P value.

Nine of the articles [1,3,5,6,8-11,14] involved tests for noninferiority against the noninferiority margin. All except [3] found a significant result.

Seven of the articles [2,4,7,12,13,15,16] tested for differences instead. When this did not provide a significant result, they claimed noninferiority.

Ensuring Assay Sensitivity

Four of the studies [2,13-15] had a pre-post design and were able to detect a significant difference between the start and end scores. This means that the studies had assay sensitivity. In some studies [1,5,6-9,16], one of several previously validated questionnaires was used—Patient Assessment of Communication during Telemedicine (PACT) [1], Liebowitz Social Anxiety Scale (LSAS) [5], Rapid HIV Pretest Information Comprehension [6], Clinician Administered PSTS Scale (CAPS) [7], State Trait Anger Expression Inventory-2 (STAXI-2) [8], Clinician Administered PSTS Scale (CAPS) [9], Novaco Anger Scale total score (NAS-T) [9], and AHA PALS Core Case Testing Checklist [16]. Since these questionnaires had previously shown significant results, it might be argued that this ensures assay sensitivity. It can be argued that [4] is in the same category, since measuring number of days in the therapeutic range is typically used when evaluating anticoagulants in other studies.

In [3], the authors included a noninferiority test but did not get a significant result. A difference test would show a significantly worse outcome for the treatment group, and the study does therefore, albeit indirectly, have assay sensitivity. In [10-12], we were not able to identify attempts at proving assay sensitivity.

Table 2. Articles included in review.

Reference	Margin	Reasoning	Test
Agha et al, 2009 [1]	0.15 SD	Guided by Cohen	Noninferiority
Chenkin et al, 2008 [2]	10% from mean	Clinically + prior studies	Difference
de Vries et al, 2010 [3]	Cohen's *d* = 0.2	Defined	Noninferiority
Harper & Pollock, 2011 [4]	Unclear: a) Within 5%, b) Lower bound 95% CI	No reason given	Difference
Hedman et al, 2011 [5]	Absolute value + Cohen's *d* = 0.5	Clinically + prior studies	Noninferiority
Merchant et al, 2009 [6]	10% from mean	Typical in medical trials	Noninferiority
Morland et al, 2011 [7]	Not set	Not relevant	Difference
Morland et al, 2010 [8]	Absolute value	Clinically	Noninferiority
Morland et al, 2009 [9]	Absolute value	Clinically + prior studies	Noninferiority
Mpotos et al, 2011 [10]	10 percentage points difference in proportions	No reason given	Noninferiority
Munger et al, 2008 [11]	RR=0.95	No reason given	Noninferiority
Péres-Ferre et al, 2010 [12]	Not set	No reason given	Difference
Robinson et al, 2010 [13]	Not set	Not relevant	Difference
Russell et al, 2011 [14]	Absolute value	Clinically	Noninferiority
Titov et al, 2010 [15]	Not set	Not relevant	Difference
Weeks & Molsberry, 2008 [16]	Lower bound 90% CI	No reason given	Difference

Discussion

As the results show, there are considerable variations in the way the noninferiority trials are performed. The 16 included articles should encompass the majority of the studies that claim to be noninferiority trials within the field of telemedicine and e-health, but a few that have not been indexed in PubMed might have been missed. While the study method seems to be growing in popularity, it is still in its infancy. Most current use of noninferiority trials is within biomedicine, and there are, as we have shown, only a few examples of use within telemedicine and e-health. While noninferiority trials within biomedicine can serve as an inspiration, differences between the fields make it difficult to copy the approaches used in biomedical trials. Below, we discuss some of the central elements of noninferiority trials and how they can be applied to studies within telemedicine and e-health.

Setting the Margin

To prove that something is equal, or not inferior, we need to define what we mean by equality or noninferiority. This is mainly a clinical issue that primarily should be assessed by experts within the field. Some very rough guidelines have been referred to, and values within 10-20% appear to be considered fairly equal in the literature. What is clinically relevant cannot be decided by this value alone. In some cases, a 10% difference can have enormous impact, while in other cases this value is clinically irrelevant. Only five of the articles included in our review referred to the concept of clinical relevance.

There are other guidelines stating that the margin should be set so that a majority of the effect between the control (C) and the nontreatment (P) should be preserved. In trials where the nontreatment group is not included, the researcher will have to estimate the effect of C-P based on previous trials. This is not a luxury that many telemedicine/e-health trials have.

Proving Noninferiority

When performing a traditional hypothesis test, a *P* value higher than the significance level does not provide evidence that the null hypothesis is true. It simply means that the evidence is not strong enough to reject the null hypothesis with sufficient confidence. Indeed, it is possible that a study that results in a *P* value above the significance level will be a positive contribution to a future meta-analysis in proving that there is a difference. The most surprising result of our review is that almost half (7 of 16) of the articles seem to disregard this fact. They actually performed tests of difference, and their main argument for noninferiority was that the difference test gave an insignificant result.

Assay Sensitivity

One of the main driving forces in the popularity of noninferiority and equality testing within biomedicine is that it enables doing evidence-based medicine without including a nontreatment group. In some cases it might be ethically unacceptable to introduce a placebo. In other cases, this is primarily a question of cost saving. It might be fair to say that the increasing use of noninferiority and equality testing is related to the growth of so-called explanatory or pragmatic trials, where the main question is not whether a treatment is effective but whether the treatment is worthwhile using in a clinical setting [27,28].

Ideally, assay sensitivity should be proven by a previous trial or a meta-analysis of multiple previous trials. It is difficult to replicate studies in this fashion within the field of telemedicine/e-health, and none of the studies examined in our review did this. However, 7 of the 16 studies did use previously validated questionnaires, an alternative that in many cases actually may be sufficient.

There were also four studies that included a placebo/no treatment. For simply proving assay sensitivity, this is definitely sufficient. It is, however, a bit contrary to the original purpose of equality and noninferiority tests, which is to be able to do without a placebo/no treatment group.

The review did also identify three studies where there were no explicit indications in the articles that assay sensitivity had been established. The authors might, however, have carried out such procedures without reporting it.

Recommendations

As the analysis shows, the fundamentals of noninferiority testing can be daunting to use in practice, especially for authors that are new to this type of analysis. We recommend that authors pay close attention to the extended CONSORT guidelines for noninferiority testing to the extent that they are applicable for the study in question [22]. We believe that one of the articles by Morland et al [8] provides a good example of how a noninferiority article can be performed and reported. Morland et al [9] also provide a more detailed methodological discussion concerning noninferiority trial design.

Performing a noninferiority study requires, as with any choice of statistical analysis, strict adherence to protocol to avoid fishing for positive results, which will dramatically affect the probability of type II errors. In particular, the noninferiority margin must be set before the study starts. Setting the margin after investigating the data means the investigator essentially can obtain any result wanted. Similarly, if an investigator performs a standard superiority trial and finds a nonsignificant result, the study should never be transformed into a noninferiority study. The intent of determining noninferiority must be clear from the outset.

When there is sparse evidence for assay sensitivity, such as if there are few studies to base the analysis on, noninferiority testing may not be the best option. Assay sensitivity is essential for doing a proper noninferiority study since without it, the study could end up proving that the intervention is no worse than doing nothing (ie, does no harm). In such settings, it should be considered if another type of design is more appropriate, eg, an economic evaluation.

Conclusions

Noninferiority testing clearly has a place within telemedicine and e-health. It is, however, always a much more daunting task to prove that something (like a difference) does not exist than to prove that it does exist. As we have discussed in our review, noninferiority trials are not a magic shortcut to solving this fundamental challenge.

While several of the trials included in this review are of a high quality, the review also brings to light an apparent lack of awareness of the pitfalls of performing noninferiority trials. We recommend more stringent adherence to the basic principles of noninferiority testing. We have discussed some points that should be given specific attention, including the importance of not mistaking a failed difference test for proof of noninferiority and the importance of setting a clinically relevant noninferiority margin.

Authors' Contributions

The study was conceived and designed by PEK. All authors analyzed the data and wrote the paper.

Conflicts of Interest

None declared.

An ethics statement was not required for this work. No direct funding was received for this study. The authors were personally salaried by their institutions during the period of writing (though no specific salary was set aside or given for the writing of this paper).

References

1. Agha Z, Schapira RM, Laud PW, McNutt G, Roter DL. Patient satisfaction with physician-patient communication during telemedicine. Telemed J E Health 2009 Nov;15(9):830-839. [doi: 10.1089/tmj.2009.0030] [Medline: 19919189]
2. Chenkin J, Lee S, Huynh T, Bandiera G. Procedures can be learned on the Web: a randomized study of ultrasound-guided vascular access training. Acad Emerg Med 2008 Oct;15(10):949-954. [doi: 10.1111/j.1553-2712.2008.00231.x] [Medline: 18778380]
3. de Vries W, Turner NM, Monsieurs KG, Bierens JJ, Koster RW. Comparison of instructor-led automated external defibrillation training and three alternative DVD-based training methods. Resuscitation 2010 Aug;81(8):1004-1009. [doi: 10.1016/j.resuscitation.2010.04.006] [Medline: 20483519]
4. Harper P, Pollock D. Improved anticoagulant control in patients using home international normalized ratio testing and decision support provided through the Internet. Intern Med J 2011 Apr;41(4):332-337. [doi: 10.1111/j.1445-5994.2010.02282.x] [Medline: 20546057]
5. Hedman E, Andersson G, Ljótsson B, Andersson E, Rück C, Mörtberg E, et al. Internet-based cognitive behavior therapy vs. cognitive behavioral group therapy for social anxiety disorder: a randomized controlled non-inferiority trial. PLoS One 2011;6(3):e18001 [FREE Full text] [doi: 10.1371/journal.pone.0018001] [Medline: 21483704]

6. Merchant RC, Clark MA, Mayer KH, Seage Iii GR, DeGruttola VG, Becker BM. Video as an effective method to deliver pretest information for rapid human immunodeficiency testing. Acad Emerg Med 2009 Feb;16(2):124-135. [doi: 10.1111/j.1553-2712.2008.00326.x] [Medline: 19120050]
7. Morland LA, Hynes AK, Mackintosh MA, Resick PA, Chard KM. Group cognitive processing therapy delivered to veterans via telehealth: a pilot cohort. J Trauma Stress 2011 Aug;24(4):465-469. [doi: 10.1002/jts.20661] [Medline: 21793047]
8. Morland LA, Greene CJ, Rosen CS, Foy D, Reilly P, Shore J, et al. Telemedicine for anger management therapy in a rural population of combat veterans with posttraumatic stress disorder: a randomized noninferiority trial. J Clin Psychiatry 2010 Jul;71(7):855-863. [doi: 10.4088/JCP.09m05604blu] [Medline: 20122374]
9. Morland LA, Greene CJ, Rosen C, Mauldin PD, Frueh BC. Issues in the design of a randomized noninferiority clinical trial of telemental health psychotherapy for rural combat veterans with PTSD. Contemp Clin Trials 2009 Nov;30(6):513-522. [doi: 10.1016/j.cct.2009.06.006] [Medline: 19576299]
10. Mpotos N, Lemoyne S, Calle PA, Deschepper E, Valcke M, Monsieurs KG. Combining video instruction followed by voice feedback in a self-learning station for acquisition of Basic Life Support skills: a randomised non-inferiority trial. Resuscitation 2011 Jul;82(7):896-901. [doi: 10.1016/j.resuscitation.2011.02.024] [Medline: 21444145]
11. Munger MA, Stoddard GJ, Wenner AR, Bachman JW, Jurige JH, Poe L, et al. Safety of prescribing PDE-5 inhibitors via e-medicine vs traditional medicine. Mayo Clin Proc 2008 Aug;83(8):890-896. [doi: 10.4065/83.8.890] [Medline: 18674473]
12. Pérez-Ferre N, Galindo M, Fernández MD, Velasco V, Runkle I, de la Cruz MJ, et al. The outcomes of gestational diabetes mellitus after a telecare approach are not inferior to traditional outpatient clinic visits. Int J Endocrinol 2010;2010:386941 [FREE Full text] [doi: 10.1155/2010/386941] [Medline: 20628517]
13. Robinson E, Titov N, Andrews G, McIntyre K, Schwencke G, Solley K. Internet treatment for generalized anxiety disorder: a randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010;5(6):e10942 [FREE Full text] [doi: 10.1371/journal.pone.0010942] [Medline: 20532167]
14. Russell TG, Buttrum P, Wootton R, Jull GA. Internet-based outpatient telerehabilitation for patients following total knee arthroplasty: a randomized controlled trial. J Bone Joint Surg Am 2011 Jan 19;93(2):113-120. [doi: 10.2106/JBJS.I.01375] [Medline: 21248209]
15. Titov N, Andrews G, Davies M, McIntyre K, Robinson E, Solley K. Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010;5(6):e10939 [FREE Full text] [doi: 10.1371/journal.pone.0010939] [Medline: 20544030]
16. Weeks DL, Molsberry DM. Pediatric advanced life support re-training by videoconferencing compared to face-to-face instruction: a planned non-inferiority trial. Resuscitation 2008 Oct;79(1):109-117. [doi: 10.1016/j.resuscitation.2008.05.008] [Medline: 18617310]
17. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
18. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
19. Sagan C. The demon-haunted world: science as a candle in the dark. New York: Random House; 1996.
20. Wellek S. Testing statistical hypotheses of equivalence and noninferiority. Second edition. International Statistical Review 2012;80:188. [doi: 10.1111/j.1751-5823.2012.00179_12.x]
21. D'Agostino R, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. Stat Med 2003 Jan 30;22(2):169-186. [doi: 10.1002/sim.1425] [Medline: 12520555]
22. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ, CONSORT Group. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. JAMA 2006 Mar 8;295(10):1152-1160. [doi: 10.1001/jama.295.10.1152] [Medline: 16522836]
23. Food and Drug Administration. Guidance for Industry. Rockville, MD: Center for Drug Evaluation and Research; 2001. Statistical Approaches to Establishing Bioequivalence URL: http://www.fda.gov/downloads/Drugs/.../Guidances/ucm070244.pdf [accessed 2012-09-21] [WebCite Cache ID 6AqUslP02]
24. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. BMJ 1996 Jul 6;313(7048):36-39. [Medline: 8664772]
25. James Hung HM, Wang SJ, Tsong YL, Lawrence J, O'Neil RT. Some fundamental issues with non-inferiority testing in active controlled trials. Stat Med 2003 Jan 30;22(2):213-225. [doi: 10.1002/sim.1315] [Medline: 12520558]
26. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, N.J: L. Erlbaum Associates; 1988.
27. King KM, Thompson DR. Pragmatic trials: is this a useful method in nursing research? J Clin Nurs 2008 Jun;17(11):1401-1402. [doi: 10.1111/j.1365-2702.2008.02297.x] [Medline: 18482137]
28. Ware JH, Hamel MB. Pragmatic trials - guides to better patient care. N Engl J Med 2011 May 5;364(18):1685-1687. [doi: 10.1056/NEJMp1103502] [Medline: 21542739]

Edited by G Eysenbach; submitted 10.05.12; peer-reviewed by L Morland, CM Chu, A Garg; comments to author 31.05.12; revised version received 15.07.12; accepted 04.09.12; published 28.09.12.

Please cite as:
Kummervold PE, Johnsen JAK, Skrøvseth SO, Wynn R
Using Noninferiority Tests to Evaluate Telemedicine and E-Health Services: Systematic Review
J Med Internet Res 2012;14(5):e132
URL: http://www.jmir.org/2012/5/e132/
doi:10.2196/jmir.2169
PMID:23022989

Original Paper

Electronic Symptom Reporting Between Patient and Provider for Improved Health Care Service Quality: A Systematic Review of Randomized Controlled Trials. Part 2: Methodological Quality and Effects

Monika Alise Johansen[1], MSc; Gro K Rosvold Berntsen[1], MD, PhD; Tibor Schuster[2], PhD; Eva Henriksen[1], MSc; Alexander Horsch[2,3], PhD

[1]Norwegian Centre for Integrated Care and Telemedicine, University Hospital of North Norway, Tromsø, Norway
[2]Institute of Medical Statistics and Epidemiology, Technische Universität München, Munich, Germany
[3]Department of Telemedicine and e-Health, Institute of Clinical Medicine, University of Tromsø, Tromsø, Norway

Corresponding Author:
Monika Alise Johansen, MSc
Norwegian Centre for Integrated Care and Telemedicine
University Hospital of North Norway
Sykehusveien 23
pb35
Tromsø, 9038
Norway
Phone: 47 906 84 407
Fax: 47 777 54 099
Email: monika.alise.johansen@telemed.no

Abstract

Background: We conducted in two parts a systematic review of randomized controlled trials (RCTs) on electronic symptom reporting between patients and providers to improve health care service quality. Part 1 reviewed the typology of patient groups, health service innovations, and research targets. Four innovation categories were identified: consultation support, monitoring with clinician support, self-management with clinician support, and therapy.

Objective: To assess the methodological quality of the RCTs, and summarize effects and benefits from the methodologically best studies.

Methods: We searched Medline, EMBASE, PsycINFO, Cochrane Central Register of Controlled Trials, and IEEE Xplore for original studies presented in English-language articles between 1990 and November 2011. Risk of bias and feasibility were judged according to the Cochrane recommendation, and theoretical evidence and preclinical testing were evaluated according to the Framework for Design and Evaluation of Complex Interventions to Improve Health. Three authors assessed the risk of bias and two authors extracted the effect data independently. Disagreement regarding bias assessment, extraction, and interpretation of results were resolved by consensus discussions.

Results: Of 642 records identified, we included 32 articles representing 29 studies. No articles fulfilled all quality requirements. All interventions were feasible to implement in a real-life setting, and theoretical evidence was provided for almost all studies. However, preclinical testing was reported in only a third of the articles. We judged three-quarters of the articles to have low risk for random sequence allocation and approximately half of the articles to have low risk for the following biases: allocation concealment, incomplete outcome data, and selective reporting. Slightly more than one fifth of the articles were judged as low risk for blinding of outcome assessment. Only 1 article had low risk of bias for blinding of participants and personnel. We excluded 12 articles showing high risk or unclear risk for both selective reporting and blinding of outcome assessment from the effect assessment. The authors' hypothesis was confirmed for 13 (65%) of the 20 remaining articles. Articles on self-management support were of higher quality, allowing us to assess effects in a larger proportion of studies. All except one self-management interventions were equally effective to or better than the control option. The self-management articles document substantial benefits for patients, and partly also for health professionals and the health care system.

Conclusion: Electronic symptom reporting between patients and providers is an exciting area of development for health services. However, the research generally is of low quality. The field would benefit from increased focus on methods for conducting and reporting RCTs. It appears particularly important to improve blinding of outcome assessment and to precisely define primary outcomes to avoid selective reporting. Supporting self-management seems to be especially promising, but consultation support also shows encouraging results.

(J Med Internet Res 2012;14(5):e126) doi:10.2196/jmir.2216

KEYWORDS

Electronic symptom reporting; physician-patient relationship; patient participation; shared decision making; review; consultation; monitoring; self-management; bias

Introduction

This paper presents the second part of a comprehensive review of randomized controlled trials (RCTs) focusing on electronic communication between patient and provider to improve health care service quality. This patient-provider partnership is defined by patients or parents reporting symptoms or health information electronically [1]. The patient reports to health care personnel, an institution, or a system, where the receiver processes and interprets the data and provides feedback to the patient [1]. The general purpose is improved health care service quality and efficiency, for example, by improving or avoiding a consultation [1].

Part 1 of this review identified the following typology of the field in terms of [2] patient groups, health service innovations, and research targets:

- Five specific patient groups mainly based on the International Classification of Primary Care (ICPC) definitions [3]: cancer, respiratory and lung diseases, cardiovascular diseases, psychiatry, and diabetes.
- Four health service innovation categories: consultation support, monitoring with clinician support, self-management with clinician support, and therapy.
- Research targets: consultation support studies primarily aimed to improve patient-centered care and secondarily to provide health benefits. Monitoring studies focused on health benefits, patient-centeredness outcomes, and reduced health care costs. Self-management studies mainly aimed for health benefits and secondarily patient-centered outcomes.

This part of the review looked into the methodological quality on the RCTs, and summarized effects and benefits of electronic symptom reporting of the methodologically best RCTs.

Effects and Benefits of Electronic Symptom Reporting

It is possible to achieve effects of electronic symptom reporting at the health care professional, health care system, and patient levels.

At the *health care professional level*, electronic symptom reporting might support the diagnostic process, and thus also make better use of the health professional's time. Determining the patient's main problem or concern is often demanding for the physician [4]. The way in which patients present their problems, and the sequence, importance, and severity of symptoms influence the physician's interpretation. Likewise, studies of interview styles show that physicians elicit only about 50% of the medical information considered important in a consultation [5]. Health care professionals may also be challenged by patients' difficulties in correctly remembering symptom levels beyond the past several days [6] and older patients' omission of many symptoms [7] during a consultation. On the other hand, we know that people in general report a higher number of and more serious symptoms when using computer-mediated communication than in face-to-face encounters or phone conversations [8] (p. 28-29).

At the *health care system level*, time and money might be saved [9]. Trials of electronic symptom reporting suggest that it may be possible to substitute about one-third or more of face-to-face consultations in primary care settings [10,11]. It is probably also possible to reduce the number of consultations in specialist care. Internationally, up to 24% of surgeries are cancelled the same day as they are scheduled [12-14], which is a major expense for health care systems [13]. Patient information might be outdated, inadequate, or even wrong at the time of surgery [15,16], and nearly half of the cancellations could have been avoided with an adequate patient information review and update [13].

At the *patient level*, it is possible to improve documentation of key variables that affect service quality and safety [17]. Patients embrace the idea of reporting symptoms electronically before their visit to the doctor [18-21] and believe it will improve the quality of care and effectiveness during the encounter [19,20]. Wald et al demonstrated that 70% of 2027 patients actually submitted symptom information before consultation, and that patients felt more prepared for the visit and that their provider had more accurate information about them [22]. These findings give rise to the assumption that electronic symptom reporting might be a useful tool to strengthen patient empowerment. Patients who report symptoms electronically prior to a consultation are given a chance to convey their problems in a less-stressful situation. This may result in patients having a preformed clear and concise understanding of their own clinical problems, while at the same time it provides updated patient information and documentation that can be saved in the electronic patient record. This may improve the diagnostic process and result in better patient management and care planning.

However, our preliminary screening found that studies in the field typically are small in terms of number of patients involved and are best described as feasibility studies [1]. Many of the studies focused on technologies rather than health effects, and most of them seem to have been underpowered to document clinical effects or specific benefits for health care professionals, health care systems, or patients [1]. No systematic review has yet addressed this topic, to the best of our knowledge, which makes it difficult for innovators and researchers to assess which of these choices are most promising and have the strongest potential for development on a larger scale.

Methodological Quality

We wanted to limit our work to the most mature stage of a complex intervention before implementation, the RCTs [23,24]. Unfortunately, the overall quality of RCT reporting is not always satisfactory [25]. Studies of low methodological quality typically tend to report better treatment effects than do studies of high quality [26-28]. Despite the development of guidelines to improve RCT reporting [29], it is still necessary to assess the methodological quality of RCTs, in our case for studies on electronic symptom reporting.

Objectives

The overall aim of this review was to systematically assemble the knowledge gained from RCTs focusing on electronic communication between patient and provider to improve health care service quality.

The objective for this second part of the review was to (1) assess the methodological quality of the RCTs identified in the first part of the review, and (2) summarize effects and benefits of electronic symptom reporting from data published in the methodologically best RCT articles. The benefits will be presented with regard to patients, health care professionals, and health care systems.

Methods

The review in general followed the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) recommendations [30]. To further improve the quality, we consulted the Cochrane handbook [31] for data extraction and assessment of methodological quality. The group conducting the review has multidisciplinary background, including experience in medical and epidemiological research (GB, AH, TS), RCT methodology and statistics (TS, GB, AH), telemedicine and medical informatics (MAJ, EH, AH, TS), theoretical knowledge of electronic symptom reporting (MAJ, EH), and experience from earlier review work (AH, GB, TS).

Inclusion and Exclusion Criteria

We included original RCTs if patients or parents reported recent symptoms or health information electronically, either to clinical health care personnel or to a system, where the receiver processed and interpreted the data for health care purposes and provided feedback. Feedback did not need to be given electronically. The focus was on asynchronous systems that can be established within the health care system. If the control group reported symptoms or health information, this information was not received by the health care professional or system. For a detailed description of the inclusion and exclusion criteria, see part 1 of the review [2].

Search Methods for Identification of Studies

We searched Medline, EMBASE, PsycINFO, the Cochrane Central Register of Controlled Trials, and IEEE Xplore to retrieve RCTs about human medicine presented in the English language, published from 1990 to November 2011. For a detailed description of the search methods, see part 1 of the review [2].

Selection of Studies and Data Extraction

Studies (abstract and full text) were selected independently by two authors (MAJ and EH), and all disagreements were resolved by consensus discussions. We extracted 84 variables from each included article, guided by the Cochrane data collection checklist [31] (Table 7.3.a in the Cochrane handbook). Variables defined as especially relevant for this specific review were included.

Extracted variables focused on the assessment of methodological quality, including evaluation of the risk of bias in the results, and outcome measures and results relevant to electronic symptom reporting. A full presentation of the extracted variables and the citations can be found on the website of the Norwegian Centre for Integrated Care and Telemedicine [32]. A comprehensive description regarding selection of studies and data extraction and management is given in part 1 of the review [2].

Assessment of Methodological Quality

We assessed the methodological quality of each article, including risk of bias and three additional variables reflecting feasibility, theoretical evidence, and preclinical testing.

Risk of bias was assessed according to Cochrane's recommended domain-based evaluation, the criteria for judging risk of bias [31] (Table 8.5.d in the Cochrane handbook), and judged as low risk, unclear risk, or high risk.

Theoretical evidence and preclinical testing are both recommended as part of a framework for the design and evaluation of complex interventions to improve health [23,24]. An RCT should rest firmly on both a theoretical foundation and practical testing of how that theory can be applied in a specific context. Without such prior exploration, a nonsignificant finding may result from several causes that have nothing to do with the intervention itself, leading to a wrong conclusion. Thus, we included three additional quality assessment variables, referring to (1) whether implementing the intervention as planned is feasible or likely in a real-life setting [31], (2) theoretical evidence that the intervention might have the desired effects, and (3) preclinical testing, referring to the process of operationalizing theories through pilot trials and feasibility or acceptability testing [23,24].

The "assessment of the overall risk of bias involves consideration of the relative importance of different [bias] domains" and the review author's judgments "about which domains are most important in the current review" [31] (chapter 8.7). Given the nature of telemedicine and eHealth innovations, blinding of participants and personnel is extremely challenging. We thus did not consider this bias to be crucial to the quality judgment of the articles. As it is difficult to blind participants and personnel in electronic symptom reporting settings, we attached more importance to the blinding of outcome assessment. This kind of blinding is both possible and realistic to achieve, and might affect the study result. We also considered selective reporting to be important, since it indicates post hoc selection of a subset of the original analyses performed [31] (chapter 8.14.1), while typically omitting the negative analyses answering the original research question. Selective reporting thus causes publication bias, as negative results tend to be left unreported, and spurious random findings are highlighted instead. We considered a low risk of bias for selective reporting and for blinding of outcome assessment to be the best indicators for identification of studies with high methodological quality.

Incomplete outcome data refers both to attrition and to exclusion of participants through as-treated or other subgroup analyses. When assessing selective reporting, we accepted that a primary outcome variable could be represented through a group of well-defined measures, as long as authors reported all measures and time points properly and completely in the results section. On the other hand, we assessed studies as having a high risk for selective reporting if we found any incongruence between the published protocol and the reported primary results, or if the variable used to make power calculations was not part of the reported primary outcome measures. We applied the same logic if the authors' main conclusions did not rely on previously defined outcome parameters and therefore had to be regarded as limitedly interpretable post hoc findings.

The blinding of outcome assessment risk of bias was judged as high if there was no information indicating involvement of any independent personnel for assessment of outcome other than those performing the intervention. If the patients were the outcome assessors for the primary outcome, and all the patients had access to a common online discussion page, we regarded this as having knowledge about which intervention they and other patients received.

The Cochrane criteria for the unclear risk judgment is primarily defined as "insufficient information to permit judgment of 'low risk' or 'high risk'" [31] (Table 8.5.d in the Cochrane handbook). This means that everything in the unclear category might be of good or bad quality. However, the fact that the author did not report satisfactorily for us to make a judgment is in itself a bias, which is why we combined the high and unclear category in the analysis of the total bias results.

Three of the authors (TS, GB, MAJ) assessed the risk of bias independently. In all cases of disagreement, a discussion took place until consensus was achieved.

Effects of Electronic Symptom Reporting

We agree that it is not acceptable to "present analyses and interpretations based on all studies, ignoring flaws identified during the assessment of risk of bias" [31] (chapter 8.8.1). Thus, we excluded articles found to be at high risk or unclear risk for both selective reporting and blinding of outcome assessors from the subsequent analysis of reported effects.

For all other articles, we extracted the primary outcome and present it in an effect table according to the article's health service innovation category. Some articles defined more than one outcome variable as their primary outcome. In these cases, we chose the first variable presented in the article's text to be included in our effect table.

Since only a few of these RCTs had a follow-up after the intervention, we chose the immediate postintervention outcome when extracting effects. We calculated within- and between-group pre- to postintervention differences and report the extracted P value for the between-group difference. Studies were defined as either equivalence studies (authors hypothesized that the study arms would be equivalent in terms of the effect measure) or as superiority studies (authors hypothesized that one arm would be superior to the other or others in terms of the chosen effect measure). If the authors' hypothesis was confirmed, we classified the study as positive; otherwise it was negative. Two of the authors (MJ and GB) extracted the effect data independently. In case of disagreement regarding what to extract and how to interpret the results, a discussion to reach consensus was reached.

In addition to the primary outcome effects, other extracted results from the articles with acceptable quality are reported for each health service innovation according to who might benefit: patients, health professionals, or the health care system. The reporting makes use of the Institute of Medicine (IOM) definitions stating that health care should be safe, effective in terms of health benefits, patient centered, timely, efficient (reduced time, reduced health care costs for the health system, and resource utilization of the health professional), and equitable [33]. The extracted outcome variables are based on the Cochrane recommendations, in addition to a set of variables that we developed. The cross-link between who benefits, the extracted outcome variables, and the areas of health service quality defined by IOM is presented in Table 1 in part 1 of the review [2].

Results

Main Background Data

Of 642 records identified and 444 abstracts reviewed, 32 articles presenting 29 studies were included [34-65] (see Figure 1 in part 1 [2]). The 32 articles were published from 2002 to 2011, with 24 of them being published in the last 5 years; 27 studies were conducted in Western countries, 12 of these in the United States.

All except 2 studies were designed as parallel studies with random allocation of patients. Of the parallel studies, 4 had three arms [38,47,59,61] and the others had two. A total of 2 studies were based on cluster randomization, 1 on randomized primary care practices [63,64], and 1 on clinics [39]. All studies focused on both genders, and the studies included on average 60% females (varying from 37.5% to 93%).

Methodological Quality

Even if we accept that patients and personnel were not blinded, no articles met all the quality requirements, and many articles satisfied few methodological quality criteria (Table 1, low risk or yes). Only 2 of the articles, Bergström et al [60] and Schwarz et al [54], had a low risk for all types of bias except blinding of participants and personnel. However, they did not fulfill the preclinical testing requirements. For 3 of the articles, Boyes et al [35], Santamore et al [53], and Williams et al [64], we found no types of bias to be at low risk.

All articles had interventions that could be implemented as planned in a real-life setting. Thus, this aspect is not included in Table 1. Theoretical evidence was provided in almost all articles. However, preclinical testing was properly provided for only about a third of the articles.

The quality assessment with regard to random sequence allocation was the bias domain with the best results, with three-quarters of the articles judged to have low risk. For only about half of the articles, the risk of bias for allocation concealment, incomplete outcome data, and selective reporting was judged to be low risk. With regard to selective reporting, 25% of the articles used several primary outcomes, and 28% had not defined or remained unclear regarding the primary outcome (Tables 4-7 in part 1, [2]).

For barely more than one-fifth of the articles, we judged the blinding of outcome assessment bias to be low risk. High risk of bias due to inadequate concealment of the allocated intervention from participants and personnel during the study is very challenging in telemedicine and eHealth research. This was, not surprisingly, achieved for only 1 of the included articles, Yardley et al [56].

Effects of Electronic Symptom Reporting

We excluded 12 articles assigned a high risk or unclear risk for both selective reporting and blinding of outcome assessments from the following effect report. We excluded 3 consultation support articles: Berry et al [34], Boyes et al [35], and Stevens et al [39]. We also excluded 8 monitoring articles (representing 6 studies): Chan et al [42], Chan et al [43], Jan et al [45], Kearney et al [41], Lewis et al (quality of life study) [50], Nguyen et al [51], Prabhakaran et al [46], and Santamore et al [53]. Only the secondary analysis of 1 self-management study was excluded (Williams et al [64]).

The reported effects (Table 2, Table 3, Table 4, and Table 5) show that the authors' hypothesis was confirmed in 13 (65%) of the 20 remaining articles. Interpreting the hypothesis as negative (no) or positive (yes) for primary outcome depended on whether the intervention hypothesis relative to the control condition was stated as equivalent or superior. We considered 4 of the studies to be equivalence studies, in all of which the authors' hypothesis was confirmed.

Overall Picture of Evidence

Multimedia Appendix 1 shows that the 20 RCTs with acceptable quality included a total of 3991 patients (ie, 200 patients on average per study). The average number of patients per RCT per combination of patient group and health innovation category ranged from 40 (chronic obstructive pulmonary disease monitoring) to 886 (diabetes self-management). The average per innovation category is comparable for consultation support (181), monitoring (162), and self-management (249), while it is much smaller for therapy (55). Evidence appears most advanced in the self-management category, with a total of 9 RCTs including more than half of the total number of patients.

Main Research Focuses and Study Results

Table 6 gives an overview of the main research focuses and study results for the 20 articles included in the effect review. *Study* is now equivalent to *article*, since none of these studies were reported in more than 1 article after exclusion by quality. Self-management appears to be the most promising health service innovation category, since the hypothesis was confirmed for 8 of the 9 studies.

In the monitoring category, 2 of the asthma studies confirmed their hypothesis, while we lack positive results for chronic obstructive pulmonary disease and cardiovascular monitoring. Also, the hypothesis was confirmed in 2 studies on consultation support in the cancer patient group.

Table 1. Judgments of methodological quality in the reviewed randomized controlled trials of electronic symptom reporting[a].

Article	Theoretical evidence/ preclinical testing	Random sequence generation[b]	Allocation concealment[b]	Blinding of participants and personnel[b,c]	Blinding of outcome assessment[b,d]	Incomplete outcome data[b]	Selective reporting[b,d]
Berger et al [59]	yes/yes	+	+	–	–	+	+
Bergström et al [60]	yes/unclear	+	+	–	+	+	+
Berry et al [34]	yes/yes	+	+	–	–	+	–
Boyes et al [35]	yes/no	○	○	–	–	–	–
Carrasco et al [52]	yes/no	+	–	–	–	+	+
Chan et al [42]	yes/no	+	○	–	–	○	–
Chan et al [43]	yes/yes	+	○	–	–	–	–
DeVito Dabbs et al [55]	yes/yes	+	–	–	–	+	+
Glasgow et al [63]	yes/yes	+	+	–	–	○	+
Guendelman et al [44]	yes/yes	○	+	–	–	+	+
Jan et al [45]	yes/no	○	+	–	–	+	–
Kearney et al [41]	yes/yes	+	+	–	–	–	–
Leveille et al [40]	yes/no	○	○	–	–	+	+
Lewis et al [49] (hospitalization)	yes/unclear	+	+	–	+	○	–
Lewis et al [50] (quality of life)	unclear/unclear	+	+	–	–	○	–
Nguyen et al [58]	yes/yes	+	+	–	–	–	+
Nguyen et al [51]	yes/no	+	+	–	–	+	–
Oerlemans et al [62]	yes/no	+	–	–	–	+	+
Prabhakaran et al [46]	no/no	+	+	–	–	+	–
Rasmussen et al [47]	yes/no	–	+	–	–	○	+
Ruland et al [36]	yes/yes	○	○	–	○	○	+
Ruland et al [37]	yes/yes	+	–	–	+	+	+
Santamore et al [53]	no/no	○	○	–	–	○	–
Schwarz et al [54]	yes/no	+	+	–	+	+	+
Stevens et al [39]	yes/no	+	○	–	–	–	–
van der Meer et al [57]	yes/unclear	+	+	–	–	+	+
Velikova et al [38]	yes/no	+	○	–	+	+	+
Vernmark et al [61]	yes/no	+	○	–	+	+	+
Wagner et al [65]	yes/no	+	○	–	–	+	+
Willems et al [48]	yes/no	+	–	–	–	+	+
Williams et al [64]	no/no	○	○	–	–	○	○
Yardley et al 2010 [56]	yes/no	+	+	+	+	–	○

[a] Articles were identified in a comprehensive search in Medline, EMBASE, PsycINFO, Cochrane Central Register of Controlled Trials, and IEEE Xplore from 1990 to November 2011, and were published in the time period 2002–2011.

[b] + = low risk, ○ = unclear risk, and – = high risk.

[c] Blinding of participants and personnel is extremely challenging in telemedicine and eHealth innovations. We thus did not consider this bias to be crucial to the quality judgment of the articles.

[d] We considered a low risk of bias for selective reporting and for blinding of outcome assessment to be the best indicators for identification of studies with high methodological quality.

Table 2. Effects reported in consultation support randomized controlled trials of acceptable quality of electronic symptom reporting, in alphabetic order of first author[a].

Article	Primary outcome	Intervention hypothesis relative to control	Number at randomization[b]			Within-group pre–post change[b]			Between-group pre–post change[b,c]		Hypothesis confirmed?
			C	I1	I2	C	I1	I2	C-I1	C-I2	
Leveille et al [40]	Number of patients discussing chronic condition with physician during consultation	Superior	120	121	NA[d]	–80.5%	–84.9%	NA	4.4% P = .37	NA	No
Ruland et al [36]	Congruence (weighted) between patients' preconsultation-reported health issues and issues discussed in consultation	Superior	25	27	NA	–12.8	–33	NA	20.2 P < .01	NA	Yes
Ruland et al [37]	Number of patients' symptoms and problems addressed by physicians as documented in inpatients' records	Superior	70	75	NA	–7.9	–10	NA	2.1 P < .001	NA	Yes
Velikova et al [38]	Health-related quality of life, functional assessment of cancer	Superior	72	144	70	NR[e]	NR	NR	–0.02 P = .3	NR	No

[a] Articles were identified in a comprehensive search in Medline, EMBASE, PsycINFO, Cochrane Central Register of Controlled Trials, and IEEE Xplore from 1990 to November 2011, and were published in the time period 2002–2011.

[b] C = control group, I1 = intervention group 1, I2 = intervention group 2.

[c] P values for difference between groups.

[d] Not applicable.

[e] Not reported.

Table 3. Effects reported in monitoring with clinical support randomized controlled trials of acceptable quality of electronic symptom reporting, in alphabetic order of first author[a].

Article	Primary outcome	Intervention hypothesis relative to control	Number at randomization[b]			Within-group pre–post change[b]			Between-group pre–post change[b,c]		Hypothesis confirmed?
			C	I1	I2	C	I1	I2	C-I1	C-I2	
Carrasco et al [52]	Number of patients exhibiting poor hypertension control	Superior	143	142	NA[d]	64.3%	68.3%	NA	–4.0% $P = .47$	NA	No
Guendelman et al [44]	Patients experiencing limitations in activity due to asthma last 14 days	Superior	68	66	NA	25%	35%	NA	–9% $P = .03$	NA	Yes
Lewis et al [49]	Number of hospitalizations	Superior	20	20	NA	–7	–4	NA	–3 $P = .16$	NA	No
Rasmussen et al [47]	FEV_1[e]	Superior	100	100	100	0.004	0.187	0.035	–0.183 $P < .001$	–0.031 Non-significant	I1: Yes
Schwarz et al [54]	Mean number of hospital readmissions in group	Superior	51	51	NA	–0.33	–0.32	NA	–0.01 $P = .9$	NA	No
Willems et al [48]	Asthma-specific quality of life	Superior	54	55	NA	–0.06	–0.29	NA	0.23 $P = .386$	NA	No

[a] Articles were identified in a comprehensive search in Medline, EMBASE, PsycINFO, Cochrane Central Register of Controlled Trials, and IEEE Xplore from 1990 to November 2011, and were published in the time period 2002–2011.

[b] C = control group, I1 = intervention group 1, I2 = intervention group 2.

[c] *P* values for difference between groups.

[d] Not applicable.

[e] Forced expiratory volume of air in the first second of expiration.

Table 4. Effects reported in self-management with clinical support randomized controlled trials of acceptable quality of electronic symptom reporting, in alphabetic order of first author[a].

Article	Primary outcome	Intervention hypothesis relative to control	Number at randomization[b]			Within-group pre–post change[b]			Between-group pre–post change[b,c]		Hypothesis confirmed?
			C	I1	I2	C	I1	I2	C-I1	C-I2	
Berger et al [59]	Social Phobia Scale	Equivalence	27	27	27	16.2	16.3	17.9	–0.1 *P* = .90	–1.7 *P* = .90	Yes
Bergström et al [60]	Panic Disorder Severity Scale	Equivalence	60	53	NA[d]	NR[e]	NR	NA	*P* = .95	NA	Yes
DeVito Dabbs et al [55]	Perceived self-care agency (in follow-up after surgery)	Superior	17	17	NA	–15.1	–19.44	NA	4.34 *P* = .003	NA	Yes
Glasgow et al [63]	Number of laboratory procedures completed in accordance with national diabetes guidelines	Superior	28[f]/417	24[f]/469	NA	–0.09	–0.37	NA	0.28 *P* = .001	NA	Yes
Nguyen et al [58] (dyspnea)	Patients' self-report of dyspnea with activities of daily living, Likert scale	Equivalence	24	26	NA	–4	–2.5	NA	–1.5 *P* = .51	NA	Yes
Oerlemans et al [62]	Cognitive Scale for Functional Bowel disorders	Superior	38	38	NA	1.84	8.99	NA	–7.15 *P* > .05	NA	No
van der Meer et al [57]	Asthma-related quality of life	Superior	99	101	NA	–0.18	–0.56	NA	0.38 *P* < .001	NA	Yes
Vernmark et al [61]	Symptom reduction: Beck Depression Inventory	Equivalence	29	29	30	5.2	9.9	11.9	–4.7 *P* = .06	–6.7 *P* = .002	Yes
Yardley et al [56][g]	Patient enablement scores, after 4 weeks.	Superior	346	368	NA	2	3	NA	–1 *P* = .03	NA	Yes

[a] Articles were identified in a comprehensive search in Medline, EMBASE, PsycINFO, Cochrane Central Register of Controlled Trials, and IEEE Xplore from 1990 to November 2011, and were published in the time period 2002–2011.

[b] C = control group, I1 = intervention group 1, I2 = intervention group 2.

[c] *P* values for difference between groups.

[d] Not applicable.

[e] Not reported.

[f] Cluster randomized by primary care physicians.

[g] 214 replied with regard to patient enablement, 95 in the intervention group, 119 in control.

Table 5. Effects reported in therapy randomized controlled trials of acceptable quality of electronic symptom reporting[a].

Article	Primary outcome	Intervention hypothesis relative to control	Number at randomization[b]			Within-group pre–post change[b]			Between-group pre–post change[b,c]		Hypothesis confirmed?
			C	I1	I2	C	I1	I2	C-I1	C-I2	
Wagner et al [65]	Intrusion measured by the Impact of Event Scale	Superior	26	29	NA[d]	3.77	11.5	NA	–7.73 *P* < .1	NA	Yes[e]

[a] Articles were identified in a comprehensive search in Medline, EMBASE, PsycINFO, Cochrane Central Register of Controlled Trials, and IEEE Xplore from 1990 to November 2011, and were published in the time period 2002–2011.

[b] C = control group, I1 = intervention group 1, I2 = intervention group 2.

[c] *P* values for difference between groups.

[d] Not applicable.

[e] Authors considered *P* < 0.1 to be statistically significant.

Table 6. Main research focus and overview of confirmed (+) and not confirmed (–) hypothesis for articles included in effect review of randomized controlled trials of electronic symptom reporting, by health service innovation category and patient group[a].

Patient group	Consultation support	Monitoring with clinical support	Self-management with clinical support	Therapy
Cancer	3 articles (2 excluded): More symptoms identified and discussed [36,37]++, [38]-	0 articles (1 excluded)		
Respiratory and lung diseases: asthma		3 articles (4 excluded): Improved asthma outcome (symptoms or quality of life): Children: [44]+ Adults: [47]+ Both: [48]-	1 article: Improved asthma-related quality of life [57]+	
Respiratory and lung diseases: COPD[b]		1 article (2 excluded): Reduced health care use [49]-	1 article: Reduced dyspnea [58]+[c]	
Respiratory and lung diseases: other			2 articles: Maximized lung transplant-related health outcomes [55]+ Self-care in management of minor respiratory symptoms [56]+	
Cardiovascular disease		2 articles (1 excluded): Improved hypertension [52]- Reduced health care use and costs [54]-		
Psychiatry	0 articles (1 excluded)		4 articles: Symptom reduction for (1) social phobia [59]+[c], (2) depression [61]+[c] More effective cognitive behavior therapy for patients with (1) panic disorder [60]+[c], (2) irritable bowel syndrome [62]-	1 article: More effective cognitive behavioral therapy for bereaved people with complicated grief [65]+
Diabetes			1 article (1 excluded) Quality of care [63]+	
Mixed	1 article: More symptoms identified and discussed [40]-			
Number of articles excluded due to low quality	3	8	1	0

[a] Articles were identified in a comprehensive search in Medline, EMBASE, PsycINFO, Cochrane Central Register of Controlled Trials, and IEEE Xplore from 1990 to November 2011, and were published in the time period 2002–2011.

[b] Chronic obstructive pulmonary disease.

[c] Hypothesis was to demonstrate equivalence between intervention and control.

Effects in Consultation Support

The outcomes were categorized by IOM's quality domains [33]. In the consultation support category, all studies provided patient-centered care, ensuring that patient-reported symptoms guided the clinical decisions. Except for the study where nurses coached patients [40], symptom reporting was generally conducted while the patient was present at the clinic, and a summary of the reported symptoms was made available to the physician [36-38]. These summaries were found effective in identifying and prompting discussion of troublesome symptoms, which made it possible to focus the conversation on issues relevant to the patient's problems [36-38].

The electronic symptom reporting systems also showed positive outcomes for patient symptom distress [37], symptom management [37], and health-related quality of life [38]. Patients supported the gathering of symptom information by computerized survey [36,40] and spent a median of 9 minutes reporting [36].

Most clinicians found the summaries useful for identifying problems and providing communication [38], which reduced the need for symptom management support [37]. Benefits for the health care system were mainly that visit duration was similar with and without use of the summaries [38].

The one trial with patient coaching [40] did not show benefits for patients or health professionals regarding detection of symptoms and quality of life, but patients in the intervention group reported that they received significantly more advice about their health and referrals to specialist.

Effects in Monitoring With Clinical Support

Only 2 monitoring studies reported benefits for patients, while nearly no benefits for the health system and none for the health professionals were reported. The 2 studies identifying health benefits for the patient focused on asthma outcomes for children [44] and adults [47] respectively. Both studies included a strong self-management element. For the latter study, some side effects for the health care system and patient need to be resolved [47].

All the studies, except 1, that aimed to demonstrate reduced health care costs belong to the monitoring group. However, with one exception, no health care costs or health care system benefits were identified: there was no improvement in total number of home care services or informal social support [54], number of consultations [52], occurrence of emergency room visits [44], hospital or specialist team use [49], number of hospital admissions [44,52,54], or mean costs per patient [66]. However, primary care contacts were reduced for patients with chronic obstructive pulmonary disease [49].

Effects in Self-management

Articles on self-management support were of higher quality, allowing a larger proportion of studies to be assessed with respect to effects. All self-management interventions were found equally effective to or better than the control option, with only one exception [62]. Substantial benefits for patients, and partly also for health professionals and health care systems, have been documented in this area.

Patient health benefits were reported for follow-up after lung transplantation [55], improved asthma-related quality of life [57], and reduced dyspnea associated with chronic obstructive pulmonary disease [58]. Patient-centered aspects of diabetes care [63], improved level of enablement through Web-based decision support of minor symptoms [56], and satisfaction [56,58] were also documented. Patient health benefits were documented in the psychiatry category for Internet-delivered treatment for social phobia [59], panic disorder [60], major depression [61], and partly for irritable bowel syndrome, with respect to catastrophizing thoughts [62]. Patients also reported a high level of satisfaction [59].

For health professionals, resource utilization was reported. There was some reduction in the number of physician consultations, due to increased asthma control [57] and when patients used the Web-based decision support system providing tailored advice for minor respiratory symptoms [56]. Regarding the latter study, it is important to take into account that the control group used a webpage consisting of advice previously shown to be effective in reducing the number of consultations [56]. In addition, the Internet-delivered treatment of panic disorder used considerably less therapist time than the cognitive behavioral therapy group treatment [60]. However, therapist time for email therapy for major depression was almost 10 times longer than time for guided self-help [61].

At the health care system level, health care cost benefits were analyzed and reported for Internet treatment of panic disorder, which was nearly 4 times cheaper than group treatment [60].

Effects in Therapy

Patients receiving email therapy for complicated grief improved significantly relative to participants in the waiting list condition, and were quite satisfied with the treatment [65]. Only 20% missed face-to-face contact with a therapist, and 85% had positive attitudes toward being treated via the Internet instead of face-to-face [65].

Discussion

Results are discussed with respect to methodological quality of the included RCTs and the effects and benefits gained from electronic symptom reporting between patient and health care provider.

Principal Findings on Methodological Quality and Effects or Benefits

Overall, the research field appears to be characterized by a comparably large number of low-quality articles that have serious methodological drawbacks.

In total 25% of the articles had multiple primary outcomes, and 28% had not defined or remained unclear regarding the primary outcome.

We extracted effect data only from articles with acceptable quality, which represented 62.5% of all included articles. About half of the articles in the consultation support and monitoring categories were excluded due to low quality, whereas only 1 of the 10 articles on self-management had to be excluded (1 of 2 articles describing the same study). The study hypotheses were confirmed in 13 of the 20 remaining articles. The hypotheses were confirmed in all 4 equivalence studies.

Overall, articles on self-management support were of a higher quality, allowing a larger proportion of studies to be assessed with respect to effects. All the self-management interventions are equally effective to or better than the control option, with one exception [62]. Substantial benefits for patients, and partly also for health professionals and the health care system, have been documented in this area.

In the monitoring trials, health benefits were identified for asthmatic children [44] and adults [47]. Both of these interventions included self-management elements with computer-tailored feedback. Of the 6 monitoring studies, 5 also addressed health care costs, but with one small exception, no cost benefits were identified.

The cancer studies in consultation support are encouraging, since it was found to provide patient-centered care, ensuring that patient-reported symptoms guided the clinical decisions.

Interpretation of Results

According to our requirements, seven of eight quality criteria should be fulfilled for a study to be considered methodologically correct (accepting that patients and personnel are not blinded). Unfortunately, none of the included articles received positive scores on all criteria, and many articles met just few of them. This lack of adequate methodology negatively affects the overall quality of the RCTs, as pointed out in other studies and reviews [25-27,67,68].

A total of 9 studies had an unclear primary outcome description [2], and all of these studies were excluded from the review of intervention effects due to low quality. Therefore, it is obvious that the field would benefit from a better definition of primary outcome to raise study quality in general, and to avoid selective reporting in particular.

Of the 9 studies with unclear primary outcome and excluded due to low quality, 6 belong to the monitoring category. None of the self-management studies were excluded (only a secondary analysis article where the primary analysis article still is included), which may reflect that this area is more mature. Self-management has already proven to be quite efficient for many long-term diseases [69,70], including psychiatric conditions [71], and various Internet-based setups for self-management have already been used and evaluated for many years.

The heterogeneity in intervention and research targets limits the possibilities to draw reliable conclusions with respect to the effects. Furthermore, designing, conducting, and reporting high-quality RCTs in this field in general is a great challenge, as they have to deal with complex interventions. The complex interventions include several components acting both independently and interdependently [23], and are thus difficult to analyze. If the result is negative, it is hard to judge whether this is because the trial was inadequately developed or applied, or applied in an inappropriate context, or used an inappropriate study design, especially regarding control groups and outcomes [23]. On the other hand, if the result is positive, there is no guarantee that the results can be generalized to a different context [23], not even within the same patient group. For example, in the study of Rasmussen et al where the “study showed that its use resulted in closer monitoring, immediate feedback, adequate medication, and better compliance and that all these initiatives together produced better asthma control” [47], it might be a challenge to judge how the different components affected each other, and how one can repeat the study expecting to obtain the same positive effect. This implies that even implementing an innovation that has been demonstrated to provide a positive effect requires attention and examination of the effects, rate of uptake, intervention stability, and so on [24]. To improve the uptake and impact of technologies in medical care, a holistic framework based on existing eHealth frameworks has recently been suggested [72]. This approach is aware of the existing interdependencies between technology, human characteristics, and the socioeconomic environment, and may be useful for innovating health care, also in future implementations [72]. Another possible next step in the quality assessment process of evaluating possible health service innovations is to use the Model for Assessment of Telemedicine as a guide. This model assists decision makers, before bringing services into everyday use, in predicting medical, social, economic, and ethical issues related to use of the service [73].

Limitations and Strengths of the Review

An important strength of this review is that the methodological quality assessment was based on Cochrane’s recommended domain-based evaluation for assessment of risk of bias [31], in addition to including three not commonly used quality assessment variables. These variables focus on the feasibility [31] and the theoretical basis for evaluation of complex interventions such as theoretical evidence and preclinical testing recommended by the Framework for Design and Evaluation of Complex Interventions to Improve Health [23,24]. Another important strength is that the risk of bias assessment was conducted by three independent researchers, spending several working days on consensus discussions. A third strength is that effect variables were extracted by two independent persons. A fourth strength is that we followed Cochrane’s recommendations for identifying the types of bias that are most important for the review. A fifth strength is that we took into account the Cochrane warning not to present effects for all studies, and in this way took seriously the flaws identified during the methodology assessment.

However, when conducting the bias evaluation, we could have split the bias regarding blinding of participants and personnel in two—that is, considering study participants and personnel separately. The reason for this is that some of the trials blinded patients to which intervention they received, but did not blind the involved health care personnel.

Some of the assessed articles had sources of bias outside those specified by the Cochrane Collaboration. An example is multiple end-point criteria and hypotheses without adequate adjustment for multiple statistical tests, which may cause problems with final interpretation of results [74]. Another example is unclear statistical analyses of cluster randomized trials. However, we felt that the specified Cochrane bias criteria were sufficiently detailed to discriminate between high- and low-quality articles, which is why we did not systematically extract and present information about other sources of bias.

Interpretation of evidence depends on many factors and is rarely straightforward [75]. Therefore, the interpretation also depends on the reviewers’ experience and background. However, as we always had two or three authors performing the reviewing subtasks independently, with consensus discussions for resolving disagreements, the results of the review should vouch for a high degree of validity and reliability.

Where 8 of the articles used more than one primary outcome, we decided to extract the first variable presented in the text to use in our effect measurement. All 8 had sufficient methodological quality, and 4 demonstrated a positive effect, while the other 4 did not. However, 3 of the 4 that did not demonstrate a positive effect: the self-management study by Oerlemans et al [62] and the consultation support studies by Leveille et al [40] and Velikova et al [38] included other primary outcomes that were positive and significant. We decided to extract the first variable because it is difficult to evaluate a mix of several outcomes; nevertheless, this may be considered a limitation.

Despite using a very comprehensive search strategy, we have reason to believe that we did not quite succeed in covering the area of psychiatry adequately (see part 1 of the review for a more detailed discussion [2]). If the 6 psychiatry studies had not been included, the self-management category would have been reduced, and been less convincing, and the focus for self-management would mainly have been on respiratory and lung diseases. On the other hand, if we had conducted a search that covered the psychiatry field better, we hypothesize, based on reading the studies from the reference lists, that more studies would have been added to both the self-management and the therapy groups. However, the overall quality and effects of electronic symptom reporting within the field of psychiatry are unclear. Therefore, this field deserves its own future review.

As a result of the heterogeneous outcome data in the studies, a meta-analysis was not possible.

Future Research

Studies of low quality are typically associated with an overestimation of benefits [26-28]. Consequently, improving the methodological quality in the field is essential, and future reviews are necessary to identify whether the methodological quality is improving. This concerns preclinical testing, allocation concealment, incomplete outcome data, and especially risk of bias introduced by selective reporting and nonblinding of outcome assessments being most relevant for this review. Even the blinding of patients and personnel might be achieved in specific studies.

In addition to better definition of the primary outcome to avoid selective reporting, we also recommend improving the account of how the proposed intervention should work, how the intervention links to the outcome measures, and the use of intention-to-treat analysis. As we rarely found that principles such as patient empowerment and patient-centered care were appropriately taken into consideration, this is also a recommendation for future studies.

We also encourage researchers to carefully consider whether it is necessary to demonstrate that the intervention is superior, or if it is sufficient to demonstrate that it is equivalent, as some of the studies designed as superiority studies would have had positive results if they had been designed as equivalence studies. Examples of such studies are the monitoring study of Willems at al [48] and the self-management study of Oerlemans et al [62].

More than half of the monitoring, self-management, and therapy interventions lasted 4 months or less (see part 1), which might be too brief to achieve the intended effect for long-term conditions, especially for complex interventions where both patient and provider often need some time to get used to the technology. Some of the negative studies may have had too short a time frame for an effect to materialize. For example, in Leveille et al’s study of nurse coaching, 38% of intervention participants had less than 2 weeks between completing the screening survey and their indexed appointment, so a longer intervention period might have led to better outcomes [40]. Appreciating the current evidence, we recommend running pilot trials to determine the time frame for effects to appear, and then designing interventions of the necessary length to better document the effects within the electronic symptom reporting field.

The self-management support trials were very successful and showed the most promising results, and should thus be an important guide for further research.

In the consultation support category, two related questions need to be investigated in future studies: (1) does completion of questionnaires, simply giving patients the opportunity to express how they feel, have a positive effect on patient well-being, regardless of whether the results are fed back to physicians? [35,38], and (2) does completion of assessment schemes prior to consultation result in patients recalling their answers and bringing up more symptoms or problems in the consultations [37,39], even if the physician does not read the reported symptoms? A yes to the second question could perhaps explain why the recognition rates in the group where the physician received the summary after the consultation were higher than recognition rates in usual-care samples [39].

The therapy category comprises innovations where the whole treatment, and all communication between therapists and patients, is conducted exclusively electronically. Unfortunately, we identified only 1 therapy article, so we cannot say anything about the general effects without conducting a new search, using specific psychiatry and Internet therapy-related terminology, as suggested in part 1 of the review [2].

An important contribution to the field would be to identify theoretical models that link the health service innovations, and their various components, with expected effects for patients, health professionals, and the health care system in a way that may support the design of the next generation of studies.

Implications for Practice

The number of studies within each combination of patient group and health service innovation is too small to draw final conclusions. However, if we look at electronic symptom reporting with regard to health service innovations, consultation support and self-management seem to bear various potential benefits for all stakeholders, at the patient, health care provider, and health care system levels.

Symptom reporting in cancer consultation support seems to require little staff effort to empower the patient, and there is little reason to doubt the accuracy of real-time reported symptom data when compared with average or retrospective ratings [6,76]. Reporting data has proven effective in identifying symptoms and prompting discussion of troublesome symptoms, and it allows for focusing the clinic visit conversation on issues relevant to the patient's problems [36-38]. Electronically reported data makes it possible to easily create clinical databases from which symptom and quality-of-life data can be retrieved, processed, and used in future consultations, surveillance, or for other epidemiological purposes.

Electronic symptom reporting for consultation support should also hold potential for other health conditions—for example, in raising sensitive issues the patient might find difficult to disclose in a face-to-face setting, such as stigma associated with sexually transmitted diseases or mental health problems [8]. However, in the future we expect more generic (not diagnosis-specific) symptom reporting to support consultations to guide both the patient and the clinician. When the patient conducts the reporting, data might be automatically analyzed to provide diagnostic aid for patients, or links to further reading, to prepare patients and thus facilitate more active participation in the treatment [77]. The health care provider, on the other hand, might prepare for recommended examinations on the basis of patient symptoms [78] and suggest solutions to problems and possible diagnoses based on comparable cases [79-81]. The patient and the clinician might then use the information to make a shared decision [82], with improved quality in terms of knowledge and values [83]. Positive effects for the health system seem possible as well—for example, through better information flow, which may avoid unnecessary allocation of resources, and through substitution of face-to-face consultations [10,11].

The positive effects on patients' self-management should encourage health care providers to promote future services based on the best practice of these innovations. Some of the self-management studies also point to improved cost effectiveness, as shown by Bergström et al with their Internet treatment for panic disorder, which was nearly 4 times cheaper than the group treatment [60]. Vernmark et al did not formally analyze cost effectiveness, but they reported guided self-help as the clinically most feasible option to implement, whereas individualized email therapy turned out to be more costly. However, both treatment effect sizes are in the range of what can be expected from face-to-face treatments [61], so cost-effective solutions might easily be offered to patients over the Internet, providing a more equitable service regardless of geographic location.

This review's positive result regarding self-management studies should be seen in light of the comprehensive Cochrane review on interactive health communication applications for people with chronic diseases [84]. These applications combine health information with either social support, decision support, or behavior support. The Cochrane review showed that these applications have a significant positive effect on knowledge, clinical outcomes, continuous behavioral outcomes, and the patient's feeling of being better socially supported [84].

Conclusion

Even in the subgroup of RCTs, the research methods in the included trials are of low quality. The field would benefit from an increased focus on methods of conducting and reporting RCTs. It appears particularly important to improve blinding of outcome assessment and to more precisely define the primary outcomes to avoid selective reporting.

Electronic communication between patients and health care providers is an exciting area of development for innovative health services, in line with current policies strengthening patient-centered service delivery models and information and communication technologies to increase efficiency and quality. Supporting self-management seems to be especially promising, but results from consultation support trials are also encouraging.

Acknowledgments

This study was partly funded by grants from the Health North Authorities (TFP-661-07), Norwegian Research Council grants (1749349), and The Norwegian Centre for Integrated Care and Telemedicine (NST). We want to express our thanks to colleagues at NST for useful advice, support, and feedback.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Number of patients and articles, and the average number of patients per article, in each combination of patient group and health innovation category.

[PDF File (Adobe PDF File), 92KB - jmir_v14i5e126_app1.pdf]

References

1. Johansen MA, Henriksen E, Berntsen G, Horsch A. Electronic symptom reporting by patients: a literature review. Stud Health Technol Inform 2011;169:13-17. [Medline: 21893705]

2. Johansen MA, Henriksen E, Horsch A, Schuster T, Berntsen GKR. Electronic Symptom Reporting Between Patient and Provider for Improved Health Care Service Quality: A Systematic Review of Randomized Controlled Trials. Part 1: State of the Art. J Med Internet Res 2012 Oct 03;14(5):e118 [FREE Full text] [doi: 10.2196/jmir.2214]
3. KITH. Norwegian Centre for Informatics in Health and Social Care in cooperation with Norsk selskap for allemennmedisin. 2004. ICPC-2 Kodekort [International Classification of Primary Care, Norwegian version]. Translated and adapted to Norwegian conditions by KITH URL: http://www.kith.no/upload/1895/ICPC-2-kodekort_fastsatt-04.2004.pdf [accessed 2012-05-29] [WebCite Cache ID 6815fPfhJ]
4. Haugli L, Finset A. [Physician-patient relations in functional disorders]. Tidsskr Nor Laegeforen 2002 Apr 30;122(11):1123-1125 [FREE Full text] [Medline: 12043058]
5. Roter DL, Hall JA. Physician's interviewing styles and medical information obtained from patients. J Gen Intern Med 1987;2(5):325-329. [Medline: 3655958]
6. Broderick JE, Schwartz JE, Vikingstad G, Pribbernow M, Grossman S, Stone AA. The accuracy of pain and fatigue items across different reporting periods. Pain 2008 Sep 30;139(1):146-157. [doi: 10.1016/j.pain.2008.03.024] [Medline: 18455312]
7. Brody EM, Kleban MH. Physical and mental health symptoms of older people: who do they tell? J Am Geriatr Soc 1981 Oct;29(10):442-449. [Medline: 7276408]
8. Johnsen J-AK, Gammon D. Connecting with ourselves and others online: psychological aspects of online health communication. In: Wilson EV , editor. Patient-Centered E-Health. New York, NY: Medical Information Science Reference; 2009:26-46.
9. Bergmo TS, Kummervold PE, Gammon D, Dahl LB. Electronic patient-provider communication: will it offset office visits and telephone consultations in primary care? Int J Med Inform 2005 Sep;74(9):705-710. [doi: 10.1016/j.ijmedinf.2005.06.002] [Medline: 16095961]
10. Adamson SC, Bachman JW. Pilot study of providing online care in a primary care setting. Mayo Clin Proc 2010 Aug;85(8):704-710. [doi: 10.4065/mcp.2010.0145] [Medline: 20516427]
11. Kummervold PE, Trondsen M, Andreassen H, Gammon D, Hjortdahl P. [Patient-physician interaction over the internet]. Tidsskr Nor Laegeforen 2004 Oct 21;124(20):2633-2636 [FREE Full text] [Medline: 15534640]
12. Aguirre-Córdova JF, Chávez-Vázquez G, Huitrón-Aguilar GA, Cortés-Jiménez N. [Why is surgery cancelled? causes, implications, and bibliographic antecedents]. Gac Med Mex 2003;139(6):545-551. [Medline: 14723050]
13. Trentman TL, Mueller JT, Fassett SL, Dormer CL, Weinmeister KP. Day of surgery cancellations in a tertiary care hospital: a one year review. J Anesth Clin Res 2010;1(3):109. [doi: 10.4172/2155-6148.1000109]
14. Yoon SZ, Lee SI, Lee HW, Lim HJ, Yoon SM, Chang SH. The effect of increasing operating room capacity on day-of-surgery cancellation. Anaesth Intensive Care 2009 Mar;37(2):261-266. [Medline: 19400490]
15. González-Arévalo A, Gómez-Arnau JI, delaCruz FJ, Marzal JM, Ramírez S, Corral EM, et al. Causes for cancellation of elective surgical procedures in a Spanish general hospital. Anaesthesia 2009 May;64(5):487-493. [doi: 10.1111/j.1365-2044.2008.05852.x] [Medline: 19413817]
16. Schofield WN, Rubin GL, Piza M, Lai YY, Sindhusake D, Fearnside MR, et al. Cancellation of operations on the day of intended surgery at a major Australian referral hospital. Med J Aust 2005 Jun 20;182(12):612-615. [Medline: 15963016]
17. Porter SC, Kohane IS, Goldmann DA. Parents as partners in obtaining the medication history. J Am Med Inform Assoc 2005;12(3):299-305 [FREE Full text] [doi: 10.1197/jamia.M1713] [Medline: 15684127]
18. Sciamanna CN, Diaz J, Myne P. Patient attitudes toward using computers to improve health services delivery. BMC Health Serv Res 2002 Sep 11;2(1):19 [FREE Full text] [Medline: 12225617]
19. Benoit A, Dykes P, Chang F, Gertman P, Vandever W, Li Q, et al. Using electronic questionnaires to collect patient reported history. AMIA Annu Symp Proc 2007:871. [Medline: 18693972]
20. Johansen MA, Berntsen G, Shrestha N, Bellika JG, Johnsen JA. An exploratory study of patient attitudes towards symptom reporting in a primary care setting. Benefits for medical consultation and syndromic surveillance? Methods Inf Med 2011;50(5):479-486. [doi: 10.3414/ME11-02-0005] [Medline: 21897995]
21. Buzaglo JS, Millard JL, Ridgway CG, Ross EA, Antaramian SP, Miller SM, et al. An Internet method to assess cancer patient information needs and enhance doctor-patient communication: a pilot study. J Cancer Educ 2007;22(4):233-240. [doi: 10.1080/08858190701634870] [Medline: 18067435]
22. Wald JS, Businger A, Gandhi TK, Grant RW, Poon EG, Schnipper JL, et al. Implementing practice-linked pre-visit electronic journals in primary care: patient and physician use and satisfaction. J Am Med Inform Assoc 2010;17(5):502-506 [FREE Full text] [doi: 10.1136/jamia.2009.001362] [Medline: 20819852]
23. Campbell NC, Murray E, Darbyshire J, Emery J, Farmer A, Griffiths F, et al. Designing and evaluating complex interventions to improve health care. BMJ 2007 Mar 3;334(7591):455-459. [doi: 10.1136/bmj.39108.379965.BE] [Medline: 17332585]
24. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. BMJ 2000 Sep 16;321(7262):694-696. [Medline: 10987780]
25. Augestad KM, Berntsen G, Lassen K, Bellika JG, Wootton R, Lindsetmo RO, Study Group of Research Quality in Medical Informatics and Decision Support (SQUID). Standards for reporting randomized controlled trials in medical informatics: a systematic review of CONSORT adherence in RCTs on clinical decision support. J Am Med Inform Assoc 2012;19(1):13-21 [FREE Full text] [doi: 10.1136/amiajnl-2011-000411] [Medline: 21803926]

26. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998 Aug 22;352(9128):609-613. [doi: 10.1016/S0140-6736(98)01085-X] [Medline: 9746022]
27. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995 Feb 1;273(5):408-412. [Medline: 7823387]
28. Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy. I: Medical. Stat Med 1989 Apr;8(4):441-454. [Medline: 2727468]
29. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P, CONSORT Group. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. Ann Intern Med 2008 Feb 19;148(4):295-309. [Medline: 18283207]
30. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. Open Med 2009;3(3):e123-e130. [Medline: 21603045]
31. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0 [updated March 2011]. 2011 Mar. URL: http://www.cochrane-handbook.org/ [accessed 2012-07-23] [WebCite Cache ID 69Nal1bMs]
32. Johansen MA. The Norwegian Centre for integrated Care and Telemedicine. 2012. List of Variables URL: http://telemed.custompublish.com/getfile.php/1936271.357.dvtfsvvepq/Variabelliste.pdf [accessed 2012-05-08] [WebCite Cache ID 67VZUbsga]
33. Committee on Quality of Health Care in America, Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press; 2001.
34. Berry DL, Blumenstein BA, Halpenny B, Wolpin S, Fann JR, Austin-Seymour M, et al. Enhancing patient-provider communication with the electronic self-report assessment for cancer: a randomized trial. J Clin Oncol 2011 Mar 10;29(8):1029-1035. [doi: 10.1200/JCO.2010.30.3909] [Medline: 21282548]
35. Boyes A, Newell S, Girgis A, McElduff P, Sanson-Fisher R. Does routine assessment and real-time feedback improve cancer patients' psychosocial well-being? Eur J Cancer Care (Engl) 2006 May;15(2):163-171. [doi: 10.1111/j.1365-2354.2005.00633.x] [Medline: 16643264]
36. Ruland CM, White T, Stevens M, Fanciullo G, Khilani SM. Effects of a computerized system to support shared decision making in symptom management of cancer patients: preliminary results. J Am Med Inform Assoc 2003;10(6):573-579 [FREE Full text] [doi: 10.1197/jamia.M1365] [Medline: 12925545]
37. Ruland CM, Holte HH, Røislien J, Heaven C, Hamilton GA, Kristiansen J, et al. Effects of a computer-supported interactive tailored patient assessment tool on patient care, symptom distress, and patients' need for symptom management support: a randomized clinical trial. J Am Med Inform Assoc 2010;17(4):403-410 [FREE Full text] [doi: 10.1136/jamia.2010.005660] [Medline: 20595307]
38. Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. J Clin Oncol 2004 Feb 15;22(4):714-724. [doi: 10.1200/JCO.2004.06.078] [Medline: 14966096]
39. Stevens J, Kelleher KJ, Gardner W, Chisolm D, McGeehan J, Pajer K, et al. Trial of computerized screening for adolescent behavioral concerns. Pediatrics 2008 Jun;121(6):1099-1105 [FREE Full text] [doi: 10.1542/peds.2007-1878] [Medline: 18519478]
40. Leveille SG, Huang A, Tsai SB, Allen M, Weingart SN, Iezzoni LI. Health coaching via an internet portal for primary care patients with chronic conditions: a randomized controlled trial. Med Care 2009 Jan;47(1):41-47. [doi: 10.1097/MLR.0b013e3181844dd0] [Medline: 19106729]
41. Kearney N, McCann L, Norrie J, Taylor L, Gray P, McGee-Lennon M, et al. Evaluation of a mobile phone-based, advanced symptom management system (ASyMS) in the management of chemotherapy-related toxicity. Support Care Cancer 2009 Apr;17(4):437-444. [doi: 10.1007/s00520-008-0515-0] [Medline: 18953579]
42. Chan DS, Callahan CW, Sheets SJ, Moreno CN, Malone FJ. An Internet-based store-and-forward video home telehealth system for improving asthma outcomes in children. Am J Health Syst Pharm 2003 Oct 1;60(19):1976-1981. [Medline: 14531243]
43. Chan DS, Callahan CW, Hatch-Pigott VB, Lawless A, Proffitt HL, Manning NE, et al. Internet-based home monitoring and education of children with asthma is comparable to ideal office-based care: results of a 1-year asthma in-home monitoring trial. Pediatrics 2007 Mar;119(3):569-578 [FREE Full text] [doi: 10.1542/peds.2006-1884] [Medline: 17332210]
44. Guendelman S, Meade K, Benson M, Chen YQ, Samuels S. Improving asthma outcomes and self-management behaviors of inner-city children: a randomized trial of the Health Buddy interactive device and an asthma diary. Arch Pediatr Adolesc Med 2002 Feb;156(2):114-120. [Medline: 11814370]
45. Jan RL, Wang JY, Huang MC, Tseng SM, Su HJ, Liu LF. An internet-based interactive telemonitoring system for improving childhood asthma outcomes in Taiwan. Telemed J E Health 2007 Jun;13(3):257-268. [doi: 10.1089/tmj.2006.0053] [Medline: 17603828]
46. Prabhakaran L, Chee WY, Chua KC, Abisheganaden J, Wong WM. The use of text messaging to improve asthma control: a pilot study using the mobile phone short messaging service (SMS). J Telemed Telecare 2010 Jul;16(5):286-290. [doi: 10.1258/jtt.2010.090809] [Medline: 20576744]

47. Rasmussen LM, Phanareth K, Nolte H, Backer V. Internet-based monitoring of asthma: a long-term, randomized clinical study of 300 asthmatic subjects. J Allergy Clin Immunol 2005 Jun;115(6):1137-1142. [doi: 10.1016/j.jaci.2005.03.030] [Medline: 15940125]
48. Willems DC, Joore MA, Hendriks JJ, Nieman FH, Severens JL, Wouters EF. The effectiveness of nurse-led telemonitoring of asthma: results of a randomized controlled trial. J Eval Clin Pract 2008 Aug;14(4):600-609. [doi: 10.1111/j.1365-2753.2007.00936.x] [Medline: 19126178]
49. Lewis KE, Annandale JA, Warm DL, Rees SE, Hurlin C, Blyth H, et al. Does home telemonitoring after pulmonary rehabilitation reduce healthcare use in optimized COPD? A pilot randomized trial. COPD 2010 Feb;7(1):44-50. [doi: 10.3109/15412550903499555] [Medline: 20214462]
50. Lewis KE, Annandale JA, Warm DL, Hurlin C, Lewis MJ, Lewis L. Home telemonitoring and quality of life in stable, optimised chronic obstructive pulmonary disease. J Telemed Telecare 2010 Jul;16(5):253-259. [doi: 10.1258/jtt.2009.090907] [Medline: 20483881]
51. Nguyen HQ, Gill DP, Wolpin S, Steele BG, Benditt JO. Pilot study of a cell phone-based exercise persistence intervention post-rehabilitation for COPD. Int J Chron Obstruct Pulmon Dis 2009;4:301-313 [FREE Full text] [Medline: 19750190]
52. Carrasco MP, Salvador CH, Sagredo PG, Márquez-Montes J, González de Mingo MA, Fragua JA, et al. Impact of patient-general practitioner short-messages-based interaction on the control of hypertension in a follow-up service for low-to-medium risk hypertensive patients: a randomized controlled trial. IEEE Trans Inf Technol Biomed 2008 Nov;12(6):780-791. [doi: 10.1109/TITB.2008.926429] [Medline: 19000959]
53. Santamore WP, Homko CJ, Kashem A, McConnell TR, Bove AA. Using a telemedicine system to decrease cardiovascular disease risk in an underserved population: design, use, and interim results. Conf Proc IEEE Eng Med Biol Soc 2007;2007:3701-3704. [doi: 10.1109/IEMBS.2007.4353135] [Medline: 18002801]
54. Schwarz KA, Mion LC, Hudock D, Litman G. Telemonitoring of heart failure patients and their caregivers: a pilot randomized controlled trial. Prog Cardiovasc Nurs 2008;23(1):18-26. [Medline: 18326990]
55. DeVito Dabbs A, Dew MA, Myers B, Begey A, Hawkins R, Ren D, et al. Evaluation of a hand-held, computer-based intervention to promote early self-care behaviors after lung transplant. Clin Transplant 2009;23(4):537-545. [doi: 10.1111/j.1399-0012.2009.00992.x] [Medline: 19473201]
56. Yardley L, Joseph J, Michie S, Weal M, Wills G, Little P. Evaluation of a Web-based intervention providing tailored advice for self-management of minor respiratory symptoms: exploratory randomized controlled trial. J Med Internet Res 2010;12(4):e66 [FREE Full text] [doi: 10.2196/jmir.1599] [Medline: 21159599]
57. van der Meer V, Bakker MJ, van den Hout WB, Rabe KF, Sterk PJ, Kievit J, SMASHING (Self-Management in Asthma Supported by Hospitals‚ ICT‚ Nurses and General Practitioners) Study Group. Internet-based self-management plus education compared with usual care in asthma: a randomized trial. Ann Intern Med 2009 Jul 21;151(2):110-120. [Medline: 19620163]
58. Nguyen HQ, Donesky-Cuenco D, Wolpin S, Reinke LF, Benditt JO, Paul SM, et al. Randomized controlled trial of an internet-based versus face-to-face dyspnea self-management program for patients with chronic obstructive pulmonary disease: pilot study. J Med Internet Res 2008;10(2):e9 [FREE Full text] [doi: 10.2196/jmir.990] [Medline: 18417444]
59. Berger T, Caspar F, Richardson R, Kneubühler B, Sutter D, Andersson G. Internet-based treatment of social phobia: a randomized controlled trial comparing unguided with two types of guided self-help. Behav Res Ther 2011 Mar;49(3):158-169. [doi: 10.1016/j.brat.2010.12.007] [Medline: 21255767]
60. Bergström J, Andersson G, Ljótsson B, Rück C, Andréewitch S, Karlsson A, et al. Internet-versus group-administered cognitive behaviour therapy for panic disorder in a psychiatric setting: a randomised trial. BMC Psychiatry 2010;10:54 [FREE Full text] [doi: 10.1186/1471-244X-10-54] [Medline: 20598127]
61. Vernmark K, Lenndin J, Bjärehed J, Carlsson M, Karlsson J, Oberg J, et al. Internet administered guided self-help versus individualized e-mail therapy: A randomized trial of two versions of CBT for major depression. Behav Res Ther 2010 May;48(5):368-376. [doi: 10.1016/j.brat.2010.01.005] [Medline: 20152960]
62. Oerlemans S, van Cranenburgh O, Herremans PJ, Spreeuwenberg P, van Dulmen S. Intervening on cognitions and behavior in irritable bowel syndrome: A feasibility trial using PDAs. J Psychosom Res 2011 Mar;70(3):267-277. [doi: 10.1016/j.jpsychores.2010.09.018] [Medline: 21334498]
63. Glasgow RE, Nutting PA, King DK, Nelson CC, Cutter G, Gaglio B, et al. Randomized effectiveness trial of a computer-assisted intervention to improve diabetes care. Diabetes Care 2005 Jan;28(1):33-39. [Medline: 15616230]
64. Williams GC, Lynch M, Glasgow RE. Computer-assisted intervention improves patient-centered diabetes care by increasing autonomy support. Health Psychol 2007 Nov;26(6):728-734. [doi: 10.1037/0278-6133.26.6.728] [Medline: 18020845]
65. Wagner B, Knaevelsrud C, Maercker A. Internet-based cognitive-behavioral therapy for complicated grief: a randomized controlled trial. Death Stud 2006 Jun;30(5):429-453. [doi: 10.1080/07481180600614385] [Medline: 16610157]
66. Willems DC, Joore MA, Hendriks JJ, Wouters EF, Severens JL. Cost-effectiveness of a nurse-led telemonitoring intervention based on peak expiratory flow measurements in asthmatics: results of a randomised controlled trial. Cost Eff Resour Alloc 2007 Jul;5:10 [FREE Full text] [doi: 10.1186/1478-7547-5-10] [Medline: 17662113]

67. Schulz KF, Chalmers I, Altman DG, Grimes DA, Doré CJ. The methodologic quality of randomization as assessed from reports of trials in specialist and general medical journals. Online J Curr Clin Trials 1995 Aug 26;Doc No 197:[81 paragraphs]. [Medline: 8542098]
68. Lau PW, Lau EY, Wong del P, Ransdell L. A systematic review of information and communication technology-based interventions for promoting physical activity behavior change in children and adolescents. J Med Internet Res 2011;13(3):e48 [FREE Full text] [doi: 10.2196/jmir.1533] [Medline: 21749967]
69. 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 [FREE Full text] [Medline: 11816692]
70. Von Korff M, Gruman J, Schaefer J, Curry SJ, Wagner EH. Collaborative management of chronic illness. Ann Intern Med 1997 Dec 15;127(12):1097-1102. [Medline: 9412313]
71. Bilsker D, Goldner EM, Jones W. Health service patterns indicate potential benefit of supported self-management for depression in primary care. Can J Psychiatry 2007 Feb;52(2):86-95. [Medline: 17375863]
72. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
73. Kidholm K, Ekeland AG, Jensen LK, Rasmussen J, Pedersen CD, Bowes A, et al. A model for assessment of telemedicine applications: mast. Int J Technol Assess Health Care 2012 Jan;28(1):44-51. [doi: 10.1017/S0266462311000638] [Medline: 22617736]
74. Bender R, Lange S. Adjusting for multiple testing--when and how? J Clin Epidemiol 2001 Apr;54(4):343-349. [Medline: 11297884]
75. Lau J, Ioannidis JP, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. BMJ 2006 Sep 16;333(7568):597-600. [doi: 10.1136/bmj.333.7568.597] [Medline: 16974018]
76. Broderick JE, Schwartz JE, Schneider S, Stone AA. Can End-of-day reports replace momentary assessment of pain and fatigue? J Pain 2009 Mar;10(3):274-281. [doi: 10.1016/j.jpain.2008.09.003] [Medline: 19070550]
77. Jimison HB, Sher PP, Jimison JJB. Decision support for patients. In: Berner ES , editor. Clinical Decision Support Systems: Theory and Practice. New York, NY: Springer; 2007:249-261.
78. Tanaka Y, Shibata S, Ohtuka Y, Hattori M, Aoshima T, Tohyama S, et al. Subjective symptoms acquisition system in a health promotion system for the elderly. Committee of System Development, Council of Japan AMHTS Institutions. Methods Inf Med 2000 Aug;39(3):238-240. [Medline: 10992751]
79. Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. JAMA 1998 Oct 21;280(15):1339-1346. [Medline: 9794315]
80. Johnston ME, Langton KB, Haynes RB, Mathieu A. Effects of computer-based clinical decision support systems on clinician performance and patient outcome. A critical appraisal of research. Ann Intern Med 1994 Jan 15;120(2):135-142. [Medline: 8256973]
81. Elvidge K. Improving pain & symptom management for advanced cancer patients with a clinical decision support system. Stud Health Technol Inform 2008;136:169-174. [Medline: 18487726]
82. Drake RE, Deegan PE, Woltmann E, Haslett W, Drake T, Rapp CA. Comprehensive electronic decision support systems. Psychiatr Serv 2010 Jul;61(7):714-717. [doi: 10.1176/appi.ps.61.7.714] [Medline: 20592007]
83. Drake RE, Cimpean D, Torrey WC. Shared decision making in mental health: prospects for personalized medicine. Dialogues Clin Neurosci 2009;11(4):455-463. [Medline: 20135903]
84. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive Health Communication Applications for people with chronic disease. Cochrane Database Syst Rev 2005(4):CD004274. [doi: 10.1002/14651858.CD004274.pub4] [Medline: 16235356]

Abbreviations

ICPC: International Classification of Primary Care
IOM: Institute of Medicine
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses
RCT: Randomized Controlled Trial

Edited by G Eysenbach; submitted 09.06.12; peer-reviewed by E Ammenwerth, P Kastner, K Antypas; comments to author 27.06.12; revised version received 11.07.12; accepted 12.07.12; published 03.10.12.

Please cite as:
Johansen MA, Berntsen GKR, Schuster T, Henriksen E, Horsch A
Electronic Symptom Reporting Between Patient and Provider for Improved Health Care Service Quality: A Systematic Review of Randomized Controlled Trials. Part 2: Methodological Quality and Effects
J Med Internet Res 2012;14(5):e126
URL: http://www.jmir.org/2012/5/e126/
doi:10.2196/jmir.2216
PMID:23032363

Original Paper

What Young People Want From a Sexual Health Website: Design and Development of Sexunzipped

Ona McCarthy[1], BA(Hons), MSc; Kenneth Carswell[2], BA(Hons), Mphil, DClinPsy; Elizabeth Murray[1], FRCGP, PhD; Caroline Free[3], MBChB, MRCGP, PhD; Fiona Stevenson[1], MA, PhD; Julia V Bailey[1], BSc, MSc, PhD, MBBS, MRCGP

[1]e-Health Unit, Research Department of Primary Care & Population Health, University College London, London, United Kingdom

[2]Barts & The London School of Medicine and Dentistry, Queen Mary, University of London, London, United Kingdom

[3]Department of Nutrition and Public Health Intervention Research, London School of Hygiene & Tropical Medicine, London, United Kingdom

Corresponding Author:
Ona McCarthy, BA(Hons), MSc
e-Health Unit
Research Department of Primary Care & Population Health
University College London
Upper 3rd Floor
Royal Free Hospital, Rowland Hill Street
London, NW3 2PF
United Kingdom
Phone: 44 (0)20 7794 0500
Fax: 44 (0)20 7794 1224
Email: omccarthy@tccr.org.uk

Abstract

Background: Sexual health education in the United Kingdom is of variable quality, typically focusing on the biological aspects of sex rather than on communication, relationships, and sexual pleasure. The Internet offers a unique opportunity to provide sexual health education to young people, since they can be difficult to engage but frequently use the Internet as a health information resource.

Objectives: To explore through qualitative research young people's views on what elements of a sexual health website would be appealing and engaging, and their views on the content, design, and interactive features of the Sexunzipped intervention website.

Methods: We recruited 67 young people aged 16–22 years in London, UK. We held 21 focus groups and 6 one-to-one interviews to establish sexual health priorities, views on website look and feel, and what features of a sexual heath website would attract and engage them. Two researchers facilitated the focus groups, using a semistructured topic guide to lead the discussions and asking open questions to elicit a range of views. The discussions and interviews were audio recorded and detailed notes were made on key topics from the audio recording. Young people's views influenced design templates for the content and interactive features of Sexunzipped.

Results: Young people particularly wanted straightforward information on sexual pleasure, sexually transmitted infections and pregnancy, how to communicate with partners, how to develop skills in giving pleasure, and emotions involved in sex and relationships. Focus group participants wanted social interaction with other young people online and wanted to see themselves reflected in some way such as through images or videos.

Conclusions: While it is challenging to meet all of young people's technological and design requirements, consultation with the target audience is valuable and necessary in developing an online sexual health intervention. Young people are willing to talk about sensitive issues, enjoy the discussions, and can offer key insights that influence intervention development.

(J Med Internet Res 2012;14(5):e127) doi:10.2196/jmir.2116

KEYWORDS

Sex education; adolescents; young adults; qualitative research

Introduction

Sexual health information provided in UK schools is taught within the statutory requirements of the National Science Curriculum [1]. It is variable in quality and tends to focus on the biological facts about reproduction, condom use, and sexually transmitted infections [1]. Consequently, information on skills building, sexual communication, relationship decision making, and sexual pleasure are generally missing from young people's sexual learning experience. Including these elements in young people's sexual education would align with one definition of sexual health put forth by the World Health Organization:

> *...a state of physical, emotional, mental and social well-being in relation to sexuality; it is not merely the absence of disease, dysfunction or infirmity. Sexual health requires a positive and respectful approach to sexuality and sexual relationships, as well as the possibility of having pleasurable and safe sexual experiences, free of coercion, discrimination and violence. For sexual health to be attained and maintained, the sexual rights of all persons must be respected, protected and fulfilled. [2]*

Young people can be seen as a hard-to-reach group for health promotion. Internet interventions may be one way of reaching young people, as they frequently use the Internet to search for health information [3-5] and appreciate the privacy, convenience, and ease of accessing information on the Web [4,5]. These potential advantages are particularly valued for stigmatized or sensitive subjects such as sexual health. A further potential advantage of using the Internet for sexual health promotion is the relative ease of covering a broad range of information including sexual pleasure. Some argue that, since one of the main motivations for engaging in sex is pleasure, pleasure must be addressed in the discourse of sexual health promotion [6]. A systematic review of interactive computer-based interventions found that they can help increase sexual knowledge and have a positive effect on safer-sex self-efficacy, intention, and behavior [7].

To determine the feasibility of promoting sexual health on the Internet to young people, we first set out to determine young people's views about whether an Internet intervention could help meet their sexual health needs and, if so, how best this could be done. This paper reports on our findings and on how they informed the development of an Internet intervention for sexual health known as Sexunzipped [8]. Additional details on integrating psychological theory into the design of this online intervention have been described elsewhere [9].

Methods

Design and Setting

We conducted 21 focus groups in London, UK, at young people's community sexual health clinics and held 6 one-to-one interviews in the Research Department of Primary Care and Population Health, University College London. Ethical approval was obtained from the University College London Research Ethics Committee (ref number 1023/001).

Participants

Young people aged 16–22 years were eligible to participate in the study.

Recruitment

We used multiple methods of recruitment, including approaching young people in sexual health clinic waiting rooms, distributing flyers around clinics and colleges, approaching key contacts involved in youth work, and placing advertisements on youth websites such as the UK Youth Parliament and college webpages. Snowballing was also used, where initial participants were asked to recruit potentially interested friends.

Participants were initially asked to attend only 1 focus group but were invited to subsequent groups if they were particularly engaged. Repeated attendance at focus groups over time allowed these participants to build on previous discussions. They were able to comment on refinements made to earlier website content and interactive formats.

We asked 3 focus group participants who were particularly engaged to attend a one-to-one interview. We also recruited 3 interviewees from online advertisements on youth websites and through key youth work contacts.

All participants provided written consent to their participation before attending a focus group or interview and were offered a £10 (US $16) cash incentive if they stayed for the duration.

Data Collection

The focus group discussions were used to explore (1) young people's sexual health priorities, (2) their views on website look and feel, and (3) what features of an interactive sexual heath website would attract and engage them.

All focus groups were moderated by 2 facilitators who introduced the project, explained the purpose of the focus group, set ground rules of mutual respect and confidentiality, and ensured that all participants had provided written informed consent. Participants provided anonymous demographic information by completing a paper form. The discussions and interviews were audio recorded with participants' permission.

Semistructured topic guides were used to guide the focus group discussions. These guides evolved as data were collected. Open-ended questioning was used to elicit a wide range of views, and participants were asked to elaborate on relevant topics beyond the topic guide. Less-vocal participants were encouraged to share their views. We developed the Sexunzipped website in parallel with the emerging data from the focus groups, with early website design templates prompting discussion in subsequent focus groups. This enabled participants to comment on the presentation and content of the intervention, leading to iterative changes as a form of participatory design. Focus group discussions ranged from 60 to 90 minutes, depending on the quality of the information gathered and participants' attention spans.

One-to-one interviews lasted 60 minutes and used open-ended questioning. The interviews were used to gather views on early versions of website content and to generate anonymous quotations on sexual health issues to use throughout the website.

To test the website content, we asked participants to read and comment on sample content and to give feedback on paper versions of interactive activities. To generate the quotations, the interviewer asked participants for their views on potentially controversial sex-related topics, such as “What do you think about girls carrying condoms?”. Anonymous quotations for the website were also generated by adapting comments taken from youth website discussion boards.

Data Analysis

One researcher made detailed notes on key themes from the focus group audio recordings. The analytic notes were reviewed and discussed by the core research team consisting of a clinical psychologist, a health services researcher, and a sexual health clinician. Key findings were documented, and emergent themes were tested in subsequent focus groups until the core team was satisfied that the information gained was sufficient. Information from the one-to-one content testing and quotation generation was used to refine subsequent website content.

Results

Participant Characteristics

A total of 67 people aged 16–22 years living in the United Kingdom participated in 21 focus groups and 6 one-to-one interviews over 16 months (Table 1). We used 3 one-to-one interviews to test website content and 3 to generate anonymous quotations for use throughout the website.

Of the 67 participants, 48 (72%) were aged 16–17 years, and 50 (75%) were female. Participants were ethnically diverse, with 26 (39%) describing themselves as white British or white other, 26 (39%) as black British, 2 (3%) Asian British, 7 (10%) as mixed, and 6 (9%) as other. The median age at which they first had sex was 15 years for both young women and men, with 33 (66%) young women ever having had sex and 15 (88%) young men ever having had sex. Of the 67 participants, 57 (85%) were in education or training. Most participants lived in inner London, with 2 living outside but near London. Of the 67 participants, 20 (30%) young people participated more than once.

Table 1. Self-reported participant characteristics (n = 67).

Demographic characteristic	n	%	
Age (years)			
	16–17	48	72%
	18–22	19	28%
Gender			
	Female	50	75%
	Male	17	25%
Ever had sex			
	Yes	48	72%
	No	19	28%
Ethnic group			
	White British/white other	26	39%
	Black British	26	39%
	Asian British	2	3%
	Mixed	7	10%
	Other	6	9%
Education/work			
	In education or training	57	85%
	Working	3	4%
	Other	7	10%

Young People’s Views on the Creation of a Sexual Health Website

Focus group participants appreciated being consulted and were eager to share their views. They supported the creation of a sexual health website with input from young people because they felt that current sites did not fully address their needs. They provided a wealth of information about their sexual health priorities and what features of such a website would be attractive and engaging to them.

Content

Participants wanted straightforward information on sexual pleasure, sexually transmitted infections and pregnancy, how to communicate with partners, how to develop skills in giving pleasure, and emotions involved in sex and relationships.

Young people wanted information on the pleasurable aspects of sex, presented in a nonjudgmental way (eg, it's acceptable to have casual sex as long as both people consent and are happy about it) while recognizing the importance of more serious, risk-focused sexual health topics. They were enthusiastic about including information on masturbation, and sexual positions and practices, and wanted general tips on how to be a better lover. They also wanted to hear others' experiences about giving and getting pleasure.

Participants wanted help with developing skills in communicating about sex and dealing with emotions within sexual relationships. They also wanted practical tips on how to communicate about sensitive issues such as negotiating condom use and talking about what kind of sex they do or do not like.

Participants thought a sexual health website should include information relevant for young people at different stages of sexual experience and exploration. They suggested it should cover information and activities for those who had not yet had sex, such as information on sex for the first time and advice on how to avoid unwanted events, as well as content for those who are looking to explore new sexual practices. Participants felt inclusion of more advanced sexual practices such as anal sex and fisting would not be intimidating or create feelings of inadequacy as long as content was balanced with other sexual health information. They speculated that advanced sexual content would attract young people to the website because such content is usually absent from youth-targeted resources, and if users were not interested in certain areas of the website, they simply would not read it.

Young people wanted website content presented in a straightforward style that sounds honest, accurate, and not "preachy." They recommended this style should not sound too serious and should not come across as "school work." They preferred content written in the voice of someone credible and knowledgeable but also someone they can relate to and respect such as an older sibling.

They preferred the use of uncomplicated words and phrases, for example, sex instead of sexual intercourse. For those situations where an uncomplicated word is unavailable, they suggested providing a definition, such as "transmit = to pass on to someone." They were also adamant that "youth-speak" or slang (eg, coz for because) should not be used, as the website would come across as trying too hard, although young people felt that it could sometimes be appropriate to use simpler versions of words such as lube for lubrication and come for ejaculate.

Young people did not want what they interpreted as clichés such as "I feel like I should be having sex because everyone else is" included on the site. They recognized an existence of double standards for men and women regarding sex (eg, it's acceptable for men to have numerous partners but not for women) and valued a space to address these issues.

Website Look and Feel

There was no clear consensus on website colors and logo design, with some preferring bold colors and others neutral tones. Participants wanted images of people, scenarios, and "real-life stuff" they could relate to such as young people partying in clubs, kissing, and hanging out on the street. They felt that images of young people should reflect the diversity of the UK population.

Participants wanted a website to include specific sexual health images such as photographs of sexually transmitted infections and contraception options. Young people expected the website logo to reflect sex and not require a lot of thought to understand its significance. They suggested the website name should be clear, be memorable, and suggest the content of the website. Young people wanted the site to feel active, with new activities and information added each week. They disliked pages containing a lot of text, which they said that they would not read.

Engaging Website Features

Focus group participants wanted social interaction with other young people such as through discussion boards. A post-and-comment format used on social networking websites such as Facebook was a popular suggestion, since it is familiar and enables users to interact. Young people said that all communication must be anonymous so as not to disclose their identity if "embarrassing things" were posted.

Young people wanted videos to feature on a sexual health website but specified that these videos must represent real people talking about real experiences. If this is impossible, high-quality actors must be used; if the people and situations in videos are not believable and realistic, the young people would not relate to them and would not find them valuable. One suggestion was to include short clips of young people sharing real stories on a range of issues with a means of commenting on the videos such as on the website YouTube.

Another feature that appealed to young people was a dramatic story format in which the user can choose different courses of action at pivotal points in the narrative. Participants liked the idea that dramas could have different endings depending on what decisions they made along the way and said they would be disappointed if they received the same feedback after different story endings.

Design and Content of the Sexunzipped Website

Paying particular attention to young people's desire for interactivity, we integrated young people's views with psychological theories of behavior change to create the Sexunzipped theory-based intervention website [9]. Reflecting young people's suggestions and priorities, the website content is divided into three major sections: Relationships, Safer Sex, and Sexual Pleasure. Each section contains a combination of interactive quizzes and decision-making activities, as well as text-based information. Links to other related topics on the site are provided at the end of all activities and text-based pages.

Information Pages

The text-based information pages throughout each of the three sections cover the major issues raised by the young people under the headings Relationships, Safer Sex, and Sexual Pleasure (Figure 1).

Interactive Quizzes

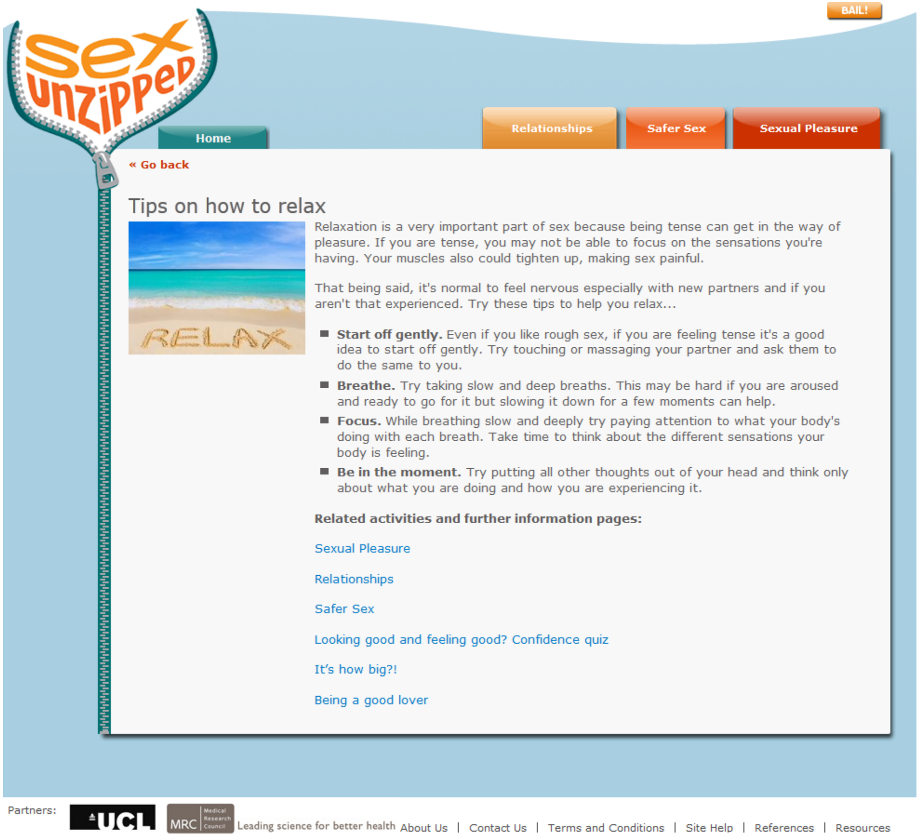
Drawing from behavior change theory [9], the quizzes on Sexunzipped were designed to prompt active learning and reflection by providing feedback according to responses given by users (Figure 2). Response options were designed to encourage users to reflect on their views, emotions, and experiences, as well as providing information on social norms, and encouraging beliefs and attitudes associated with safer-sex behavior. Incorporating what we learned from the focus groups, we designed and wrote the quizzes to be entertaining as well as educational, keeping in mind that young people said they valued an honest tone that does not “try too hard.”

Interactive Decision-Making Activities

The decision-making activities were designed to provoke self-reflection about sexual behavior, focusing on problematic situations or dilemmas concerning relationships, or risky or regretted sexual situations (Figure 3). Participants had strong feelings about the length of these activities, which they said must be short and not require a lot of writing. To address this, we created a format whereby users could drag and drop prespecified response options but could also add their own.

The website also includes quotations with aliases reflecting a range of views intending to give the feel of a peer-to-peer exchange of views (Figure 4).

Figure 1. Example information page on the Sexunzipped website.



Tips on how to relax

Relaxation is a very important part of sex because being tense can get in the way of pleasure. If you are tense, you may not be able to focus on the sensations you're having. Your muscles also could tighten up, making sex painful.

That being said, it's normal to feel nervous especially with new partners and if you aren't that experienced. Try these tips to help you relax...

- **Start off gently.** Even if you like rough sex, if you are feeling tense it's a good idea to start off gently. Try touching or massaging your partner and ask them to do the same to you.
- **Breathe.** Try taking slow and deep breaths. This may be hard if you are aroused and ready to go for it but slowing it down for a few moments can help.
- **Focus.** While breathing slow and deeply try paying attention to what your body's doing with each breath. Take time to think about the different sensations your body is feeling.
- **Be in the moment.** Try putting all other thoughts out of your head and think only about what you are doing and how you are experiencing it.

Related activities and further information pages:

Sexual Pleasure

Relationships

Safer Sex

Looking good and feeling good? Confidence quiz

It’s how big?!

Being a good lover

Figure 2. Interactive quiz: example feedback.

Sex myths

Masturbation is only for sex obsessed people

No, this is so very false.

Masturbation is done by all kinds of people not just those who are "sex obsessed".

People masturbate whether they're having frequent sex or not having sex at all.

It is very normal and is healthy to do every day...even many times a day.

A UK survey found that 45% of young people (aged 16-20) had masturbated in the last month with 66% having masturbated at some point in their life.

Figure 3. Decision-making activity: example question with drag-and-drop feature.

Not enjoying sex?

Think of a time when you've been sexual with someone and really enjoyed it (this can include kissing or any type of sexual practice). What was good about the person you were with?

(Click and drag your choices. To enter your own, click +. To enter more of your own, trash some options.)

Figure 4. Example quotations regarding sexually transmitted infection (STI).

STI myth busting

You can usually tell the type of person that may have an STI

Emily, 22 "No. Your instincts may tell you something but no, you can't tell for sure."

Beth, 16 "Maybe. It's mostly people who have sex a lot with just anyone that normally get STIs."

There is no way of telling by just looking at someone if they have an STI. A lot of people think they can tell by someone's reputation. The only way you can tell for sure is by getting tested.

Discussion

Principal Results

Participants were actively engaged in the focus group discussions and enjoyed taking part. They offered key insights into the content, look, and feel and interactive features of the Sexunzipped sexual health website. The resulting intervention has been tested in a pilot online trial (trial registration number ISRCTN55651027).

The strongest message to emerge from this research was the importance of providing young people with honest sexual health information that features content on sexual pleasure and relationships. Participants desired a website that reflects a mature, trustworthy, and true-to-life feel, conveyed through the voice of a knowledgeable youth worker or older sibling. The need for the website to appear trustworthy [10,11] along with the potential to share and compare real-life experiences [12] is consistent with previous findings. Our focus group findings align with other research, which found that young people are interested in receiving sexual health information through digital technologies such as the Internet and mobile phones [11,13], and are interested in knowing more about sexual pleasure, feelings, and emotions [14].

The integration of pleasure and relationship information with traditional risk-focused sexual health promotion was seen as novel and attractive, with participants preferring this approach to their experiences of sexual health education delivered by schools or by health authorities. Giving equal weight to traditional risk-based messages and sexual pleasure (including masturbation, for example) would give the message that both are important for healthy sexual lives [15,16].

A wide variety of sexual health websites targeted at young people are available on the Internet, for example, those developed in consultation with young people [17], those including information on sexual pleasure [18,19] and those offering multimedia features such as videos of people talking about their experiences [20,21]. The Sexunzipped website combines these elements in that it was developed with young people, features sexual pleasure, and includes theory-based interactive activities.

Addressing What Young People Want in a Website

Perhaps the greatest challenge was to integrate young people's preferences with psychological theory [9] within technical and budgetary constraints to create an acceptable and engaging sexual health website.

Participants expressed their desire for direct social interaction with other young people online (eg, via discussion boards and instant messaging). These features are expensive, require moderation, and raise ethical issues, for example, regarding online bullying, or disclosure of illegal or dangerous activities. There was also a risk that if no one used the discussion board, the site would appear unpopular and therefore unattractive. Young people wanted to see themselves reflected on the site (eg, in videos and drama), but this is expensive to produce and is difficult to pitch correctly. Young people are accustomed to the actively changing landscape on the Internet and wanted new activities and information added on a regular basis, which is resource intensive.

Consultation with users is essential in the development of websites targeted at young people, as this group can be particularly influenced by look and feel. This research and previous research suggest that listening to and meeting young people's desires in terms of website design and content is essential in engaging them [10,12]. However, websites may quickly look and feel out-of-date due to the rapidly changing online environment, which necessitates regular user consultation and updating. This requires careful financial planning by those involved in site maintenance in order to support the continued user consultation and updating.

Limitations

Focus group participants were London-based young people, which could mean that any differences in sexual health priorities due to geographical location may be underrepresented in the content of Sexunzipped. However, we thought that young people in other areas of the United Kingdom would be able to recognize and relate to inner-city youth culture to some degree. The Sexunzipped website design was intended to appeal to a broad range of young people, avoiding any specific styles that could alienate particular groups or become quickly out-of-date.

The sample contained 75% (50/67) women because greater numbers of young women attend community sexual health clinics. This may have resulted in views more representative of female needs and priorities. We made attempts to include more young men such as contacting male youth organizations, but this group proved to be less accessible than young women.

Although most focus groups included participants who did not know each other, the snowballing method of recruitment meant that the sample could have resulted in participants having similar views.

Conclusions

Consultation with the target audience is valuable and necessary in developing sexual health interventions, particularly in developing interventions for young people. It can be difficult to pitch content correctly, but it is possible with repeated consultation with users. It is challenging to incorporate all of young people’s technological desires, but digital technology makes it possible to create an engaging online intervention that includes a variety of interactive formats covering a broad range of topics. Young people are willing to talk about sensitive issues, enjoy the discussions, and can offer key insights that influence intervention development.

Acknowledgments

The authors would like to thank all young people and clinic staff involved in the study for their help, in particular Brook Sexual Health Centres and Pulse. This study was funded by the UK Medical Research Council.

Conflicts of Interest

None declared.

References

1. UK Department for Education. 2010. Consultations (Sex and Relationships Education Guidance) URL: http://www.education.gov.uk/consultations/index.cfm?action=conResults&consultationId=1637&external=no&menu=3 [WebCite Cache ID 66AKH2npW]
2. World Health Organization. 2006. Defining Sexual Health: Report of a Technical Consultation on Sexual Health, 28-31 January 2002, Geneva URL: http://www.who.int/reproductivehealth/publications/sexual_health/defining_sexual_health.pdf [accessed 2012-03-14] [WebCite Cache ID 66AJLCqgH]
3. Gray NJ, Klein JD. Adolescents and the internet: health and sexuality information. Curr Opin Obstet Gynecol 2006 Oct;18(5):519-524. [doi: 10.1097/01.gco.0000242954.32867.76] [Medline: 16932046]
4. Kanuga M, Rosenfeld WD. Adolescent sexuality and the internet: the good, the bad, and the URL. J Pediatr Adolesc Gynecol 2004 Apr;17(2):117-124. [doi: 10.1016/j.jpag.2004.01.015] [Medline: 15050988]
5. Skinner H, Biscope S, Poland B, Goldberg E. How adolescents use technology for health information: implications for health professionals from focus group studies. J Med Internet Res 2003 Dec 18;5(4):e32 [FREE Full text] [doi: 10.2196/jmir.5.4.e32] [Medline: 14713660]
6. Philpott A, Knerr W, Boydell V. Pleasure and prevention: when good sex is safer sex. Reproductive Health Matters 2006 Nov;14(28):23-31. [doi: 10.1016/S0968-8080(06)28254-5] [Medline: 17101419]
7. 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]
8. University College London. 2012. Sexunzipped URL: http://www.sexunzipped.co.uk/ [accessed 2012-09-19] [WebCite Cache ID 6AnTQ8n6r]
9. Carswell K, McCarthy O, Murray E, Bailey J. Integrating psychological theory into the design of an online intervention for sexual health: the Sexunzipped website. JMIR Res Protoc 2012 (forthcoming). [doi: 10.2196/resprot.2114]
10. Livingstone, S. The challenge of engaging youth online: contrasting producers' and teenagers' interpretations of websites. Eur J Commun 2007;22(2):165-184 [FREE Full text]
11. Selkie EM, Benson M, Moreno M. Adolescents' views regarding uses of social networking websites and text messaging for adolescent sexual health education. Am J Health Educ 2011 Dec;42(4):205-212. [Medline: 22229150]

12. Franck LS, Noble G. Here's an idea: ask the users! Young people's views on navigation, design and content of a health information website. J Child Health Care 2007 Dec;11(4):287-297. [doi: 10.1177/1367493507083941] [Medline: 18039731]
13. Levine D, McCright J, Dobkin L, Woodruff AJ, Klausner JD. SEXINFO: a sexual health text messaging service for San Francisco youth. Am J Public Health 2008 Mar;98(3):393-395. [doi: 10.2105/AJPH.2007.110767] [Medline: 18235068]
14. Forrest S, Strange V, Oakley A, The Ripple Study Team. What do young people want from sex education? The results of a needs assessment from a peer-led sex education programme. Cult Health Sex 2004 Jul;6(4):337-354. [doi: 10.1080/13691050310001645050] [Medline: 21972906]
15. Kaestle CE, Allen KR. The role of masturbation in healthy sexual development: perceptions of young adults. Arch Sex Behav 2011 Oct;40(5):983-994. [doi: 10.1007/s10508-010-9722-0] [Medline: 21293916]
16. Ingham R. 'We didn't cover that at school': education against pleasure or education for pleasure? Sex Educ 2005 Nov;5(4):375-388. [doi: 10.1080/14681810500278451]
17. Marie Stopes International. Likeitis URL: http://www.likeitis.org/indexuk.html [accessed 2012-06-01] [WebCite Cache ID 686EuBqAr]
18. Scarleteen/Heather Corinna. 2012. Scarleteen: Sex Ed for the Real World URL: http://www.scarleteen.com/ [accessed 2012-06-01] [WebCite Cache ID 686FF30C4]
19. UK National Health Service, Brighton & Hove City Council. Swish: So What is Sexual Health? URL: http://www.swish.org.uk/index.cfm?request=a2300 [accessed 2012-06-01] [WebCite Cache ID 686FRy1HX]
20. youthhealthtalk.org. 2012. Sexual Health URL: http://www.youthhealthtalk.org/Sexual_Health_of_Young_People/ [accessed 2012-06-01] [WebCite Cache ID 686Ffgsqa]
21. Channel 4. 2012. Sexperience URL: http://sexperienceuk.channel4.com/ [accessed 2012-06-01] [WebCite Cache ID 686FvcgZ3]

Edited by J Keelan; submitted 19.03.12; peer-reviewed by G Kok, M Gerressu, C Fairley; comments to author 24.05.12; revised version received 12.06.12; accepted 14.06.12; published 12.10.12.

Please cite as:
McCarthy O, Carswell K, Murray E, Free C, Stevenson F, Bailey JV
What Young People Want From a Sexual Health Website: Design and Development of Sexunzipped
J Med Internet Res 2012;14(5):e127
URL: http://www.jmir.org/2012/5/e127/
doi:10.2196/jmir.2116
PMID:23060424

Original Paper

Economic Evaluation of Internet-Based Interventions for Harmful Alcohol Use Alongside a Pragmatic Randomized Controlled Trial

Matthijs Blankers[1,2], PhD; Udo Nabitz[1], PhD; Filip Smit[3,4], PhD; Maarten WJ Koeter[2], PhD; Gerard M Schippers[2], PhD

[1]Department Jellinek, Arkin Mental Health Care, Amsterdam, Netherlands

[2]Amsterdam Institute for Addiction Research, Department of Psychiatry, Academic Medical Centre, University of Amsterdam, Amsterdam, Netherlands

[3]Trimbos Institute, Netherlands Institute of Mental Health and Addiction, Utrecht, Netherlands

[4]EMGO Institute, Department of Epidemiology and Biostatistics, VU University Medical Centre, Amsterdam, Netherlands

Corresponding Author:
Matthijs Blankers, PhD
Department Jellinek
Arkin Mental Health Care
PO Box 75848
Amsterdam, 1070 AV
Netherlands
Phone: 31 611785270
Fax: 31 205905692
Email: m.blankers@amc.uva.nl

Abstract

Background: Internet interventions with and without therapist support have been found to be effective treatment options for harmful alcohol users. Internet-based therapy (IT) leads to larger and longer-lasting positive effects than Internet-based self-help (IS), but it is also more costly to provide.

Objective: To evaluate the cost effectiveness and cost utility of Internet-based interventions for harmful use of alcohol through the assessment of the incremental cost effectiveness of IT compared with IS.

Methods: This study was performed in a substance abuse treatment center in Amsterdam, the Netherlands. We collected data over the years 2008–2009. A total of 136 participants were included, 70 (51%) were female, and mean age was 41.5 (SD 9.83) years. Reported alcohol consumption and Alcohol Use Disorders Identification Test (AUDIT) scores indicated harmful drinking behavior at baseline. We collected self-reported outcome data prospectively at baseline and 6 months after randomization. Cost data were extracted from the treatment center's cost records, and sex- and age-specific mean productivity cost data for the Netherlands.

Results: The median incremental cost-effectiveness ratio was estimated at €3683 per additional treatment responder and €14,710 per quality-adjusted life-year (QALY) gained. At a willingness to pay €20,000 for 1 additional QALY, IT had a 60% likelihood of being more cost effective than IS. Sensitivity analyses attested to the robustness of the findings.

Conclusions: IT offers better value for money than IS and might therefore be considered as a treatment option, either as first-line treatment in a matched-care approach or as a second-line treatment in the context of a stepped-care approach.

Trial Registration: Netherlands Trial Register NTR-TC1155; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=1155 (Archived by WebCite at http://www.webcitation.org/6AqnV4eTU)

(J Med Internet Res 2012;14(5):e134) doi:10.2196/jmir.2052

KEYWORDS

Cost-benefit analysis; randomized controlled trial; alcohol-induced disorders; self-help; computer-assisted therapy

Introduction

Harmful alcohol use is the number-3 leading contributor to global burden of disease [1] and causes 3.8% of global mortality [2], as well as losses in gross domestic product [3]. The majority of people with alcohol use disorders are not receiving any form of treatment, leading to a treatment gap [4]. Among the possible means of bridging this treatment gap is the use of accessible and efficient treatment, delivered over the Internet. Internet interventions with and without therapist support [5-8] have been found to be effective treatment options for harmful alcohol users and could perhaps be used sequentially in a stepped-care format. Internet-based therapy (IT) leads to larger and longer-lasting positive effects than Internet-based self-help (IS) in the treatment of depression [9,10], anxiety [10,11], and problem drinking [8]. However, IT is more costly to provide and more demanding for both participants and therapists. Thus, the research question is, “Does the additional positive result of therapist support outweigh its additional cost?” We present an economic evaluation assessing the cost effectiveness and cost utility of IT compared with IS for harmful alcohol use. Recently, studies have been published on the cost effectiveness of Internet-based (self-help) interventions for depression [12], weight management [13], and harmful alcohol use [14]. The cost effectiveness of therapist support in Internet-based alcohol interventions has not yet been supported, however.

Methods

Study Design and Participants

We collected data for the cost effectiveness analysis alongside a pragmatic randomized controlled trial on the effectiveness of IT relative to IS and a waiting list, conducted in the Netherlands in 2008–2009. Because in economic evaluation the preferred comparison is between the intervention of interest (IT) and its best alternative, in this case IS, we do not present waiting list data in this paper.

We recruited applicants through jellinek.nl, a substance abuse treatment center website with 650,000 visitors annually [8]. For inclusion, applicants had to (1) be between 18 and 65 years old, (2) live in the Netherlands with health care insurance coverage, (3) have Internet access at home, (4) score above 8 on the Alcohol Use Disorders Identification Test (AUDIT) [15], (5) report a weekly consumption of more than 14 standard (10 g ethanol) drinking units, and (6) provide informed consent. Exclusion criteria were (1) prior substance abuse treatment, (2) a history of alcohol delirium or drug overdose, (3) a history of severe cardiovascular or gastrointestinal diseases, (4) a history of schizophrenia, epilepsy, or suicidal tendencies, (5) extensive substance use in the last month, and (6) unavailability of more than 2 weeks during the study. Of the 1720 who were assessed, 832 applicants were eligible for inclusion; 205 participants were included. Compared with all 832 eligible applicants, the 205 included participants reported somewhat higher baseline AUDIT scores, but this difference was not significant (mean 18.9, SD 4.98 vs mean 19.5, SD 5.13, $t_{204} = 1.617$, $P = .11$). In the IT group, 48 received the allocated intervention, that is, they participated in treatment exercises and chat therapy. In IS, 57 received the intervention, which consisted of exercises only. Outcome data were collected at baseline, 3 months, and 6 months after randomization (Figure 1). The study design [16] and outcomes of the randomized controlled trial [8] were published elsewhere.

Interventions

Both IT and IS were based on a cognitive behavioral therapy and motivational interviewing treatment protocol [17]. In IS, participants were introduced to various treatment exercises. Without a therapist’s support, participants acquired skills and knowledge about coping with craving, drinking lapses, and peer pressure. IT was driven by 7 synchronous text-based chat-therapy sessions with a personal (Internet) cognitive behavioral therapy-trained therapist, lasting 40 minutes each, and accompanied by homework assignments. Each of the chat-therapy sessions had its own theme: monitoring and goal setting, self-control, and relapse prevention, for example.

Cost Measures

In this economic evaluation, we used the societal perspective. All costs related to IT and IS interventions, health care uptake, opportunity costs of the participant’s time, and productivity losses were included. All costs (Table 1) are expressed in euros and were indexed to the reference year 2010 using an inflation correction based on the Harmonized Index of Consumer Prices (HICP) [18].

Figure 1. CONSORT trial flow diagram for the randomized controlled trial. AUDIT = Alcohol Use Disorders Identification Test, ITT = intention-to-treat analysis, IS = Internet-based self-help, IT = Internet-based therapy.

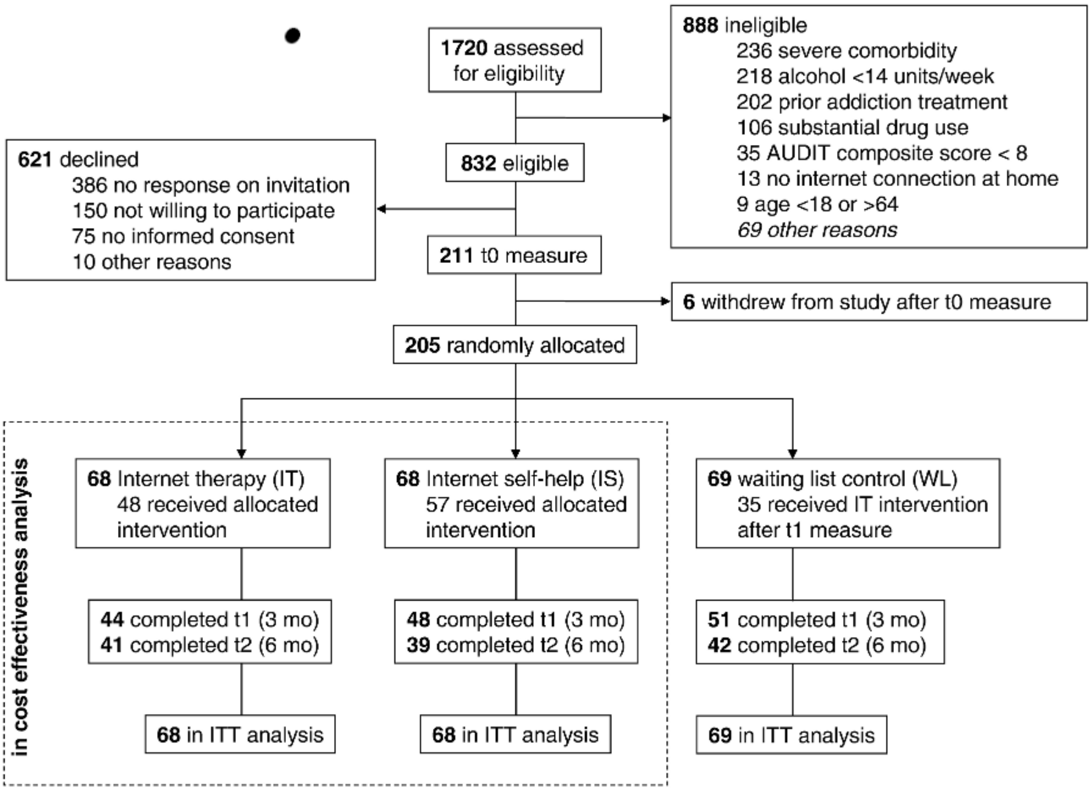


Table 1. Unit costs and average quantities per participant for Internet-based therapy and Internet-based self-help.

Cost type	Unit	Internet therapy		Internet self-help	
		No. of units	€/unit	No. of units	€/unit
Intervention costs					
Therapist therapy	Hour	2.49	79.20	NA[a]	NA
Therapist administration	Hour	0.55	79.20	NA	NA
Software development	Participant	1	23.25	1	4.87
ICT[b] service	Participant	1	14.92	1	2.49
Software overhead	Participant	1	4.27	1	4.27
Total intervention costs	Participant	1	283.21[c]	1	11.63[c]
Participant’s leisure time	Hour	10.33	9.18	2.43	9.18
Work absenteeism[d]	Hour	32.12	22.21–52.91[e]	18.35	22.21–52.91[e]
Work presenteeism[f]	Hour	8.15	22.21–52.91[e]	12.15	22.21–52.91[e]

[a] Not applicable.

[b] Information and computer technology.

[c] Average intervention cost per participant. Individual costs varied and depended on the amount of intervention uptake.

[d] Average number of work hours lost in the 6 months preceding measurement due to participants not going to work (eg, sick leave).

[e] Range of unit cost. The unit value was dependent on sex and age of the participant and based on 2010 Harmonized Index of Consumer Prices inflation-corrected average hourly wages [19].

[f] Average number of work hours lost in the 6 months preceding measurement due to participants not functioning well professionally while at work.

IT and IS intervention costs consisted of software development costs, information and computer technology service costs, overhead costs (based on the treatment center's cost records), and—for IT only—therapist-related costs. We collected the cost data over the years 2004-2009. Information and computer technology service costs were based on averaged annual costs and included server rental costs, software security costs, and a monthly information and computer technology support fee. Overhead costs were based on actual time investment estimations. Time invested was multiplied by labor costs based on collective labor agreement wages, with 50% additional employer costs for overhead and insurance. Development, information and computer technology service, and overhead costs were divided by the monthly recorded number of participants (IT: 25; IS: 50). Therapist costs were based on the actual chat-contact time, with an added 10 minutes per chat session for supervision and administrative work. Therapist work time was valued based on average sex-, age-, and profession-specific labor costs in the Netherlands [19], which resulted in €1.32 per minute in 2010. It is worth noting that the cost price for labor was in line with the costs (€80) for a single contact session with a primary care psychologist in the Netherlands in 2009 [20].

We restricted participant costs to a valuation of their time investment, valued as leisure time at €9.18 per hour [19], assuming that the therapy was not received during their office hours. Time investment for participants per treatment session, including homework, was 20 minutes (based on user inquiry) plus the therapy duration in the case of IT chat therapy. We collected data on productivity losses stemming from absenteeism and presenteeism using the Short Form-Health and Labor Questionnaire (SF-HLQ), a subscale of the Trimbos/iMTA questionnaire for Costs associated with Psychiatric illness [21]. Productivity costs were collected over a 2-week period before data collection, in correspondence with the SF-HLQ manual [21]. The reported costs over this 2-week period were then extrapolated. This method was found to be valid in patients with cluster B personality disorders [22], but we did not validate it in the current population of harmful alcohol users. To value inefficient job performance, these data were combined with sex- and age-specific mean productivity cost data for the Netherlands [20]. We used an elasticity estimate of 0.8, as suggested by the Netherlands Economic Institute [23]; we assumed that in case of absence, 20% of the production had not been lost but was compensated for by a firm's internal labor reserves. Considering the limited time horizon of collected cost data in this study, duration of absenteeism was valued according to the human capital approach. We therefore regarded cost as accrued for the full period of absenteeism and not limited to a friction period [24].

Additional societal costs were calculated using a macroscopic approach based on global burden of disease and injury data [2]. For high-income countries, productivity losses are the primary contributor to total alcohol-attributable costs: productivity loss accounts for 72.1% of the overall societal costs [2]. Additional health care resource costs (12.8%) and law-enforcement costs (3.5%) were estimated based on productivity cost data. Costs due to property damage, administration, or social work services were not taken into account, as these costs are excluded in most economic evaluations. The timeframe for this study and all time-variant costs was 6 months.

Effect Measures

The central clinical outcome for the cost effectiveness analysis was treatment response, based on alcohol consumption during the last 7 days. In the study protocol we defined treatment response as alcohol consumption within the British Medical Association boundaries (no more than 14 standard units for women, or 21 units for men, per week) [25], with an additional provision that participants did not present with a deterioration of more than 10% on the AUDIT [15], the Quality Of Life Scale [26], or the Global Severity Index of the Brief Symptom Inventory [27]. In other words, any such deterioration precluded our definition of treatment response. Positive treatment response, meeting these criteria, should be interpreted as desirable outcome of treatment and covers the wider aspects of problem drinking beyond drinking quantities only.

The central outcome for the cost utility analysis was the number of quality-adjusted life-years (QALYs) as calculated with the 5-dimensional EuroQol (EQ-5D) [28] using Dolan's UK tariff to obtain preference-based utilities [29]. We calculated QALYs taking into account the 6-month timeframe of this study.

Data Analysis

We carried out all analyses on an intention-to-treat basis. Missing observations in costs and effects data were handled using multiple imputation. The multiple imputation software package Amelia II [30] for R [31] has been found to yield the most accurate results in the type of data used in this study [32]. Amelia II takes into account the covariance structure between all variables, as opposed to some other approaches that require explication of covariates. Using this software, we imputed the original dataset 5 times, a sufficient number of imputations according to Rubin's analysis of the required number of imputed sets needed for the missingness rates in the sample analyzed [33]. Analyses were performed on each of these 5 datasets separately, and the outcomes were then combined using Rubin's rules for combining estimates obtained from multiply imputed datasets [33] for means. For medians and for the figures, we combined (appended) the data from the imputations. The relative attrition between the trial arms was low: IT:IS = ((41/68) / (39/68)) = 1.05 (Figure 1). Analyses were performed using SPSS 17.0 (IBM Corporation, Somers, NY, USA) and R 2.11.0 [31] software.

Cost and Effect Data

We analyzed cost and effect data according to methods suggested by Drummond and colleagues [34]. For all participants, we multiplied units of health care (eg, sessions, contacts), time investments, and productivity losses by their associated costs. Differences in costs and effects between IT and IS were calculated at the 6-month follow-up measurement, because randomization had resulted in sufficient comparability across conditions at baseline (Table 2).

Table 2. Baseline characteristics of participants in Internet-based therapy (IT) and Internet-based self-help (IS).

Characteristic	IT (n = 68)	IS (n = 68)	t_{134}/ Fisher exact test	*P* value
Women, n (%)	35 (51%)	35 (51%)	0.00	1.00
Age (years), mean (SD)	41.9 (10.1)	41.1 (9.6)	0.49	.63
Education, n (%) [a]			4.49	.10
Low	2 (3%)	7 (11%)		
Medium	24 (38%)	30 (46%)		
High	38 (59%)	29 (44%)		
Employed, n (%)	58 (85%)	55 (82%)	0.25	.65
Residential urbanization level, n (%)			0.74	.75
Low	9 (13%)	6 (9%)		
Medium	21 (31%)	22 (32%)		
High	37 (55%)	40 (59%)		
AUDIT[b] composite score, mean (SD)	18.8 (4.8)	19.6 (5.6)	0.98	.33
Duration of alcohol problems (years), mean (SD)	5.2 (5.7)	5.4 (5.7)	0.23	.82
Drinks per week, mean (SD)	45.2 (26.3)	43.4 (24.0)	0.38	.71
EQ-5D[c] score	0.79 (0.20)	0.80 (0.18)	0.32	.75
Work absenteeism[d]	756 (2289)	1863 (6983)	1.24	.22
Work presenteeism[d]	1137 (2386)	794 (1922)	0.78	.44

[a] Classified according to Statistics Netherlands (CBS) and International Standard Classification of Education 1997.

[b] Alcohol Use Disorders Identification Test [15].

[c] 5-dimensional EuroQol instrument, score calculated using the measurement and valuation of health (MVH-A1) algorithm from Dolan [29].

[d] Averaged costs over the 6 months preceding baseline measurement.

Bootstrapping

We extracted 1000 nonparametric bootstrapped [35] samples (n = 68 per trial arm) from each of the 5 multiply imputed datasets. For each of these 5 × 1000 bootstrapped samples, we calculated the incremental costs, incremental effects, and incremental cost effectiveness ratio (ICER). This ICER was calculated as follows: ICER = $(C_{IT} – C_{IS})/(E_{IT} – E_{IS})$, where C is costs, E is effects, and the subscripts IT and IS refer to the two interventions. As effects, we used two outcome measures: (1) proportion of treatment responders, and (2) QALYs.

Cost Effectiveness Plane

The resulting 1000 ICERs per dataset were used for further calculations and plotted on the cost effectiveness plane [36] (Figure 2). The reference intervention (IS) is positioned in the origin of the cost effectiveness plane. The horizontal axis indicates differences in health gains between IT and IS and the vertical axis represents differences in costs. Along the horizontal and vertical axis, Figure 2 is divided into quadrants, each with a specific interpretation. ICERs that fall in the upper right quadrant indicate that IT generated better health for additional costs; the lower left quadrant indicates a reduction in health gains for fewer costs. In the upper left quadrant, IT is dominated by IS, as poorer health outcomes were obtained at additional costs. In the lower right quadrant, IT dominates IS with better health outcomes for fewer costs. The median values of the bootstrapped ICERs are presented in the Results section.

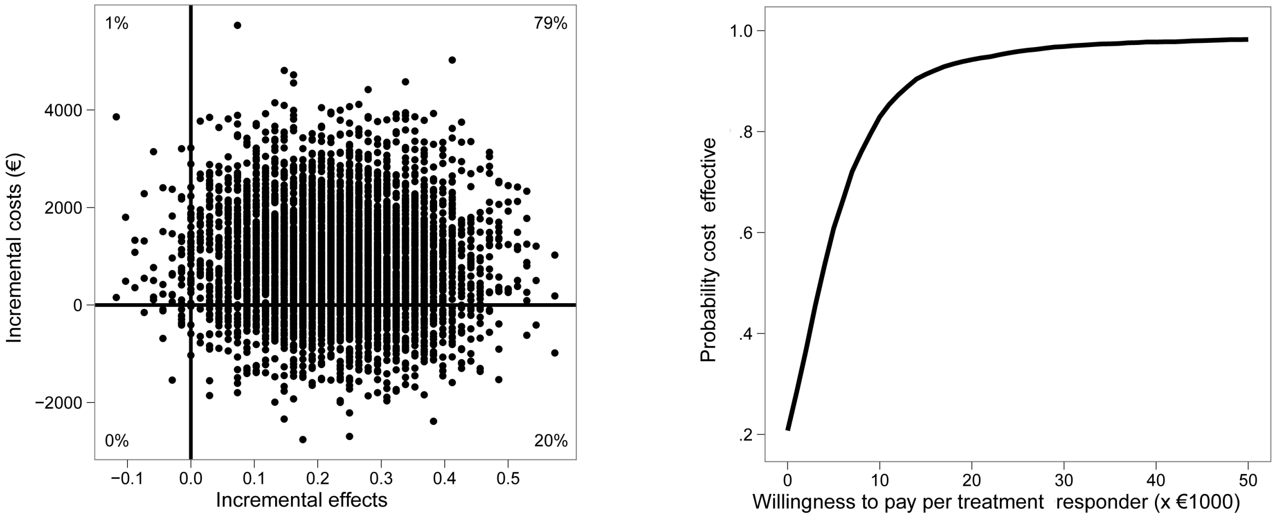


Figure 2. Cost effectiveness plane (left) and cost effectiveness acceptability curve (right) with treatment response as the effect measure.

Cost Effectiveness Acceptability Curve

Based on the distribution of the ICERs over the cost effectiveness plane, cost effectiveness acceptability curves [37] were drawn (Figure 3). The cost effectiveness acceptability curves show the probability that IT is more cost effective than IS as a function of the willingness to pay (WTP) for 1 additional unit of effect (1 treatment responder or 1 QALY). At a probability of 0.5 on the vertical axis, the indifference point is reached. Above this indifference point, IT is to be preferred over IS with regard to cost effectiveness. WTP is an unknown quantity and therefore presented as a series of increments on the horizontal axis.

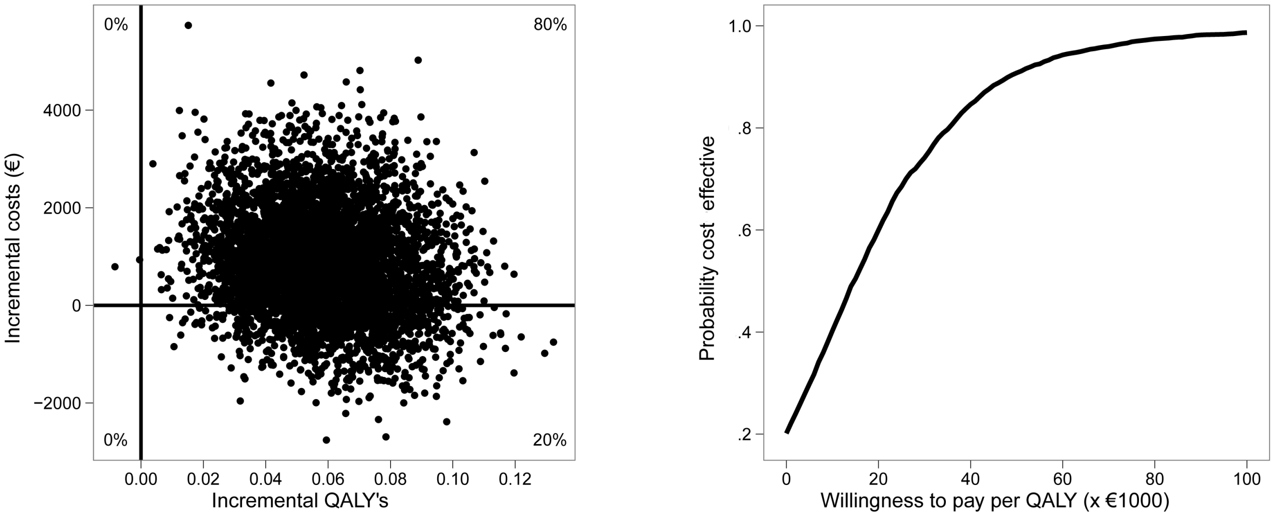


Figure 3. Cost effectiveness plane (left) and cost effectiveness acceptability curve (right) with quality-adjusted life-year (QALY) as the effect measure.

Sensitivity Analysis

To test the robustness of the economic evaluation, we performed a sensitivity analysis in which we varied the most relevant cost drivers. First, the cost effectiveness analysis was replicated from the health care provider perspective, including only health care costs in the analysis. In other alternative scenarios, the influence of the most influential cost drivers (ie, intervention costs and productivity costs) was explored. These costs drivers were raised and lowered independent of each other, in order to test the influence of adjustments on the median ICER and the likelihood that IT is more cost effective than IS.

Results

Participants

Of the 136 participants included in this cost effectiveness analysis, 68 were randomly assigned to IT and 68 to IS. The participants (n = 70, 51% women) were a mean of 41.5 (SD 9.83) years old (Table 2). Reported drinking frequencies and AUDIT composite score indicated harmful drinking behavior at baseline. None of the baseline characteristics differed markedly between the groups.

Costs

Per-participant costs in IT and IS, and bootstrapped incremental costs are presented in Table 3. All costs were estimated for the 6-month period between baseline and follow-up. Total intervention costs for IT and IS were on average €283 and €12, respectively (Table 1). Software development costs were, contrary to what is sometimes thought, prospective costs and not sunk costs. Complex software products such as the e-mental health interventions in this study needed continuous updates, bug fixes, security adjustments, and improvements to make sure they functioned with more recent browsers, operating systems, etc. Therefore, development costs were to a large extent running costs, and it is common in eHealth cost effectiveness studies to include these costs in the cost analysis (eg, [12-14]). For both groups, the largest cost drivers at follow-up were costs due to productivity losses (IT: €1331; IS: €886). The difference between IT and IS in mean costs of work absenteeism and presenteeism seems considerable and relevant (though not statistically significant), but must be seen in light of the actual number of participants who reported these costs. At baseline, 8 participants in IT and 12 participants in IS reported absenteeism; for presenteeism at baseline, these counts were 21 for IT and 15 for IS. The number of participants reporting absenteeism 6 months later dropped to 5 for IT and 2 for IS; presenteeism was reported by 7 in IT and 5 in IS. Total average societal costs for IT, €2010, were higher than the average €1120 for IS. The median difference of the societal costs between IS and IT was €845, which means that IT was more costly than IS from a societal perspective. The main incremental cost drivers were productivity costs and intervention costs.

Table 3. Costs and increments in the 6-month period preceding follow-up of the Internet-based therapy (IT) and Internet-based self-help (IS) groups[a].

Cost type	IT		IS		Bootstrapped difference	
	Mean	SD	Mean	SD	Median	95% CI[b]
Intervention costs						
Therapist labor	241	236	0	0	240	187–296
Software development	23	0	5	0	18	18–18
Software/hardware service	15	0	2	0	12	12–12
Software overhead	4	0	4	0	0	0–0
Total intervention costs	283	236	12	0	271	217–327
Participant time investment costs	95	103	22	37	72	48–99
Productivity costs						
Work absenteeism	1114	5704	536	3800	555	–967 to 2234
Work presenteeism	217	847	350	1637	–119	–609 to 256
Total productivity costs	1331	5774	886	4215	417	–1215 to 2208
Societal costs						
Additional societal costs[c]	301	1305	200	953	94	–275 to 499
Total societal costs	2010	7141	1120	5167	845	–1157 to 3048
Treatment response (proportion)	0.53		0.29		0.24	0.07–0.38
EQ-5D[d] score	0.89	0.20	0.78	0.34	0.12	0.05–0.18
ICER[e] treatment response					3683	–5703 to 20,366
ICER QALY[f]					14,710	–18,337 to 71,664

[a] All costs have been rounded for presentation in this table, and may therefore not add up exactly.

[b] Confidence interval.

[c] An estimation of real costs, based on Rehm et al [2], and includes additional health care resource costs and law-enforcement costs.

[d] 5-dimensional EuroQol instrument, score calculated using the measurement and valuation of health (MVH-A1) algorithm from Dolan [29].

[e] Incremental cost effectiveness ratio.

[f] Quality-adjusted life-year.

Effects

Table 3 shows the proportion of favorable treatment response and the EQ-5D scores. In IT, 36 / 68 = 0.53 responded well to treatment after 6 months; in IS this was 20 / 68 = 0.29. Incremental effectiveness of IT compared with IS was therefore 0.53 - 0.29 = 0.24. Dolan’s [29] EQ-5D scores for IT and IS at 6 months, which can be used for cost utility analysis, were 0.89 and 0.78, respectively. The incremental utility gain of IT relative to IS can thus be calculated as 0.89 - 0.78 = 0.12. Considering the 6-month timeframe of this study, and the fact that mortality of participants in this study was zero, the number of incremental QALYs gained with 1 IT intervention compared with 1 IS intervention can be calculated as 0.12 * (6 / 12) = 0.06 [34].

Cost Effectiveness Analysis

By dividing the incremental costs by the incremental effects, the mean ICER of IT compared with IS from the societal perspective is calculated as €845/0.24 = €3521 for 1 additional treatment responder, 6 months after inclusion. Using the bootstrapping procedure, we estimated the median ICER to be €3683. In the cost effectiveness plane (Figure 2, left), each dot represents a bootstrapped mean ICER. By calculating the proportion of simulated ICERs in each of the 4 quadrants, we found that IT had a 79% probability of leading to additional effects at additional costs relative to IS. A total of 20% fell into the dominant quadrant, indicating that there was a 20% likelihood that IT led to additional effects at lower societal costs (Table 4). The WTP at 50% was €3683 per additional treatment responder. Above a WTP of €3683 per additional treatment responder, IT must be considered cost effective in comparison with IS.

Cost Utility Analysis

The mean incremental societal costs for 1 additional QALY gained by IT compared with IS were €845 / 0.06 = €14,083. The median ICER for 1 extra QALY was estimated too be €14,710. From Figure 3 (left) it becomes clear that there was an 80% probability that IT led to a better QALY health gain at additional costs, while 20% of the ICERs fell into the dominant quadrant. The cost effectiveness acceptability curve (Figure 3, right) suggests that at a WTP of €20,000 for 1 additional QALY, the probability that IT was more cost effective than IS was at 60% (Table 4). At a WTP of €14,710 or more for 1 additional QALY, IT must be considered cost effective in comparison with IS.

Sensitivity Analysis

In Table 4, alternative costing scenarios are explored. From the health care provider perspective, the median ICER was €1157 per additional treatment responder, or €4693 per additional QALY. In other alternative costing scenarios, the main incremental cost drivers (intervention costs, costs due to productivity losses, and associated societal costs) were adjusted over a range of ±60%, in order to explore their impact on the ICERs. The results for ±40% adjustments are presented in Table 4. We found that ICERs were more sensitive to changes in productivity losses than to changes in intervention costs. Adjustments in both intervention and productivity costs led to the largest changes in ICERs. In all sensitivity scenarios in Table 4, the point of indifference from the cost effectiveness perspective between IT and IS was below a WTP of €20,000 per QALY, indicating that in any alternative scenario in this table, IT would be preferred over IS at a WTP of €20,000 or more per QALY.

Figure 4. Cost effectiveness acceptability curve after sensitivity analyses with treatment response (left) and quality-adjusted life-year (QALY) (right) as effect measures.

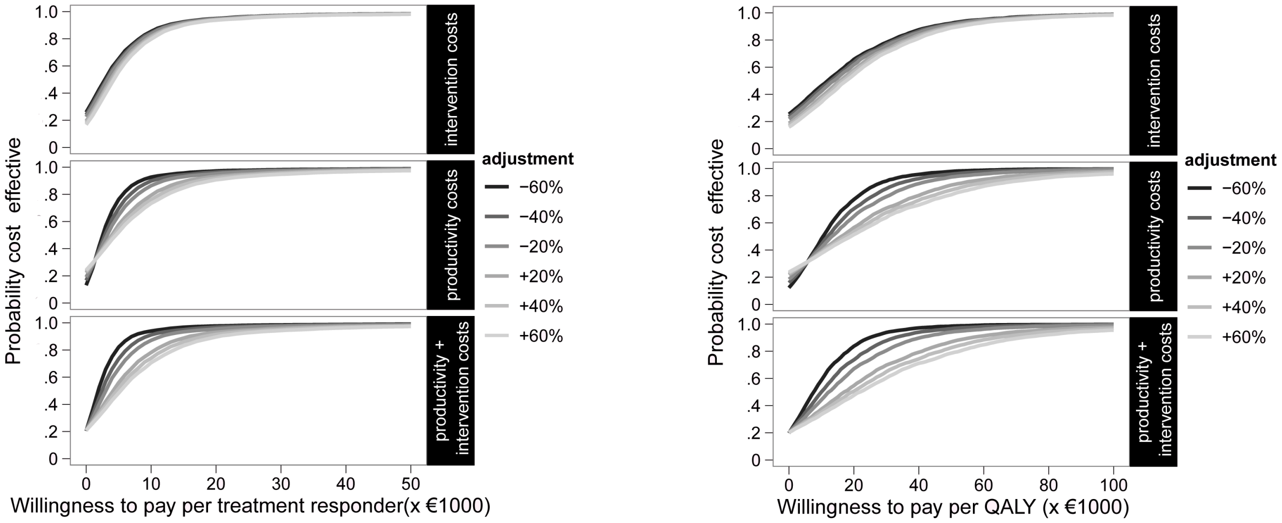


Table 4. Cost effectiveness analysis of base case, health care provider perspective, and additional sensitivity analyses.

Cost drivers	Base case: societal	Alternative case: health care provider	Sensitivity analyses					
			I[a] –40%	I +40%	P[b] –40%	P +40%	I and P –40%	I and P +40%
Incremental costs (median)	845	271	739	954	681	1012	573	1120
Treatment response								
Incremental effects (median)	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24
ICER[c] (median)	3683	1157	3187	4172	2977	4387	2494	4868
ICER (95%low)	–5703	665	–6441	–5050	–3227	–8313	–3821	–7576
ICER (95%high)	20,366	3722	19,410	21,409	14,724	25,979	13,738	26,957
WTP[d] €4000	53%	95%	57%	50%	62%	48%	66%	46%
WTP €8000	76%	98%	78%	74%	85%	69%	87%	67%
WTP €12,000	87%	99%	89%	86%	92%	82%	93%	80%
Upper right quadrant	79%	99%	76%	82%	83%	76%	79%	79%
Upper left (inferior) quadrant	1%	1%	1%	1%	1%	1%	1%	1%
Lower left quadrant	0%	0%	0%	0%	0%	0%	0%	0%
Lower right (dominant) quadrant	20%	0%	23%	17%	16%	22%	20%	20%
QALYs[e]								
Incremental QALYs (median)	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
ICER QALY (median)	14,710	4693	12,932	16,584	11,876	17,683	9946	19,436
ICER QALY (95%low)	–18,337	2783	–20,177	–16,241	–10,291	–26,220	–12,282	–24,352
ICER QALY (95%high)	71,664	10,848	67,913	75,671	52,202	91,101	48,403	94,958
WTP €10,000	40%	95%	45%	36%	44%	38%	50%	35%
WTP €20,000	60%	99%	64%	57%	70%	54%	74%	51%
WTP €40,000	85%	100%	87%	83%	93%	77%	94%	74%
Upper right quadrant	80%	100%	76%	83%	84%	77%	80%	80%
Upper left (inferior) quadrant	0%	0%	0%	0%	0%	0%	0%	0%
Lower left quadrant	0%	0%	0%	0%	0%	0%	0%	0%
Lower right (dominant) quadrant	20%	0%	23%	17%	16%	23%	20%	20%

[a] Intervention costs.

[b] Productivity costs.

[c] Incremental cost effectiveness ratio.

[d] Willingness to pay.

[e] Quality-adjusted life-year.

Discussion

Main Findings

In this cost effectiveness analysis, we found that the IT intervention led to almost double (0.53 versus 0.29) the number of treatment responders at 6 months, which was achieved at an incremental cost of €845 (equivalent to US $1008, based on purchasing power parity for the reference year 2010 [38]). Hence, 1 additional treatment responder for IT compared with IS was achieved at a median incremental cost of €3683 (US $4394). IT led to better EQ-5D health-related quality of life and 1 additional QALY was gained at a median incremental cost of €14,710 (US $17,548). At a WTP of €20,000 (US $23,859) [39] for 1 QALY gained, IT had a 60% likelihood of being more cost effective than IS. These results are somewhat sensitive to large adjustments in societal costs, but the alternative cost scenarios were also in favor of IT, assuming a WTP of, but not more than, €20,000 (US $23,859) for gaining an additional QALY.

Implications

The maximum WTP per QALY is a matter of debate, but the figure of €20,000 is conservative compared with the World Health Organization's recommendation of a maximum cost per QALY of 3 times the gross domestic product per capita (€88,000 for the Netherlands in 2010) [40]. The main findings of this study indicate that, from a cost effectiveness perspective, IT could be offered rather than IS, but the difference (60%) is not very large. In a conventional hypothesis-testing statistical approach, this difference would not be considered statistically significant, although such an approach is uncommon in cost effectiveness analyses alongside randomized controlled trials [34].

It is clear that the costs of providing IS from a health care provider perspective are only a fraction of those of providing IT. Alternatively, a stepped-care approach could be proposed, in which a client is referred to IS first and then is referred to IT in a second step when desirable results have not been achieved after IS.

Previous Studies

Smit and colleagues [14] found that Internet-based alcohol self-help has a 73% probability of dominating from a cost effectiveness point of view compared with a text-only information leaflet. They found a negative ICER (US -$13,950, ie, a cost saving), mainly due to lower costs of productivity losses in the self-help intervention. In the current study, we did not find a relative reduction in productivity losses between the two active interventions. Solberg and colleagues [41] reviewed cost effectiveness studies from the health care provider perspective for brief (non-Internet-based) alcohol interventions. The number of QALYs gained was found to be highly sensitive to the effectiveness of counseling. Screening and brief counseling compared with no intervention had an overall ICER of US $1755 (in year 2000 terms) per QALY gained. Compared with this figure, the median incremental cost per QALY for IT compared with IS is less favorable. However, the lower cost per gained QALY can be expected when a comparison between an active intervention and no intervention is made.

Strengths and Limitations

A limitation of this study stems from the generalizability of the cost data. The reported software costs were based on actual cost records, which may be different in other settings. To estimate full societal costs, we measured productivity loss cost data using the SF-HLQ, but we estimated health care costs other than the focal intervention and for law enforcement based on Rehm and colleagues [2]. Two potential issues arise: (1) data presented in Rehm et al [2] are not limited to harmful users, and (2) not all costs included as societal costs may be drinking related. Because additional health care costs and law-enforcement costs depend on productivity costs, we leveraged a potential bias in the measurement of productivity costs into these costs.

We collected data on productivity losses using the SF-HLQ, aiming at a 2-week period before data collection. Subsequently, we extrapolated the calculated costs of productivity losses. An alternative approach to calculating absence days could have been applied by (1) retrospectively asking participants about their work absenteeism in the previous 3 or 6 months, and thus collecting an alternative measure of absenteeism based on which we could have validated our extrapolation approach, or (2) measuring absenteeism, presenteeism, and the main clinical end points more frequently, in order to have more data on which to base the extrapolation. This would, however, have increased the research burden on our study participants, but it would presumably also have led to more sound data on costs (and effects). To assess the robustness of the results in terms of deviations from the calculated costs, we performed sensitivity analyses, plus an additional analysis in which we took into account only the intervention costs (health care provider cost perspective).

Another limitation was the time horizon in this analysis, which was restricted to 6 months. It is very possible that clinical effects were maintained after 6 months, although they may diminish over time. The same may be true for losses or gains in (productivity) costs. We have, however, not modeled the possible developments of effects and costs beyond the 6 months for which we have empirical data. This limits the time horizon and may jeopardize informed decision making when considering long-term effects.

By collecting patient-level cost data alongside a pragmatic randomized controlled trial, this study has both good comparability of the populations in the two interventions as a consequence of random allocation, and acceptable external validity as a result of the pragmatic approach [8]. Missing observations were multiply imputed; failing to account for missing costs data properly can produce biased results [42-44]. We subjected our base-case results to cost adjustments in a sensitivity analysis. Our main findings remained stable under alternative costing scenarios. At a WTP of €20,000 per QALY gained, IT offers outcomes for money equal to or better than those obtained with IS, and might therefore be considered as a possible treatment option, either as first-line treatment in a matched-care approach or as second-line treatment in the context of a stepped-care approach.

Acknowledgments

The randomized trial and the cost effectiveness analysis reported in this paper were funded by Grant #31160006 from the Netherlands ZonMw Addiction II Program (Risk Behavior and Dependency). The study was conducted in collaboration with the Jellinek Clinic, which is a division of Arkin, the Amsterdam-based mental health and addiction treatment center. Arkin supported the trial and facilitated its development. The two interventions evaluated in this study were developed at Jellinek. None of the funders of the project had any role in the design of the trial, selection of the measurement instruments, or preparation of the manuscript. The protocol of this randomized controlled trial has been published previous to the trial's execution. The trial was conducted in agreement with the Declaration of Helsinki and was approved by the Medical Ethics Committee of the Academic Medical Centre of the University of Amsterdam, the Netherlands. Publication of this paper was financially supported by grant #036.001.899 from the Netherlands Organization for Scientific Research (NWO) – Incentive Fund Open Access.

Conflicts of Interest

None declared.

References

1. Rehm J, Taylor B, Room R. Global burden of disease from alcohol, illicit drugs and tobacco. Drug Alcohol Rev 2006 Nov;25(6):503-513. [doi: 10.1080/09595230600944453] [Medline: 17132570]
2. Rehm J, Mathers C, Popova S, Thavorncharoensap M, Teerawattananon Y, Patra J. Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders. Lancet 2009 Jun 27;373(9682):2223-2233. [doi: 10.1016/S0140-6736(09)60746-7] [Medline: 19560604]
3. Casswell S, Thamarangsi T. Reducing harm from alcohol: call to action. Lancet 2009 Jun 27;373(9682):2247-2257. [doi: 10.1016/S0140-6736(09)60745-5] [Medline: 19560606]
4. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull World Health Organ 2004 Nov;82(11):858-866 [FREE Full text] [Medline: 15640922]
5. 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]
6. Cunningham JA, Wild TC, Cordingley J, van Mierlo T, Humphreys K. A randomized controlled trial of an internet-based intervention for alcohol abusers. Addiction 2009 Dec;104(12):2023-2032 [FREE Full text] [doi: 10.1111/j.1360-0443.2009.02726.x] [Medline: 19922569]
7. Rooke S, Thorsteinsson E, Karpin A, Copeland J, Allsop D. Computer-delivered interventions for alcohol and tobacco use: a meta-analysis. Addiction 2010 Aug;105(8):1381-1390. [doi: 10.1111/j.1360-0443.2010.02975.x] [Medline: 20528806]
8. Blankers M, Koeter MW, Schippers GM. Internet therapy versus internet self-help versus no treatment for problematic alcohol use: A randomized controlled trial. J Consult Clin Psychol 2011 Jun;79(3):330-341. [doi: 10.1037/a0023498] [Medline: 21534652]
9. Titov N, Andrews G, Davies M, McIntyre K, Robinson E, Solley K. Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010;5(6):e10939 [FREE Full text] [doi: 10.1371/journal.pone.0010939] [Medline: 20544030]
10. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
11. Robinson E, Titov N, Andrews G, McIntyre K, Schwencke G, Solley K. Internet treatment for generalized anxiety disorder: a randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010;5(6):e10942 [FREE Full text] [doi: 10.1371/journal.pone.0010942] [Medline: 20532167]
12. Gerhards SA, de Graaf LE, Jacobs LE, Severens JL, Huibers MJ, Arntz A, et al. Economic evaluation of online computerised cognitive-behavioural therapy without support for depression in primary care: randomised trial. Br J Psychiatry 2010 Apr;196(4):310-318 [FREE Full text] [doi: 10.1192/bjp.bp.109.065748] [Medline: 20357309]
13. Rasu RS, Hunter CM, Peterson AL, Maruska HM, Foreyt JP. Economic evaluation of an Internet-based weight management program. Am J Manag Care 2010 Apr;16(4):e98-104 [FREE Full text] [Medline: 20370312]
14. Smit F, Riper H, Schippers GM, Cuijpers P. cost effectiveness of a web-based self-help intervention for problem drinking: randomized trial. In: Riper H. Curbing Problem Drinking in the Digital Galaxy [PhD thesis]. Amsterdam: VU University Amsterdam; 2008.
15. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. Addiction 1993 Jun;88(6):791-804. [Medline: 8329970]
16. Blankers M, Koeter M, Schippers GM. Evaluating real-time internet therapy and online self-help for problematic alcohol consumers: a three-arm RCT protocol. BMC Public Health 2009;9:16 [FREE Full text] [doi: 10.1186/1471-2458-9-16] [Medline: 19144162]
17. De Wildt W. Achilles leefstijl 1 [Achilles Life style 1]. Zeist: Cure & Care Publishers; 2000.
18. Global-rates.com. 2010. Inflation the Netherlands (HICP) URL: http://nl.global-rates.com/economische-statistieken/inflatie/consumentenprijzen/hicp/nederland.aspx [WebCite Cache ID 5ufmHxUtI]
19. Oostenbrink JB, Bouwmans CAM, Koopmanschap MA, Rutten FFH. Handleiding voor Kostenonderzoek: Methoden en Standaard Kostprijzen voor Economische Evaluaties in de Gezondheidszorg (Geactualiseerde versie 2004) [Dutch Manual

for Costing: Methods and Standard Costs for Economic Evaluations in Health Care (actualized version 2004)]. Diemen: College voor Zorgverzekeringen; 2004.
20. Hakkaart-van Roijen L, Tan SS, Bouwmans CAM. Handleiding voor kostenonderzoek, methoden en standaard kostprijzen voor economische evaluaties in degezondheidszorg [Manual for Cost Research, Methods and Standard Cost Prices for Economic Evaluations in Health Care]. Diemen: College voor Zorgverzekeringen; 2010.
21. Hakkaart-van Roijen L, van Straten A, Donker M, Tiemens B. Manual Trimbos/iMTA Questionnaire for Costs Associated With Psychiatric Illness (TiC-P). Rotterdam: Institute for Medical Technology Assessment; 2002.
22. Soeteman DI, Hakkaart-van Roijen L, Verheul R, Busschbach JJ. The economic burden of personality disorders in mental health care. J Clin Psychiatry 2008 Feb;69(2):259-265. [Medline: 18363454]
23. de Koning J. Tuyl FAWM. De relatie tussen arbeidsduur, produktie en werkgelegenheid: Een analyse op basis van tijdreeksgegevens van drie bedrijven [The Relation Between Labor Time, Production and Employment]. Rotterdam: Netherlands Economic Institute; 1984.
24. Koopmanschap MA, Rutten FF, van Ineveld BM, van Roijen L. The friction cost method for measuring indirect costs of disease. J Health Econ 1995 Jun;14(2):171-189. [Medline: 10154656]
25. British Medical Association. Guidelines on Sensible Drinking. London: BMA; 1995.
26. Flanagan JC. A research approach to improving our quality of life. Am Psychol 1978;33:138-147. [doi: 10.1037/0003-066X.33.2.138]
27. Derogatis LR, Melisaratos N. The Brief Symptom Inventory: an introductory report. Psychol Med 1983 Aug;13(3):595-605. [Medline: 6622612]
28. EuroQol--a new facility for the measurement of health-related quality of life. The EuroQol Group. Health Policy 1990 Dec;16(3):199-208. [Medline: 10109801]
29. Dolan P. Modeling valuations for EuroQol health states. Med Care 1997 Nov;35(11):1095-1108. [Medline: 9366889]
30. Honaker J, King G, Blackwell M. Harvard University. 2008. Amelia II: A Program for Missing Data. R Package version 1 URL: http://gking.harvard.edu/amelia/
31. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2010.
32. Blankers M, Koeter MW, Schippers GM. Missing data approaches in eHealth research: simulation study and a tutorial for nonmathematically inclined researchers. J Med Internet Res 2010;12(5):e54 [FREE Full text] [doi: 10.2196/jmir.1448] [Medline: 21169167]
33. Rubin DB. Multiple Imputation for Survey Nonresponse in Surveys. Hoboken, NJ: Wiley-Interscience; 2004.
34. Drummond MF, Sculpher MJ, Torrance GW, O'Brien B, Stoddart GL. Methods for the Economic Evaluation of Health Care Programmes. Oxford: Oxford University Press; 2005.
35. Efron B. Bootstrap methods: another look at the jackknife. Ann Stat 1979;7:1-26. [doi: 10.1214/aos/1176344552]
36. Black WC. The CE plane: a graphic representation of cost effectiveness. Med Decis Making 1990;10(3):212-214. [Medline: 2115096]
37. van Hout BA, Al MJ, Gordon GS, Rutten FF. Costs, effects and C/E-ratios alongside a clinical trial. Health Econ 1994;3(5):309-319. [Medline: 7827647]
38. OECD StatExtracts. Organisation for Economic Co-operation and Development (OECD). 2011. Purchasing Power Parities and Exchange Rates in 2010 URL: http://stats.oecd.org/Index.aspx?DataSetCode=SNA_TABLE4 [WebCite Cache ID 5znuBuZga]
39. Raad voor de Volksgezondheid en Zorg. Zinnige en Duurzame Zorg [Sensible and Sustainable Care]. Zoetermeer: Raad voor de Volksgezondheid en Zorg; 2006.
40. World Health Organization. Macroeconomics and Health: Investing in Health for Economic Development. Report of the Commission on Macroeconomics and Health. Geneva: WHO; 2001.
41. Solberg LI, Maciosek MV, Edwards NM. Primary care intervention to reduce alcohol misuse ranking its health impact and cost effectiveness. Am J Prev Med 2008 Feb;34(2):143-152. [doi: 10.1016/j.amepre.2007.09.035] [Medline: 18201645]
42. Noble SM, Hollingworth W, Tilling K. Missing data in trial-based cost effectiveness analysis: the current state of play. Health Econ 2012 Feb;21(2):187-200. [doi: 10.1002/hec.1693] [Medline: 22223561]
43. Burton A, Billingham LJ, Bryan S. cost effectiveness in clinical trials: using multiple imputation to deal with incomplete cost data. Clin Trials 2007;4(2):154-161. [doi: 10.1177/1740774507076914] [Medline: 17456514]
44. Marshall A, Billingham LJ, Bryan S. Can we afford to ignore missing data in cost effectiveness analyses? Eur J Health Econ 2009 Feb;10(1):1-3. [doi: 10.1007/s10198-008-0129-y] [Medline: 18936993]

Abbreviations

AUDIT: Alcohol Use Disorders Identification Test
EQ-5D: 5-dimensional EuroQol
ICER: incremental costeffectiveness ratio
IS: Internet-based self-help

IT: Internet-based therapy
QALY: quality-adjusted life-year
SF-HLQ: Short Form-Health and Labor Questionnaire
WTP: willingness to pay

Edited by G Eysenbach; submitted 12.01.12; peer-reviewed by D Stant, N Shcherbakova; comments to author 06.02.12; revised version received 29.03.12; accepted 20.06.12; published 29.10.12.

Please cite as:
Blankers M, Nabitz U, Smit F, Koeter MWJ, Schippers GM
Economic Evaluation of Internet-Based Interventions for Harmful Alcohol Use Alongside a Pragmatic Randomized Controlled Trial
J Med Internet Res 2012;14(5):e134
URL: http://www.jmir.org/2012/5/e134/
doi:10.2196/jmir.2052
PMID:23103771

Original Paper

Evaluation of End-User Satisfaction Among Employees Participating in a Web-based Health Risk Assessment With Tailored Feedback

Sandra Vosbergen[1*], MSc; Eva K Laan[2,3*], MSc; Ersen B Colkesen[3,4], MD; Maurice AJ Niessen[3], MA; Roderik A Kraaijenhagen[3], MD, PhD; Marie-Louise Essink-Bot[2], MD, PhD; Niels Peek[1], PhD

[1]Academic Medical Center, Department of Medical Informatics, University of Amsterdam, Amsterdam, Netherlands
[2]Academic Medical Center, Department of Public Health, University of Amsterdam, Amsterdam, Netherlands
[3]NDDO Institute for Prevention and Early Diagnostics (NIPED), Amsterdam, Netherlands
[4]Academic Medical Center, Department of Cardiology, University of Amsterdam, Amsterdam, Netherlands
[*]these authors contributed equally

Corresponding Author:
Sandra Vosbergen, MSc
Academic Medical Center
Department of Medical Informatics
University of Amsterdam
P.O. Box 226600
Amsterdam, 1100DD
Netherlands
Phone: 31 20 5666179
Fax: 31 20 6919840
Email: s.vosbergen@amc.uva.nl

Abstract

Background: Web technology is increasingly being used to provide individuals with health risk assessments (HRAs) with tailored feedback. End-user satisfaction is an important determinant of the potential impact of HRAs, as this influences program attrition and adherence to behavioral advice.

Objective: The aim of this study was to evaluate end-user satisfaction with a web-based HRA with tailored feedback applied in worksite settings, using mixed (quantitative and qualitative) methods.

Methods: Employees of seven companies in the Netherlands participated in a commercial, web-based, HRA with tailored feedback. The HRA consisted of four components: 1) a health and lifestyle assessment questionnaire, 2) a biometric evaluation, 3) a laboratory evaluation, and 4) tailored feedback consisting of a personal health risk profile and lifestyle behavior advice communicated through a web portal. HRA respondents received an evaluation questionnaire after six weeks. Satisfaction with different parts of the HRA was measured on 5-point Likert scales. A free-text field provided the opportunity to make additional comments.

Results: In total, 2289 employees participated in the HRA program, of which 637 (27.8%) completed the evaluation questionnaire. Quantitative analysis showed that 85.6% of the respondents evaluated the overall HRA positively. The free-text field was filled in by 29.7 % of the respondents (189 out of 637), who made 315 separate remarks. Qualitative evaluation of these data showed that these respondents made critical remarks. Respondents felt restricted by the answer categories of the health and lifestyle assessment questionnaire, which resulted in the feeling that the corresponding feedback could be inadequate. Some respondents perceived the personal risk profile as unnecessarily alarming or suggested providing more explanations, reference values, and a justification of the behavioral advice given. Respondents also requested the opportunity to discuss the feedback with a health professional.

Conclusions: Most people were satisfied with the web-based HRA with tailored feedback. Sources of dissatisfaction were limited opportunities for providing additional health information outside of the predefined health and lifestyle assessment questionnaire and insufficient transparency on the generation of the feedback. Information regarding the aim and content of the

HRA should be clear and accurate to prevent unrealistic expectations among end-users. Involving trusted health professionals in the implementation of web-based HRAs may enhance the use of and confidence in the HRA.

(J Med Internet Res 2012;14(5):e140) doi:10.2196/jmir.2067

KEYWORDS

eHealth; satisfaction; health risk assessment; mixed methods; health behavior

Introduction

Health risk assessments (HRAs) with feedback are commonly used instruments for worksite health promotion [1]. HRAs with feedback screen for risk factors for chronic diseases and provide respondents with information about their health risks, current lifestyle behavior, and opportunities for improving their health. Several studies have already shown promising results for HRAs implemented at the worksite: they might reduce employees' health risks and improve their lifestyles [1]. This is also beneficial for the employer, as employees with healthy lifestyles are generally absent from work less often and are more productive than employees with unhealthy lifestyles [2-4].

Nowadays, HRAs are increasingly offered as web-based applications. The use of computer technology and email provides the opportunity to reach large groups of individuals and to deliver individually tailored feedback [5]. Following Kreuter et al [6], we define tailoring as "any combination of information or change strategies intended to reach one specific person, based on characteristics that are unique to that person, related to the outcome of interest, and have been derived from an individual assessment." Within the worksite setting, emails can reach diverse employee populations [7]. Emails and web pages offer a natural way of reaching employees, as it fits into their daily work routine. Employees are also able to use computer programs in private, for example, at home, and at a time that suits them [8].

Although the potential of web-based HRAs with tailored feedback at the worksite is clear, few evaluations take into account satisfaction from an end-user perspective. To date, studies that did take this into account were either focused on eHealth approaches (eg, telemedicine) [9,10] other than HRAs, studied satisfaction only as an effectiveness measure without studying factors that affect end-user satisfaction [11-13], or evaluated non−web-based health promotion programs [11,13,14]. However, end-user satisfaction is shown to be positively related to compliance to medical regimes in primary health care [15] and with initiation of health behavior change after participating in the HRA studied here [16]. Respondents who were more satisfied with the HRA were nearly three times more likely to initiate lifestyle changes after participation [16]. Yet to our knowledge, no studies have comprehensively studied satisfaction and the factors affecting it. Furthermore, more extensive evaluation of workplace health promotion programs is needed as the most effective strategies for these services have not yet been determined [17]. Evaluating the factors described within the conceptual framework of Wixom and Todd [18] can give insight into how web-based HRAs with tailored feedback are used and provide opportunities for improvement. By evaluating HRAs after implementing them in the proposed setting on a voluntary basis without remuneration, real-life personal experiences with the HRA could be assessed and used to inform the design process.

Objectives

The aim of this study was to evaluate end-user satisfaction with a web-based HRA with tailored feedback applied in worksite settings, using mixed (quantitative and qualitative) methods. Mixed methods can provide insights that may be missed when only a single research modality is used [19]. Satisfaction with different components of the HRA, and determinants affecting this satisfaction were evaluated both quantitatively and qualitatively. Qualitative methods were also used to determine opportunities for improvement.

Methods

Design and Study Population

This study was conducted between September 2007 and December 2008. Seven companies in the Netherlands invited their employees to participate in a commercial, web-based HRA with tailored feedback. The HRA was part of the companies' corporate health-management strategy. Two companies used an age-based inclusion criterion for participation in the HRA (35 years and older and 45 years and older, respectively). The other companies did not use inclusion criteria, and all employees were invited to participate. The companies' human resources departments sent the invitations for participation in the HRA by email. Employees were informed that participation was voluntary and free of charge, that all personal data would be treated confidentially, and that no individual results would be shared with their employer or with any other third parties. In case of no response to the invitation email, a single reminder was sent two weeks after the initial invitation.

Six weeks after participation in the HRA, participants received an invitation by email for an electronic evaluation questionnaire. This electronic evaluation questionnaire was not part of the HRA but was sent for research purposes.

Intervention

The HRA consisted of four components: 1) an electronic health and lifestyle assessment questionnaire, 2) a biometric evaluation, 3) a laboratory evaluation, and 4) tailored feedback consisting of a personal health risk profile and lifestyle behavioral advice communicated through a web portal.

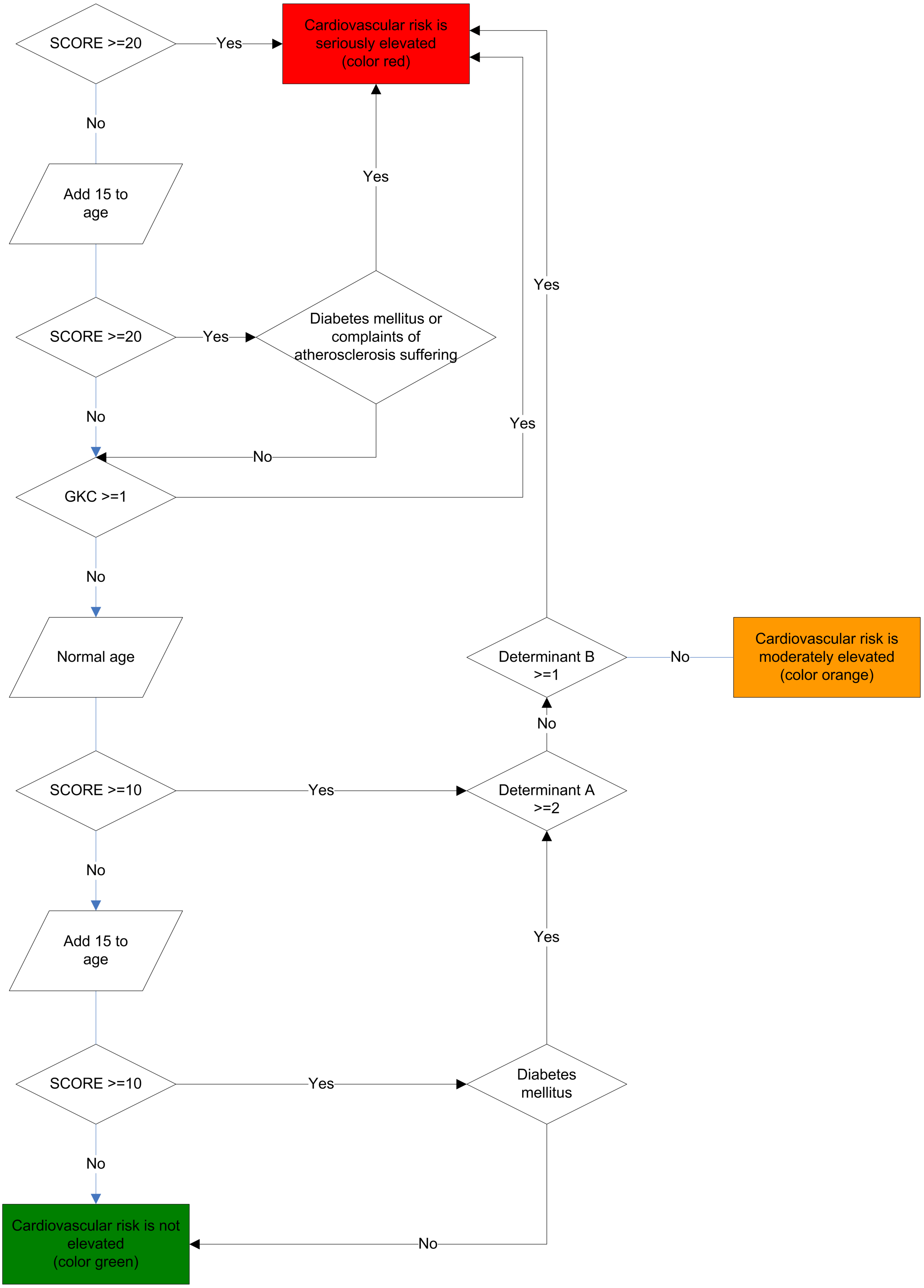
The health and lifestyle assessment questionnaire covered sociodemographic variables, family and personal medical history, health complaints, psychological functioning, and lifestyle behavior (physical activity, smoking behavior, alcohol consumption, nutrition intake, and stress). Furthermore, participants were asked to assess their own lifestyle on a scale from 1 to 10 and their health perception on a 5-point Likert scale (very good, good, neither good nor poor, poor, very poor). More details on the health and lifestyle assessment questionnaire can be found in Laan et al [20].

The biometric evaluation consisted of measurements of weight, height, waist circumference, and blood pressure taken by certified health professionals. Furthermore, samples of blood, urine, and feces were taken for lab analyses of total cholesterol, low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol, triglycerides, glucose, creatinine, and urinary albumin-to-creatinine ratio. To collect and analyze feces samples and provide feedback on the results, the developer of the HRA (NDDO Institute for Prevention and Early Diagnostics (NIPED)) received permission to screen for colorectal cancer under Dutch law, provided by the Dutch Ministry of Health, Welfare, and Sport.

After completing all HRA components, the information collected from both the biometric evaluation and the health and lifestyle questionnaire was processed by computer algorithms to compute the tailored feedback. This feedback was provided to the user immediately after completion of the questionnaire, the biometric evaluation, and after the laboratory provided feedback on the results of blood, urine, and feces. When the tailored feedback was available, the user received an email. The feedback was divided into five health-related domains (behavioral, psychological, physical, personal medical history/familiar risk, and work-related). For each of these domains, 1) a three-color system was used to explain the health risk (green: normal risk; orange: moderately elevated risk; red: seriously elevated risk), and 2) the threats associated with elevated risk (orange and red categories) and the potential gains of taking preventive action were explained. A compass metaphor was used to summarize overall health risk, with the categories “on track (color green)”, “slightly off-track (color light orange)”, “moderately off-track (color dark orange)”, and “seriously off-track (color red)”. In the remainder of this article we will refer to those categories by their colors. All risk calculations were based on prevailing practice guidelines, including the European and Dutch guidelines for cardiovascular risk management [21,22]. The feedback concluded with comprehensive suggestions of actions the participant could take. All options suggested trusted external parties the participant could go to for support for the action they might want to take. These suggestions were made based on their expressed preferences (such as for guided versus non-guided interventions, actions in groups or on their own, and actions away from or at home), and differentiated between the participant’s stage of motivation for lifestyle changes (transtheoretical model of health behavior change [23]). In case of seriously elevated health risks, the feedback included a referral to a general practitioner (GP) for further medical evaluation and treatment if necessary. For all participants, a 30-minute health counseling visit with the program physician was available on request. An example of feedback provided by the system is shown in Multimedia Appendix 1 (see also Figure 1), and a more extensive description of how the feedback is generated can be found in Multimedia Appendix 2 (see also Figure 2).

Figure 1. Screenshot of the personal health risk profile page.

Figure 2. Algorithm for cardiometabolic health risk.



Measurements

Six weeks after receiving the feedback, participants received a two-part electronic evaluation questionnaire. The first part assessed satisfaction with the HRA, and the second part assessed initiation of health behavior change after participation in the HRA. The questionnaire was sent using an email survey program, with a single reminder after one week. The findings on initiation of health behavior change have been published elsewhere [16].

The part of the evaluation questionnaire related to satisfaction with the HRA—which is the focus of this study—consisted of 10 items. Respondents were asked to appraise (1) pre-program information and communication about the HRA, (2) the registration procedure, (3) the electronic health and lifestyle assessment questionnaire, (4) the planning of the biometric evaluation visit, (5) the biometric evaluation visit, (6) the comprehensibility of the provided feedback, and (7) the optional health counseling visit with the program physician. These items refer to successive steps in the HRA procedure, each of which has to be completed before the next one is started; there is no overlap. Satisfaction with each step was measured on a 5-point ordinal ranking scale (excellent, very good, good, average, poor). In addition, overall satisfaction with the HRA was measured with two questions: (8) "How do you assess your participationin the HRA in general?" (measured on a 5-point ordinal ranking scale: excellent, very good, good, average, poor) and (9) "Would you recommend the HRA to others?" (measured on a 5-point agreement scale: definitely would, probably would, maybe, probably not, definitely not). Finally, a free-text field was available for additional comments. The corresponding instruction read "It's possible that some things were not contained in the above questions, or that you weren't able to express these things as you would have liked. If this is the case, please enter them below."

Data Analyses

Quantitative Analysis of Satisfaction With the HRA

Three items of the questionnaire were not included in the data analyses: pre-program information, the registration procedure, and the planning of the biometric evaluation visit. The first two markedly differed between companies, and no remarks in the free-text field on these components of the HRA were made. The planning of the biometric evaluation visit was excluded because this practical aspect had no direct link with the web-based component of the HRA.

Descriptive statistics were performed on all data to examine population characteristics and satisfaction with the HRA. For satisfaction with the different components of the HRA and the overall satisfaction with the HRA, the response options were dichotomized into "positive judgment" (excellent, very good, good) and "negative judgment" (average, poor) because of unbalanced ranking scales. Response bias to the evaluation questionnaire was checked by comparing respondents and non-respondents on demographic characteristics, educational level, health perception, self-rated lifestyle, body mass index, physical activity, fruit and vegetable intake, smoking behavior, alcohol consumption, and overall health risk. Response bias to the free-text field was checked by comparing respondents who made remarks in the free-text field and respondents who did not on demographic characteristics and satisfaction. For both response bias analyses, *t* test for continuous outcomes, Chi-square tests for dichotomous outcomes, and Mann-Whitney (for two groups) and Kruskal-Wallis tests (for more than two groups) for ordinal outcomes were used. The Statistical Package for the Social Sciences (SPSS) version 18 was used to perform statistical analyses.

Qualitative Analysis of Satisfaction With the HRA

To analyze the textual remarks, a codebook was developed based on concepts from the user satisfaction and technology acceptance literature [18]. User satisfaction and technology acceptance theories are used to evaluate users' perceptions about information systems to predict actual usage of these systems. The theories use a characteristics-based approach, with the potential end-user's attitude towards a system as pivotal construct. There are different theories that take varying system characteristics into account. Because we aimed to evaluate the HRA in the broadest sense, we included in our codebook all concepts as described in Wixom and Todd, 2005 Table 1, p. 88) [18]. At the start of the analysis, all domains and concepts were adopted directly from the table. Interpretations of both were specified to the HRA evaluated in the current study. Domains, concepts, and interpretations are listed in Multimedia Appendix 3.

Two researchers (SV and EL) independently categorized the remarks according to three different topic schemes: 1) the component of the HRA addressed (pre-program information and communication about the HRA, the registration procedure, the electronic health and lifestyle assessment questionnaire, the planning of the biometric evaluation visit and the visit itself, the feedback provided by the system, counseling visit with the program physician, or the HRA in general), 2) domain and concept from the codebook (see Multimedia Appendix 3), and 3) whether the remark was positive, negative, or neutral. Considering that remarks sometimes referred to different components of the HRA, they were divided into shorter remarks when necessary, one for each component of the HRA addressed. If these shorter remarks covered different concepts from the codebook, they were again divided into even smaller remarks. If in the original codebook a remark did not pertain to any concept of a domain, we looked for a concept from another domain and copied that between the domains. For example, although "feelings of control" was originally described only in the domain "outcome expectations," in our analysis this concept was also needed in the domain "information quality," so we added this concept here. If no suitable concept was found, a new one was added. Changes to the codebook made during the analyses are also described in Multimedia Appendix 3.

During four subsequent meetings, we tried to reach consensus about the categorization by comparing and discussing this. If consensus could not be reached, a third researcher (NP) was brought in to resolve the disagreement.

Analysis of Determinants

Univariate logistic regression analyses were used to assess the associations between respondent characteristics (gender, age, educational level, health perception, self-rated lifestyle, and overall health risk) and satisfaction with each of the program components (health and lifestyle assessment questionnaire, visit for biometric evaluation, comprehensibility of feedback, and health counseling visit with program physician) as well as overall satisfaction with the HRA and intention to recommend the HRA to others. Prior to the analyses, age was categorized into "younger than 50 years" and "50 years and older." Because a majority of the respondents rated their own lifestyles with the number 7, this variable was categorized into three categories: 1−6, 7, and 8−10. Baseline data also showed that the distribution of health perception was unbalanced; only a few respondents assessed their health perception as “neither good nor poor”, “poor”, or “very poor”. Therefore, health perception was categorized into “very good”, “good”, and “poor” (comprising the categories “neither good nor poor”, “poor”, and “very poor”). Finally, the response scale for intention to recommend the HRA to others was dichotomized into “positive intention” (definitely would, probably would) and “negative intention” (maybe, probably not, definitely not). Overall, 36 univariate logistic regression analyses were performed, the results of which were quantified as odds ratios.

If a respondent characteristic was found to be significantly associated with a judgment of satisfaction with (components of) the HRA, the remarks that were entered into the free text field were subsequently searched for explanations. This was done by contrasting the remarks of different groups of respondents, as defined by the characteristic. For instance, explanations of significant associations with age were searched for by contrasting the remarks made by respondents younger than 50 years with those made by respondents of 50 years and older.

Results

Study Population

A total of 6790 employees were invited to participate in the HRA. Of those, 2289 (33.7%) completed all components of the HRA measurements and thus received feedback by the system. 183 out of 2472 (7.4%) dropped out after they started with the HRA. The response to the evaluation questionnaire was 27.8% (637/2289). Table 1 shows the characteristics of the respondents who completed the evaluation questionnaire.

Table 1. Characteristics of respondents to the evaluation questionnaire (N=637).

Variable	Value
Male sex, N (%)	386 (60.6)
Age, mean ± SD	46.49 ± 8.76
≥ 50, N (%)	230 (36.1)
Education level	
low, N (%)	139 (21.8)
moderate, N (%)	191 (30.0)
high, N (%)	307 (48.2)
Health perception	
very good, N (%)	116 (18.2)
good, N (%)	415 (65.1)
not good, not bad, N (%)	102 (16.0)
bad, N (%)	4 (1)
very bad, N (%)	0 (0)
Self-rated lifestyle (1−10), mean ± SD	7.18 ± 0.98
1−6, N (%)	119 (18.7)
7, N (%)	275 (43.2)
8−10, N (%)	243 (38.1)
Body Mass Index, mean ± SD	28.09 ± 3.77
Physical activity (min/week), median [interquartile interval]	165 [60−290]
Fruit (pieces/day), mean ± SD	1.76 ± 0.43
Vegetables (g/day), mean ± SD	1.56 ± 0.50
Currently smoking, N (%)	146 (22.9)
Alcohol consumption	
less than 1 unit/week, N(%)	160 (25.1)
1–7 units/week, N(%)	268 (42.1)
8–14 units/week, N(%)	126 (19.8)
15–21 units/week, N(%)	54 (9)
22 or more units/week, N (%)	29 (5)

There were no significant differences between respondents and non-respondents to the evaluation questionnaire, except for age. On average, respondents to the evaluation questionnaire were one year older (mean ± SD: 46.5 ± 8.8 years) than those who did not respond to the evaluation questionnaire (mean ± SD: 45.5 ± 8.5 years) (*P* =.01) (data not shown). Except for age and physical activity, there were also no significant differences between respondents who made remarks in the free-text field (either a negative, positive, or neutral remark) compared to those who did not. Respondents who made remarks in the free-text field were nearly two years older (mean ± SD: 47.8 ± 8.1 years) than those who did not make remarks (mean ± SD: 45.9 ± 9.0 years) (*P* =.01) (data not shown). Respondents who made remarks in the free-text field were less physically active (median [interquartile interval (IQI)]: 150 (50–270) min/week) compared to those who did not made remarks (median [IQI]: 170 (60–300) min/week) (*P*<.01) (data not shown).

Satisfaction With the HRA

Table 2 shows the satisfaction rankings of the respondents to the evaluation questionnaire. Additional analyses showed that fewer respondents who made one or more remarks in the free-text field were satisfied overall with the HRA (134 out of 189, 70.9%) compared to those who did not make remarks (411 out of 448, 91.7%) (*P*<.01) (data not shown). Similar figures were found for the health and lifestyle assessment questionnaire, for the visit for biometric evaluation, for the comprehensibility of the feedback, and for the intention to recommend the HRA to others. However, more respondents who made remarks judged the health counseling visit with the program physician as positive (40 out of 69, 58.0%), compared to those who did not make remarks (130 out of 157, 29.0%) (*P*<.01).

Table 2. Numbers and percentages of respondents that judged positively and negatively about (components of) the HRA (N=637).

Judgments of HRA components	Number (percentage)
Health and lifestyle assessment questionnaire	
positive, N (%)	557 (87.4)
negative, N (%)	80 (13)
Biometric evaluation visit	
positive, N (%)	550 (86.3)
negative, N (%)	82 (13)
no answer	5 (1)
Comprehensibility of the feedback	
positive, N (%)	515 (80.8)
negative, N (%)	119 (18.7)
no answer	3 (1)
Health counseling visit with the program physician	
positive, N (%)	170 (75.2[a])
negative, N (%)	56 (25[a])
Overall satisfaction with the HRA	
positive, N (%)	545 (85.6)
negative, N (%)	92 (14)
Intention to recommend the HRA to others	
definitely would, N (%)	274 (43.0)
probably would, N (%)	198 (31.1)
maybe, N (%)	107 (16.8)
probably not, N (%)	44 (7)
definitely not, N (%)	14 (2)

[a] Percentage based on those respondents who actually visited the program physician (N=226).

End-users' Remarks About the HRA

The free-text field at the end of the questionnaire was used to analyze the data qualitatively; 189 out of the 637 respondents (29.7%) filled in this field. One hundred and twelve respondents made one remark, 45 respondents made two remarks, 20 respondents made three remarks, and 12 respondents made four or more remarks. In total, 315 separate remarks were made. Of those 315 remarks, 33 (10%) were positive, 249 (78.3%) were negative, and 33 (10%) were neutral.

Assessment Phase

80 remarks were made by 70 respondents about the assessment phase. Of those remarks, 3 (4%) were positive, 71 (88.8%) were negative, and 6 (8%) were neutral.

Most remarks regarding the health and lifestyle assessment questionnaire were about limitations in opportunities for expressing individual details. There were respondents who felt limited in their ability to enter all the information about their health or lifestyle that they considered relevant (eg, existing health problems like food allergies and back pain). It was also mentioned that the questionnaire was insufficiently tailored to respondents' personal situation (eg, a lack of gender-specific questions). This resulted in the perception that the assessed personal health risk profile did not adequately represent respondents' actual health status.

> *I realize you want the phrasing of the questions to be as clear as possible. In a number of cases, the answers are oversimplified. The actual situation is sometimes far removed from the possible answers, and consequently the results also give a different (more negative) picture.* [Male, age 52]

Because of this perceived restriction, these respondents indicated they preferred speaking to a health professional after completing the health and lifestyle assessment questionnaire or expected to be able to provide additional information or ask questions during the visit for biometric evaluation. However, this visit was facilitated by an external organization that had only limited information about the HRA. Their only task was to obtain the measurements. This affected respondents' satisfaction with the biometric evaluation.

> *I thought the biometric evaluation visit was quite basic. All of the procedures were carried out in a rather impersonal way and at breakneck speed. It felt a little bit like a production line.* [Female, age 38]

To improve the system with regard to the perceived restrictions, respondents suggested adding a text field to the questionnaire where they could enter additional information or ask questions.

Personal Risk Profile and Lifestyle Behavior Advice

83 remarks were made by 68 respondents about the personal risk profile and lifestyle behavior advice. Of those remarks, 1 (1%) was positive, 75 (90.4%) were negative, and 7 (8%) were neutral.

The personal risk profile was criticized because of the format of the information provided. There were respondents who perceived the risk profile as unnecessarily alarming.

> *The way in which the results were presented meant I didn't sleep very well for a number of nights, even though afterwards the company doctor and my GP said there was no reason to take any further steps.* [Female, age 40]

Feelings of anxiety and lack of confidence in the feedback provided by the system often resulted in a second opinion by the respondent's own GP.

Others criticized the lifestyle behavior advice for being too complex, and sometimes for being too trivial (eg, respondents felt the advice included complex medical terminology or perceived the advice as simple as it told them to exercise more). Furthermore, there were respondents who expected more guidance in the execution of the lifestyle behavior advice.

> *I had the impression I'd get more help. But the answers I got were things I already knew. For instance, that I'd like help in trying to lose weight. It just said I could contact my GP.* [Male, age 36]

Finally, some respondents argued that the suggestions for taking action were easy to ignore.

> *Because the communication is all on paper, it's easy to ignore any recommendations.* [Male, age 55]

The HRA was also criticized because the relation between information entered and feedback provided by the system was not always clear to respondents.

> *I'd like to see a clearer connection between output and input. For example, if there's a recommendation to eat more dairy products, does this come out of the tests or from the questionnaire? I think it's important that the client knows what a recommendation is based on.* [Male, age 40]

Especially when the feedback contained unexpected information, respondents wanted to review their answers on the questionnaire and have the possibility to link these to the provided feedback—but the HRA did not support this. Perceived contradictions in the feedback sometimes led to irritation.

> *What stands out for me is that even though I didn't get any red scores, I'm seriously off-track and was referred to the prevention consultant. This doesn't add up.* [Male, age 47]

Also, some respondents mentioned a lack of confidence in the feedback provided because it was contrary to their expectations.

Many respondents who made a remark mentioned that the personal risk profile did not provide threshold values, especially with regard to the biometric measures. As a result, some of these respondents indicated they did not know how to interpret the feedback.

> *When the results are shown, there's not enough mention of reference materials … the result is 4, but on which scale? … What's good and what isn't, and when or at what score should I be concerned?* [Male, age 42]

Applying the Advice in Practice

39 remarks were made by 28 respondents about applying the advice in practice. Of those remarks, 2 (5%) were positive, 34 (87.2%) were negative, and 3 (8%) were neutral.

Despite the fact that a health counseling visit was available to every respondent, there were respondents who criticized the HRA for the absence of such a visit. This was likely due to a communication problem. Other respondents said that web-based feedback was rather impersonal compared to feedback provided by a health care professional.

Respondents made remarks about their experiences after receiving the feedback. Most respondents who went to their GP either because the provided feedback included referral to their GP or because they wanted a second opinion, underwent another biometric evaluation. Reasons they reported for this were that the GP perceived the feedback as being unclear and unusable, disagreed with the threshold values that were used for risk stratification, or claimed that certain measures were not performed correctly.

Overall Satisfaction With the HRA

104 remarks were made by 84 respondents about overall satisfaction with the HRA. Of those remarks, 27 (26.0%) were positive, 64 (61.5%) were negative, and 13 (12.5%) were neutral.

Respondents who made remarks in the free-text field differed in their opinions about the HRA's usefulness. There were respondents who considered the HRA to be superficial, commercial, of limited value, and/or not useful without a long-term trajectory:

> *To me, it was an automated, watered-down version of my own input. I reported that I had high blood pressure and—surprise—the program affirms this. I say I sometimes don't eat two servings of fruit a day. And presto, the program says I should eat more fruit. I indicate I'm going to exercise more, but the program isn't interested in this because I've already sufficiently answered the questions on exercise.* [Male, age 53]

Furthermore, respondents indicated a preference for a more personal approach.

> *During a face to face talk, you could have given a lot more information and clarified things and also have had a more thorough physical examination.* [Female, age 42]

For a few respondents, participating in the HRA was very useful because it warned them of serious health problems such as hypertension and cancer. They underwent appropriate medical interventions and were grateful they participated, something that was illustrated by the following respondent:

> *A polyp has been removed from my intestines on two different occasions. According to the specialist, one of these would certainly have become malignant.* [Male, age 46]

Positive remarks were also made by respondents who received confirmation they were in good health, respondents who perceived the HRA as being a good point of reference for their health, and those who considered prevention in general to be useful, as illustrated by the following quote:

> *I think it's a good idea to have one's weight checked, blood tested for cholesterol, and so forth. Everyone should. Sometimes diseases, abnormalities, come to light (not now, in my case), and then it makes a big difference if this is caught in time. As far as I'm concerned, this can be done every year or every two years.* [Female, age 40]

The biometric evaluation was perceived by some respondents as the most useful component of the HRA.

Determinants Affecting Satisfaction With the HRA

Table 3 shows the associations between respondent characteristics and their satisfaction with (components of) the HRA. There were significant differences between men and women in the judgment of the health and lifestyle assessment questionnaire. Women were almost twice more likely to be dissatisfied than men (OR=0.54, *P*=.01). Qualitative analysis showed that there were no differences in the contents of remarks made by men and women in the free-text field though.

Furthermore, significant differences were found in the judgment of the comprehensibility of the feedback amongst groups with different risk profiles. None of the respondents with a green overall risk profile evaluated the comprehensibility of the feedback negatively. Those who received a light orange (OR=1.93, *P*=.02) and dark orange (OR=1.87, *P*<.01) risk profile were more likely to judge the comprehensibility of the feedback positive than those who received a red risk profile. Also, the intention to recommend the HRA varied by risk profile; those who received a green risk profile were much more likely to have the intention to recommend the HRA to others than those who received a red profile (OR=4.52, *P*=.02).

Comparisons of the remarks made by respondents with different risk profiles showed that respondents with a red risk profile perceived more difficulties with applying the lifestyle behavior advice and reported anxiety as a result of the health risk profile than respondents with other risk profiles. For instance, remarks regarding the applicability of the lifestyle behavior advice or disagreement with the provided feedback by GPs were made by respondents with a red risk profile. No remarks regarding the provided feedback were made by respondents with a green risk profile.

In general, respondents with a green risk profile made relatively more positive remarks compared to respondents with other risk profiles. Also, their remarks were about practical issues like the service provided or materials used at the biometric evaluation visit and the evaluation questionnaire itself and no issues related to the web-based component of the HRA.

Table 3. Associations of positive judgments with respondent characteristics, expressed as odd ratios with 95% confidence intervals (N=637).

	Judgment of the health and lifestyle assessment questionnaire	Judgment of the visit for biometric evaluation	Judgment of the comprehensi-bility of feedback	Judgment of the health counseling visit with the program physician	Overall satisfaction with the HRA	Intention to recommend the HRA to others
Gender						
Male	ref. category	ref. category	ref. category	ref. category	ref. category	ref. category
Female	0.54 [0.34–0.87][b]	0.79 [0.49–1.26]	0.87 [0.58−1.30]	0.73 [0.40−1.35]	0.74 [0.47–1.16]	0.69 [0.48–0.99][b]
Age						
< 50	ref. category	ref. category	ref. category	ref. category	ref. category	ref. category
50 and older	0.93 [0.58−1.52]	1.34 [0.82–2.22]	0.91 [0.61–1.38]	1.40 [0.76–2.58]	0.82 [0.52−1.28]	1.18 [0.81−1.71]
Educational level						
Low	ref. category	ref. category	ref. category	ref. category	ref. category	ref. category
Moderate	1.02 [0.55−1.91]	1.03 [0.55−1.90]	1.38 [0.80−2.17]	0.76 [0.36−1.62]	1.45 [0.79–2.68]	1.32 [0.81–2.16]
High	1.40 [0.77–2.53]	1.49 [0.82–2.67]	1.32 [0.81–2.17]	1.47 [0.68–3.16]	1.25 [0.72–2.15]	1.23 [0.79−1.92]
Health perception						
Poor	ref. category	ref. category	ref. category	ref. category	ref. category	ref. category
Good	1.04 [0.56–1.91]	0.65 [0.33−1.29]	1.09 [0.65−1.83]	0.82 [0.39–1.73]	0.76 [0.41–1.42]	0.86 [0.53−1.40]
Very good	2.23 [0.90–5.49]	1.35 [0.54−3.40]	2.38 [1.12–5.06][b]	1.96 [0.62–6.16]	1.81 [0.75–4.37]	1.64 [0.86–3.14]
Self-rated lifestyle						
1 t/m 6	ref. category	ref. category	ref. category	ref. category	ref. category	ref. category
7	1.14 [0.61–2.13]	0.48 [0.23–1.03]	1.00 [0.58–1.72]	0.739 [0.34−1.61]	0.98 [0.53−1.81]	0.70 [0.43−1.16]
8 t/m 10	1.28 [0.67–2.44]	0.52 [0.24–1.12]	1.10 [0.63–1.94]	1.03 [0.46–2.29]	0.99 [0.53−1.85]	0.96 [0.57−1.61]
Overall health risk						
Green	0.83 [0.29–2.37]	6.20 [0.81–47.2]	N/A[a]	4.14 [0.46 – 34.64]	4.07 [0.93–17.75]	4.52 [1.13–15.37][b]
Light orange	1.28 [0.60–2.72]	0.93 [0.50–1.73]	1.93 [1.10–1.87][b]	1.30 [0.56–3.02]	1.15 [0.63–2.08]	0.79 [0.49–1.29]
Dark orange	0.71 [0.40–1.25]	1.27 [0.73–2.19]	1.87 [1.19–2.94][b]	1.75 [0.88–3.49]	1.62 [0.97–2.72]	1.24 [0.81–1.90]
Red	ref. category	ref. category	ref. category	ref. category	ref. category	ref. category

[a] Odds-ratio and 95% confidence interval could not be calculated because all respondents with a green overall health risk evaluated the comprehensibility of the feedback positive.

[b] *P*< .05.

Discussion

Main Findings

This study evaluated end-user satisfaction with a web-based HRA with tailored feedback. Quantitative evaluation data showed that most respondents were satisfied with the HRA in general and with its constituent components. Nearly three-quarters (74%) of the respondents indicated that they would definitely recommend the HRA to others or considered doing so. Overall, respondents who were negative about (components of) the HRA also made critical remarks in the free-text field.

Critical remarks about the HRA found in the qualitative analysis related to perceived control over the entered information, confidence in the generated feedback, and embedding of the HRA in the health care system. Several respondents requested more insight into the generation of the feedback. This and the other results of this study suggest that transparency of the underlying computer system is important for confidence in the feedback provided. If the feedback provided by the HRA did not match the respondent's expectations, they wanted the system to provide convincing arguments for it, based on the information they had entered. Several remarks also indicated that there were respondents who had more confidence in health professionals than in a computer system. Respondents often mentioned that they went to their GP for a second opinion, as they had no confidence in the feedback provided. There were also respondents who requested the opportunity to discuss the feedback with a health professional.

Relationship to Other Studies

In a systematic review of interventions for worksite health promotion, Soler et al reported two potential adverse effects of HRAs with feedback [1]. First, the feedback provided by the HRA could cause anxiety in participants, and second, there may be false-positive results (ie, the feedback incorrectly indicates there is health problem). Both adverse effects were reported in our study. Some respondents became anxious when the health plan showed alarming results. Consequently, most of them indicated that they went to their GP for a second opinion. The GP often told these respondents there was nothing to worry about. These signals may either have been false alarms or correct warnings that were not recognized by the GPs. In principle, the latter seems more likely as both the risk calculations and feedback given by the HRA followed prevailing Dutch practice guidelines. So probably GPs were sometimes acting against their own guidelines during the visit. We think this can be explained by unfamiliarity of GPs with the HRA and perhaps a lack of transparency for GPs with respect to the decision rules underlying the feedback. Another possible explanation is the fact that GPs in the Netherlands act as "gatekeepers" to the health care system and tend to act reluctantly.

While there are an increasing number of studies on information systems targeted directly at care consumers or patients, there is more extensive literature on the use of information systems by medical professionals. A systematic review of clinical decision support systems for medical professionals by Kawamoto et al showed that clinical practice is more likely to improve if users are provided with actionable recommendations rather than mere assessments [24]. The HRA evaluated in our study does provide such actionable recommendations, which were appreciated as such by the respondents. Yet our findings also indicate that the advice should be suitably adapted to the user's needs, preferences, and characteristics. Kawamoto et al also found that the effectiveness of decision support systems might potentially benefit if recommendations given by the system are justified by providing the underlying line of reasoning or research evidence [24]. Our findings support that justifying the provided recommendations has added value for the system's end-users, especially for their confidence in the recommendations.

In a study on an automated health behavior change intervention, Bickmore et al found that relational behaviors (ie, empathy and social dialogue) improved the liking of and the satisfaction, relationship, and desire to continue with the system [25]. There were respondents in our study who indicated they preferred to discuss their feedback with a health professional face to face, which supports Bickmore's finding that adding relational behaviors to the HRA interface might increase the perceived value of the HRA.

In their qualitative study, Wolff et al found various opportunities and barriers to disease prevention counseling in primary care [26]. The three aspects participants of their study requested most were tailored information, encouragement, and follow-up. The results of our study are in line with these findings: among the respondents, there were requests for more tailoring of information, positive feedback, and in some cases respondents felt the need for a second opinion or another form of professional follow-up after using the HRA.

Strengths and Limitations

In our study, we evaluated end-user satisfaction with a web-based HRA using both quantitative and qualitative methods. As described in the introduction, using mixed methods can provide additional insights that may be missed when only a single research modality is used [19]. In our case, although respondents were generally satisfied with the HRA, the qualitative data contained predominantly critical remarks. The qualitative data showed important aspects to consider when developing or implementing a web-based HRA. Furthermore, by analyzing respondent's remarks in the broadest sense, this study gave insight into both aspects related to web-based part of the HRA as well as the implementation of the HRA.

During the categorization of remarks, all domains and most concepts of the framework were used. Only one new concept was added, and some concepts were used in more than one domain or not used at all (see Multimedia Appendix 3). Because no new domains were needed and only a few adjustments were made to the framework, we conclude that the used framework was suitable for analyzing the data of this study.

The present study does have several limitations. A large majority of participants did complete all elements of the HRA; however, 183 out of 2472 (7.4%) participants dropped out before completing the HRA and did not receive an evaluation questionnaire. Furthermore, the response to the evaluation questionnaire was 28%, which is lower than the mean response of 60% to 67% in most satisfaction surveys [27,28]. However, our response is comparable to the response for general email health surveys, which is around 34% [29]. In addition, no indication for selection bias was found, except for age.

Two companies applied an age-based criterion because they believed that HRAs are more beneficial to older employees. Companies without an age-based criterion argued they wanted to give all their employees, regardless of age, the opportunity to participate in the intervention. This might have caused a selection bias. However, complementary analysis showed this bias is small. There is a small difference in the distribution of age between all companies included in this study (age, mean ± SD: 46.5 ± 8.8) and companies without the age-based criterion (age, mean ± SD: 45.9 ± 8.9). For gender, the difference between all companies and those with the age-based criterion was respectively 386 out of 637 (60.6%) and 351 out of 553 (63.5%). Furthermore, this analysis showed that there were no significant differences between companies with and without this age-based criterion on satisfaction, except for satisfaction with the health counseling visit with the program physician (data not shown).

Self-reported data concerning health status should be interpreted with care, as they may be influenced by social factors. In addition, the phrasing of the question of the free-text field might have influenced the number and the subjects of the remarks made by respondents. The free-text field was not a mandatory field, which could explain the low percentage of respondents who actually made remarks. In general, free-text fields take more time and therefore more effort to fill in, which could also explain the low number of remarks. More important is that the free-text field was intended for additional remarks, which could have resulted in a lower number of these. Interpreting the term "additional" as "other than what was mentioned in the evaluation questionnaire" could also result in remarks unrelated to the items contained in the evaluation questionnaire. This could explain the apparent discrepancy between the results found in the quantitative and the qualitative data. Despite the limitations of the phrasing of the free-text field question, we assumed that if respondents had remarks, they would write those down anyway. Subsequently, unbalanced ordinal ranking scales were used in the evaluation questionnaire. An unbalanced scale was used to rank satisfaction with the different components of the HRA program. Therefore we dichotomized the scale into positive (excellent, very good, good) and negative (average, poor) rankings. The item "Intention to recommend the HRA to others" was ranked on a balanced scale.

Several components of the HRA (pre-program information, the registration procedure, and the planning of the biometric evaluation visit) were excluded beforehand from the analysis. Therefore, we cannot draw conclusions about the above described components of the HRA or about the influence of these components on the satisfaction rankings. Furthermore, although we believe the framework we used was suitable for the categorization of our data, some of the remarks of participants did not fully fit the framework. Therefore, extending the framework with, for example, aspects of the "Diffusion of innovations" theory [30] might have put our findings in another perspective.

Finally, the quantitative analysis showed that critical remarks in the free-text field were made by respondents who were less satisfied with the HRA. This suggests that those who are less satisfied are more inclined to make remarks, which explains the predominantly negative remarks. Still, we believe that these remarks should be taken seriously and that they could support us in improving future HRAs and increase the utilization of HRAs.

Meaning of the Study

Our study showed that in general end-user satisfaction with the web-based HRA was high, but qualitative analysis of free-text field remarks indicated that there still are opportunities to increase satisfaction. We know from a previous study that satisfaction with the HRA is positively related to initiation of health behavior change after participation [16]. This suggests that by improving satisfaction based on our findings, we also extend the effectiveness of HRAs.

From our findings, confidence seems to emerge as a key construct in the satisfaction with the HRA. Also, the confusion or disagreement of the GPs about the feedback reported by participants suggested that the evaluated HRA was not optimally embedded in the health care system. Therefore, the HRA should be seamlessly intertwined into the current health care practice. First, this might increase the familiarity of health professionals with the feedback provided by the HRA, which may encourage the use of this feedback by the GP. Second, when the HRA is offered by a health professional itself (eg, GPs), confidence of end-users in the HRA and the provided feedback might increase. Offering the HRA via the GP might also increase initial participation. Colkesen et al showed that one of the most frequently mentioned reasons for not participating in the HRA was that people were already under supervision of a physician [31]. In such cases, the physician can explain why it might still be relevant for them to participate in the HRA.

For developers of HRAs, this study provides insight into implications for HRAs and potential improvements. Transparency of the provided feedback is an important aspect to consider. Furthermore, information regarding the aim and content of the HRA should be clear and accurate, to prevent unrealistic expectations among end-users. For example, end-users should know the benefits and limitations of the HRA but also need to be educated about the advantages and disadvantages of screening in general.

Unanswered Questions and Future Research

Our study evaluated end-user satisfaction with a web-based HRA with tailored feedback and showed that respondents with elevated health risks were less satisfied with the comprehensibility of the provided feedback. Remarks showed that these respondents had difficulties applying the lifestyle advice and reported anxiety as a result of the feedback of the HRA. As lifestyle behavior change and other health-related actions will be most needed in this group of respondents, further research is needed to provide an insight into how the feedback can be improved and into the relation between respondent's satisfaction rankings and the actual health-related behavior. For instance, it is possible that respondents are more critical because of an increased awareness about their health. Future research should also focus on the relationship between confidence in the HRA and satisfaction rankings.

Furthermore, before and during use of an HRA, end-users will have various expectations with regard to an HRA's output and usefulness. This will influence their willingness to use and their eventual satisfaction with the system in question. For example, another important reason for non-participation in this HRA was that many potential users felt they were healthy and therefore not in need of screening [31]. Providing accurate information tailored to this particular group of users might increase participation. Furthermore, our study showed that during use, the expectations of participants did not match with the HRA's output, which influenced their satisfaction. Although these expectations with regard to the outcome and usefulness of a HRA might vary, it will probably be possible to distinguish clusters of (potential) users with similar expectations. If these are known, tailored HRA solutions and information about HRAs that increase participation and eventual satisfaction can be developed.

Conclusion

In general, respondents were satisfied with the web-based HRA with tailored feedback. However, information about elevated health risks was not always received well. Our study showed that respondents with elevated health risks were less satisfied with the comprehensibility of the feedback given by the HRA and were suspicious of the feedback. Furthermore, there was a lack of confidence in the HRA, as respondents felt they could not exert control over the health information and therefore they perceived that the HRA did not reflect their health status correctly. As a result, several respondents remarked they would prefer a personal face to face consult with a health professional. This suggests a need for a more responsive and flexible approach.

An important aspect in improving the web-based HRA is to increase the transparency of the generation of the provided feedback for both health professionals as well as participants. In particular, it is necessary to improve the embedding of the HRA in current health care practice. This might result in increased use and confidence in the HRA.

Acknowledgments

This study was funded by The Netherlands Organisation for Health Research and Development (ZonMw) through an unrestricted research grant for two projects: MyCARDSS, self-regulation services for cardiovascular risk management (#300020008) and The effectiveness of health check-ups for promoting health behavior (#203310006). We would like to thank all of the employees at the study worksite for participation.

Authors' Contributions

MLEB and NP were the study's principal investigators and contributed to interpreting the data. EC designed and wrote the original proposal, with support from RK. EC also carried out the data collection. MN was responsible for preparing the data for the analysis. SV and EL performed the data analysis and drafted the manuscript, which was reviewed by the other authors. All authors read and approved the final manuscript.

Conflicts of Interest

Dr. Roderik A Kraaijenhagen is medical director and co-owner of NIPED. The program that was studied was developed by this company, which currently markets the program in the Netherlands. Maurice AJ Niessen was employed full-time at NIPED. NIPED provided a PhD grant for Ersen B Colkesen. Eva K Laan was employed part-time at NIPED (four hours a week) and received a salary during this study.

Multimedia Appendix 1

Example of a personal health risk profile and lifestyle behavior advice.

[PDF File (Adobe PDF File), 121KB - jmir_v14i5e140_app1.pdf]

Multimedia Appendix 2

Simplified example of the generation of the feedback provided by the HRA.

[PDF File (Adobe PDF File), 175KB - jmir_v14i5e140_app2.pdf]

Multimedia Appendix 3

Characteristics and interpretations of the codebook used to structure the data.

[PDF File (Adobe PDF File), 126KB - jmir_v14i5e140_app3.pdf]

References

1. Soler RE, Leeks KD, Razi S, Hopkins DP, Griffith M, Aten A, Task Force on Community Preventive Services. A systematic review of selected interventions for worksite health promotion. The assessment of health risks with feedback. Am J Prev Med 2010 Feb;38(2 Suppl):S237-S262. [doi: 10.1016/j.amepre.2009.10.030] [Medline: 20117610]
2. Health Education Authority. Action on Health at Work. London, UK: HEA 1993.
3. Wilson MG, Holman PB, Hammock A. A comprehensive review of the effects of worksite health promotion on health-related outcomes. Am J Health Promot 1996;10(6):429-435. [Medline: 10163309]
4. Faculty of Public Health Medicine Committee on Health Promotion. Health Promotion in the Workplace. Guidelines for Health Promotion No. 40. London, UK: Royal College of Physicians; 1995.
5. Brug J, Oenema A, Campbell M. Past, present, and future of computer-tailored nutrition education. Am J Clin Nutr 2003 Apr;77(4 Suppl):1028S-1034S [FREE Full text] [Medline: 12663313]
6. Kreuter MW, Bull FC, Clark EM, Oswald DL. Understanding how people process health information: a comparison of tailored and nontailored weight-loss materials. Health Psychol 1999 Sep;18(5):487-494. [Medline: 10519465]
7. Franklin PD, Rosenbaum PF, Carey MP, Roizen MF. Using sequential e-mail messages to promote health behaviors: evidence of feasibility and reach in a worksite sample. J Med Internet Res 2006;8(1):e3 [FREE Full text] [doi: 10.2196/jmir.8.1.e3] [Medline: 16585028]
8. 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]
9. Aoki N, Dunn K, Johnson-Throop KA, Turley JP. Outcomes and methods in telemedicine evaluation. Telemed J E Health 2003;9(4):393-401. [doi: 10.1089/153056203772744734] [Medline: 14980098]
10. Mair F, Whitten P. Systematic review of studies of patient satisfaction with telemedicine. BMJ 2000 Jun 3;320(7248):1517-1520. [Medline: 10834899]
11. Chung M, Melnyk P, Blue D, Renaud D, Breton MC. Worksite health promotion: the value of the Tune Up Your Heart program. Popul Health Manag 2009 Dec;12(6):297-304. [doi: 10.1089/pop.2008.0044] [Medline: 20038255]
12. Nguyen HQ, Carrieri-Kohlman V, Rankin SH, Slaughter R, Stulbarg MS. Internet-based patient education and support interventions: a review of evaluation studies and directions for future research. Comput Biol Med 2004 Mar;34(2):95-112. [doi: 10.1016/S0010-4825(03)00046-5] [Medline: 14972630]
13. Renaud L, Kishchuk N, Juneau M, Nigam A, Téreault K, Leblanc MC. Implementation and outcomes of a comprehensive worksite health promotion program. Can J Public Health 2008;99(1):73-77. [Medline: 18435397]
14. Liau SY, Hassali MA, Shafie AA, Ibrahim MI. Assessing quality of a worksite health promotion programme from participants' views: findings from a qualitative study in Malaysia. Health Expect 2011 Nov 3. [doi: 10.1111/j.1369-7625.2011.00742.x] [Medline: 22050457]
15. Pascoe GC. Patient satisfaction in primary health care: a literature review and analysis. Eval Program Plann 1983;6(3-4):185-210. [Medline: 10299618]
16. Colkesen EB, Niessen MA, Peek N, Vosbergen S, Kraaijenhagen RA, van Kalken CK, et al. Initiation of health-behaviour change among employees participating in a web-based health risk assessment with tailored feedback. J Occup Med Toxicol 2011;6:5 [FREE Full text] [doi: 10.1186/1745-6673-6-5] [Medline: 21388524]
17. Marshall AL. Challenges and opportunities for promoting physical activity in the workplace. J Sci Med Sport 2004 Apr;7(1 Suppl):60-66. [Medline: 15214603]
18. Wixom BH, Todd PA. A Theoretical Integration of User Satisfaction and Technology Acceptance. Information Systems Research 2005;16(1):85. [doi: 10.1287/isre.1050.0042]
19. Johnson B, Onwuegbuzie, AJ. Mixed Methods Research: A Research Paradigm Whose Time Has Come. Educational Researcher 2004 Oct;33(7):14-26. [doi: 10.3102/0013189X033007014]
20. Laan EK, Kraaijenhagen RA, Peek N, Busschers WB, Deutekom M, Bossuyt PM, et al. Effectiveness of a web-based health risk assessment with individually-tailored feedback on lifestyle behaviour: study protocol. BMC Public Health 2012;12:200 [FREE Full text] [doi: 10.1186/1471-2458-12-200] [Medline: 22429308]
21. Burgers JS, Simoons ML, Hoes AW, Stehouwer CD, Stalman WA. [Guideline 'Cardiovascular Risk Management']. Ned Tijdschr Geneeskd 2007 May 12;151(19):1068-1074. [Medline: 17552415]
22. Graham I, Atar D, Borch-Johnsen K, Boysen G, Burell G, Cifkova R, European Society of Cardiology (ESC) Committee for Practice Guidelines (CPG). European guidelines on cardiovascular disease prevention in clinical practice: executive summary: Fourth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice. Eur Heart J 2007 Oct;28(19):2375-2414 [FREE Full text] [doi: 10.1093/eurheartj/ehm316] [Medline: 17726041]
23. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [Medline: 10170434]
24. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ 2005 Apr 2;330(7494):765. [doi: 10.1136/bmj.38398.500764.8F] [Medline: 15767266]
25. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005 Oct;59(1):21-30. [doi: 10.1016/j.pec.2004.09.008] [Medline: 16198215]

26. Wolff LS, Massett HA, Weber D, Mockenhaupt RE, Hassmiller S, Maibach EW. Opportunities and barriers to disease prevention counseling in the primary care setting: a multisite qualitative study with US health consumers. Health Promot Int 2010 Sep;25(3):265-276 [FREE Full text] [doi: 10.1093/heapro/daq030] [Medline: 20413404]
27. Asch DA, Jedrziewski MK, Christakis NA. Response rates to mail surveys published in medical journals. J Clin Epidemiol 1997 Oct;50(10):1129-1136. [Medline: 9368521]
28. Sitzia J, Wood N. Response rate in patient satisfaction research: an analysis of 210 published studies. Int J Qual Health Care 1998 Aug;10(4):311-317 [FREE Full text] [Medline: 9835247]
29. Jones R, Pitt N. Health surveys in the workplace: comparison of postal, email and World Wide Web methods. Occup Med (Lond) 1999 Nov;49(8):556-558 [FREE Full text] [Medline: 10658310]
30. Rogers EM. Diffusion of innovations. New York: Free Press; 2003.
31. Colkesen EB, Kraaijenhagen RA, Frings-Dresen MH, Sluiter JK, van Kalken CK, Tijssen JG, et al. Participation in a workplace web-based health risk assessment program. Occup Med (Lond) 2011 Dec;61(8):586-589. [doi: 10.1093/occmed/kqr122] [Medline: 21891778]

Edited by G Eysenbach; submitted 30.01.12; peer-reviewed by L Burdorf, S Langrial; comments to author 25.04.12; revised version received 14.06.12; accepted 03.08.12; published 30.10.12.

Please cite as:
Vosbergen S, Laan EK, Colkesen EB, Niessen MAJ, Kraaijenhagen RA, Essink-Bot ML, Peek N
Evaluation of End-User Satisfaction Among Employees Participating in a Web-based Health Risk Assessment With Tailored Feedback
J Med Internet Res 2012;14(5):e140
URL: http://www.jmir.org/2012/5/e140/
doi:10.2196/jmir.2067
PMID:23111097

Original Paper

eHealth Program to Empower Patients in Returning to Normal Activities and Work After Gynecological Surgery: Intervention Mapping as a Useful Method for Development

Antonie Vonk Noordegraaf[1], MD; Judith A.F Huirne[1], MD, PhD; Carina A Pittens[2], MSc; Willem van Mechelen[2], MD, PhD; Jacqueline E.W Broerse[2], PhD; Hans A.M Brölmann[1], MD, PhD; Johannes R Anema[3], MD, PhD

[1]VU University Medical Center, Department of Obstetrics and Gynaecology, Amsterdam, Netherlands

[2]VU University, Athena Institute for Research on Innovation and Communication in Health and Life Sciences, Amsterdam, Netherlands

[3]VU University Medical Center, Department of Public and Occupational Health, Amsterdam, Netherlands

Corresponding Author:
Antonie Vonk Noordegraaf, MD
VU University Medical Center
Department of Obstetrics and Gynaecology
De Boelelaan 1117, 1081 HV Amsterdam
Amsterdam,
Netherlands
Phone: 31 204446065
Fax: 31 204448387
Email: t.vonknoordegraaf@vumc.nl

Abstract

Background: Full recovery after gynecological surgery takes much longer than expected regardless of surgical technique or the level of invasiveness. After discharge, detailed convalescence recommendations are not provided to patients typically, and postoperative care is fragmented, poorly coordinated, and given only on demand. For patients, this contributes to irrational beliefs and avoidance of resumption of activities and can result in a prolonged sick leave.

Objective: To develop an eHealth intervention that empowers gynecological patients during the perioperative period to obtain timely return to work (RTW) and prevent work disability.

Methods: The intervention mapping (IM) protocol was used to develop the eHealth intervention. A literature search about behavioral and environmental conditions of prolonged sick leave and delayed RTW in patients was performed. Patients' needs, attitudes, and beliefs regarding postoperative recovery and resumption of work were identified through focus group discussions. Additionally, a literature search was performed to obtain determinants, methods, and strategies for the development of a suitable interactive eHealth intervention to empower patients to return to normal activities after gynecological surgery, including work. Finally, the eHealth intervention was evaluated by focus group participants, medical doctors, and eHealth specialists through questionnaires.

Results: Twenty-one patients participated in the focus group discussions. Sufficient, uniform, and tailored information regarding surgical procedures, complications, and resumption of activities and work were considered most essential. Knowing who to contact in case of mental or physical complaints, and counseling and tools for work reintegration were also considered important. Finally, opportunities to exchange experiences with other patients were a major issue. Considering the determinants of the Attitude–Social influence–self-Efficacy (ASE) model, various strategies based on a combination of theory and evidence were used, resulting in an eHealth intervention with different interactive functionalities including tailored convalescence recommendations and a video to communicate the most common pitfalls during the perioperative period to patients and employers. Fifteen patients in the focus groups, 11 physicians, and 3 eHealth specialists suggested points for improvement to optimize the usability of the eHealth intervention and judged it an approachable, appropriate, and attractive eHealth intervention to empower gynecological patients.

Conclusions: The IM protocol was a useful method to develop an eHealth intervention based on both theory and evidence. All patients and stakeholders judged the eHealth intervention to be a promising tool to empower gynecological patients during the perioperative period and to help them to return to normal activities and work.

(J Med Internet Res 2012;14(5):e124) doi:10.2196/jmir.1915

KEYWORDS

eHealth intervention; interactive website; Internet; patient empowerment; intervention mapping; gynecological surgery; hysterectomy; laparoscopic adnexal surgery; return to work

Introduction

Following gynecological surgery, full recovery (including returning to work) takes much longer than expected regardless of surgical technique or the level of invasiveness [1-3]. In two prospective observational studies in the Netherlands, median sick leave after gynecological procedures for benign conditions was 8 weeks [4]. Prolonged absence from work often results in a lack of social structure and meaningful activities [5,6] and can result in work disability, poorer general health, and increased risk of mental health problems [7,8]. As a result, long periods of sick leave contribute to a reduced quality of life and induce unnecessary yet substantial costs for society through lost working hours, physician consultations, medication treatment, and higher hospital admission rates [9,10].

To reduce health care costs, there is an increasing trend to limit the duration of in-hospital care and to transfer postoperative care to outpatient and primary care [11-13]. However, after discharge, gynecological care is given only on demand, detailed recommendations about resumption of work activities are not provided typically [1,14], and patients often do not know who to contact for support in case of postoperative complaints. Furthermore, family physicians frequently do not give advice about resumption of work activities and occupational or insurance physicians are only consulted if patients have paid work and these consultations take place relatively late in the course of sick leave because of legislation [15-17]. This contributes to irrational beliefs and avoidance of resumption of activities that can result in a prolonged sick leave [18].

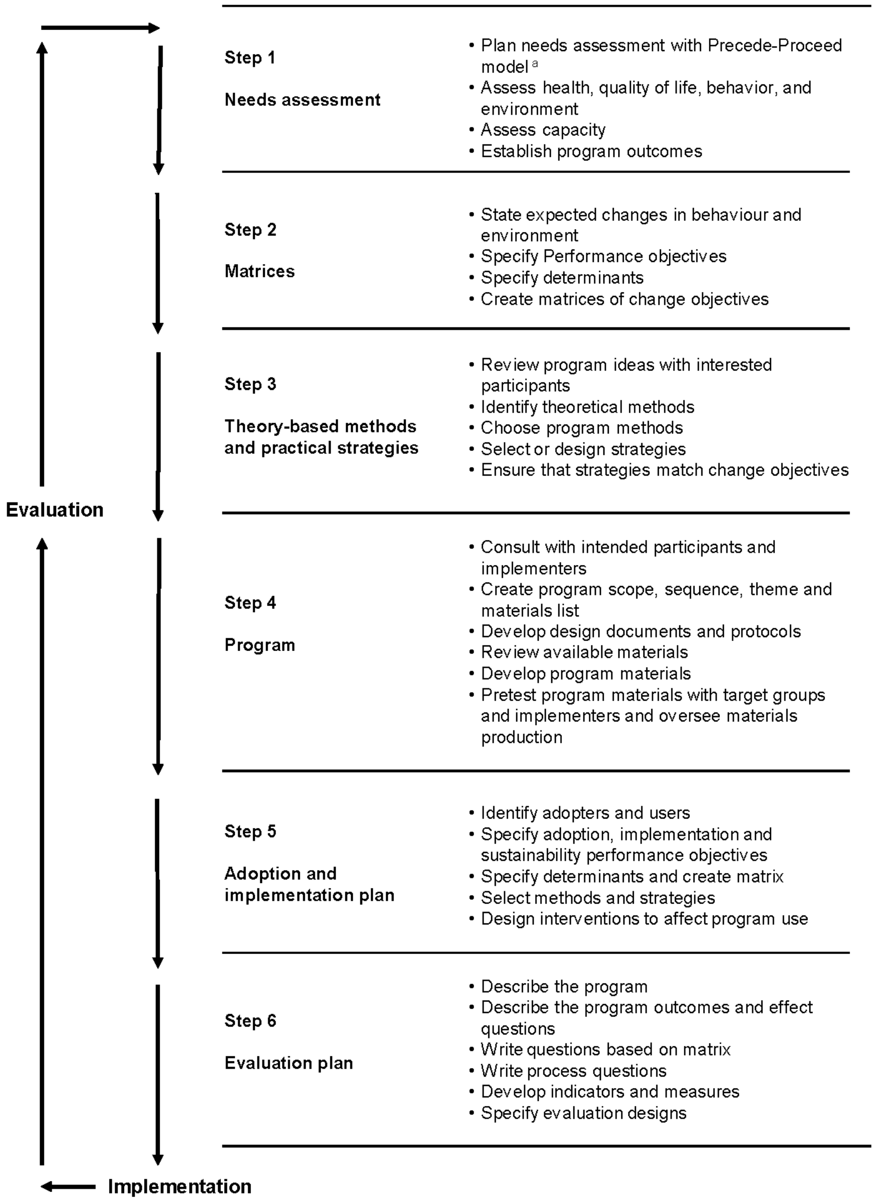
Because a significant part of the recovery and return to work (RTW) problems of patients seem to be caused by counseling and communication deficiencies, the starting point of this study was to identify these specific problems. Many interventions aiming to improve communication with and counseling of patients have focused only on health care professionals [19,20]. However, to improve communication and health outcomes, empowering patients to actively participate in their consultations with physicians is also important [21,22]. Patient empowerment refers to the enhanced ability of patients to actively understand and influence their own health status [23]. It focuses on control in patients' experiences of health, disease, and illness, as well as the roles of health care organizations, communities, and the broader health care system [24,25]. eHealth interventions seem to be a promising way to empower patients by providing personalized education (eg, detailed recommendations on resumption of work activities) and enhancing interaction between health consumers and professionals [26-28]. Patients become more actively engaged in their own state of health (eg, are aware which complications need additional consultation) and the communication between patient and health care provider becomes more efficient and equal [29-31]. Tailored eHealth interventions are more intensively used [32,33] and have a greater impact on people's behavior [33-36] than generic materials, and they provide the opportunity to deliver information to a large audience [37] at any time and with lower costs [34,38]. An important condition for a successful eHealth intervention is adequate implementation [39,40].

Therefore, the objective of this study was to develop a feasible and generally accepted eHealth intervention that empowers gynecological patients during the perioperative period about returning to normal activities and work, to obtain timely RTW, and prevent work disability. To develop this intervention, we used the intervention mapping (IM) protocol [41,42], which has been shown to be a suitable systematic and scientifically accepted method for the development and implementation of a wide range of eHealth [43-46] and RTW [47,48] interventions based on theory and stakeholders' (including patients') involvement.

Methods

Intervention mapping was used to tailor the eHealth intervention to patients' needs and wishes, taking into account the clinical evidence of the main determinants that influence patients' behavior to reach timely RTW. The project group consisted of 1 research physician, 2 gynecologists, and 2 occupational physicians. Although it is not a theoretical or conceptual framework, IM is a systematic description of a logical planning process involving 6 steps: (1) performing a needs assessment; (2) defining program objectives; (3) selection of theory-based methods and practical strategies; (4) design of the intervention program; (5) development of a plan for adoption and implementation; and (6) design of an evaluation plan (Figure 1). The iterative character of IM enables the intervention to be based on a combination of theory and evidence, which maximizes the applicability for the target population and minimizes the risk of choosing the wrong theory behind the intervention (theory failure) or of poor adoption of the intervention (program failure) [49].

Figure 1. Intervention mapping process.



[a] The Precede-Proceed model structures the analyses and correlation of quality of life, health, behavior, and environmental factors of a certain population [50].

Step 1: Needs Assessment

In needs assessment, the discrepancy between the current and the desired situation in a given group of people is studied. The needs assessment was structured by the Precede-Proceed model (PRECEDE: predisposing, reinforcing, and enabling constructs in educational diagnosis and evaluation; PROCEED: policy, regulatory, and organizational constructs in educational and environmental development), which analyzes and correlates quality of life, health, behavior, and environmental factors in a certain population [50]. The current situation has shown a large discrepancy between expected duration of physical recovery and actual RTW after gynecological surgery (even laparoscopic), whereas there is strong evidence that long periods of sick leave can result in poorer general health, increased risk of mental health problems and work disability, and induces unnecessary costs for society [2,8] The most frequently performed gynecological surgical procedures with a considerable postoperative effect on recovery and RTW (accounting for more than 17,500 procedures in the Netherlands per year) are hysterectomy (abdominal, vaginal, and laparoscopic) or laparoscopic adnexal surgery on benign indication [51]. Because approximately 67% of women aged 25-65 are in the workforce, these large numbers of surgical procedures have a great impact on absenteeism [52]. Therefore, patients who underwent these types of surgical procedures were chosen as the target group for this intervention.

To clarify and find possible explanations for prolonged sick leave, a literature search in PubMed regarding behavioral and environmental conditions of prolonged sick leave and delayed RTW in the target group was performed.

The focus group technique was considered the most suitable supplement to the literature search for identification of patient’s needs, attitudes, and beliefs regarding postoperative recovery and resumption of work. In addition to supplementing the results of the literature search, it was assumed that focus group discussions would align the results to the Dutch context and give more insight into specific content requirements of the prospective eHealth intervention that could be used during the development process. The participatory technique of focus group discussions is widely used and scientifically accepted to gain insight into public views and needs through group interaction [53,54].

Participants for the focus group discussions were recruited from the patient files of the VU University Medical Center, an academic hospital in the Netherlands. To mirror the intended target group, inclusion criteria for participation in the focus group discussions were age 18-65, a history of a laparoscopic adnexal surgery and/or hysterectomy on benign indication in 2008, and a job (paid or unpaid) of at least 8 hours per week. To create homogeneity within the focus groups but heterogeneity among the groups, the patients were recruited by means of purpose sampling into 3 groups: fast RTW, intermediate RTW, and delayed RTW. All of the participants had already returned to work after surgery (range 1-36 weeks).

The aim of the focus group discussions was to identify patients’ needs regarding perioperative care and counseling in resuming normal and work activities. In addition, patients were specifically asked for the important requirements of a useful eHealth intervention. The identification of patients’ needs and requirements occurred in 3 different steps:

1. Identifying and prioritizing patients’ perceived shortcomings in and difficulties with received perioperative care and counseling in resumption of normal and work activities.

2. Inventory of possible solutions and improvements to overcome these shortcomings and difficulties, starting with the highest prioritized bottlenecks.

3. Brainstorming about favorable content, requirements, and specific tools that should be incorporated into an eHealth intervention that aims to empower patients during the perioperative period and resumption of work activities.

The focus group discussions were all recorded and transferred into verbatim transcripts that were analyzed by open coding using the ATLAS.ti software [55].

A detailed process evaluation of the focus group discussions will be published in a separate paper [Pittens et al, unpublished data, 2012]. The study design and procedures of the focus group discussions were approved by the Medical Ethics Committee of the VU University Medical Center (2009/42, February 9, 2009). Participants signed a privacy agreement to declare voluntary participation, to give permission for processing the information for the development of an intervention (such as an eHealth intervention), and to exclude transmittal of information to others.

As a starting point for the development of the intervention, the products of this first step were the main behavioral and environmental conditions of the chosen target group contributing to prolonged sick leave.

Step 2: Matrices

The purpose of this step was to transform the identified behaviors and environmental conditions causing prolonged sick leave into behaviors and conditions that prevent a prolonged sick leave. To achieve this, performance objectives were formulated. Performance objectives describe in detail patients’ behavioral and environmental outcomes that are necessary for patients to reach the formulated behavior objective of “timely RTW.”

To select a suitable theoretical framework to reach the performance objectives, a literature search regarding main determinants of recovery and RTW was performed in PubMed. To elucidate, a suitable theoretical framework provides appropriate determinants that could be influenced to reach the behavior objective. Based on this framework, the performance objectives of the target group were elaborated into matrices with change objectives, explaining how patients and their environment will change as a result of the eHealth intervention to reach the behavior objective.

Step 3: Theory-based Methods and Practical Strategies

In this step, theoretical methods and practical strategies to address the change objectives were searched for and applied. Research has shown that the effectiveness of interventions to change behavior can be increased by the use of theory-based methods [56]. A theory-based method is a method derived from theory and research that describes a process that influences changes in determinants of behavior and environmental conditions. A practical strategy is a technique for the application of the theory-based method in ways that fit the target group and the context in which the eHealth intervention will be applied. The required theoretical framework, theory-based methods, and translation into practical strategies were determined based on the book that describes the IM approach [42], a literature search in the PubMed database, the focus group discussions, and a brainstorm session of the researchers.

Step 4: Program Plan and Design of the Intervention

During this step, information obtained in previous steps was translated into specific tailored tools and information to empower gynecological patients by the eHealth intervention. Furthermore, to obtain evidence-based information and instruments necessary to fulfill patients' needs, additional research was performed.

To verify if the eHealth intervention matched with the main target group and fitted the expectations of gynecologists, family physicians, and occupational physicians, the first concept version was evaluated by focus group participants (n = 21), physicians (n = 22), eHealth specialists (n = 3), and a representative of a patient organization (n = 1) through questionnaires. The eHealth intervention was scored on 8 main areas used to describe how the intervention functions, empowers, and can be modified to provide the best behavior change to obtain timely RTW and prevent work disability. The 8 areas included: appearance, behavior prescriptions, burdens of using the website, content, delivery, message, participation, and assessment and tailoring. Ritterband et al [57] describe these areas in detail. This model is meant to ground Internet intervention research within a scientific framework. The 8 different areas were covered in the evaluation questionnaires with 23 unique open- and close-ended questions (Appendices 4 and 5). In addition, participants were also encouraged to propose recommendations. The results of the evaluation were used to optimize the design and usability of the eHealth intervention, which resulted in the final version.

Step 5: Design of an Implementation Plan

The focus of Step 5 is adoption of the intervention by the patients and relevant stakeholders, and the development of an implementation plan. With the input of patients and stakeholders during previous steps, the researchers identified facilitating factors and barriers regarding adoption and implementation of the eHealth intervention. With this information, an implementation plan to enable an extensive evaluation of the intervention was developed and an appropriate linkage system for future implementation was composed.

Step 6: Design of an Evaluation Plan

During Step 6, the main objective of this study (ie, to develop a feasible and generally accepted eHealth intervention that empowers gynecological patients during the perioperative period into returning to normal activities and work, to obtain timely RTW, and prevent work disability) was used to compose an evaluation plan. Although the eHealth intervention was based on both theory and evidence and was developed in collaboration with the main target group and relevant stakeholders, its adoption, barriers for usage, and implementation possibilities still had to be evaluated in daily practice. In addition, the effectiveness of this eHealth intervention regarding a timely RTW to prevent work disability had to be investigated. Therefore, the project group approached 7 gynecology practices (1 university-based and 6 hospital-based) about participation in the evaluation of this intervention through implementation of the eHealth intervention as a supplement to the standard perioperative care given at their hospital. In addition, the project group formulated inclusion and exclusion criteria for patients to participate in the study and developed appropriate outcome measures to evaluate the intervention's effectiveness, adoption, usage, and implementation. Furthermore, a logistic plan to recruit patients and involve participating health care providers was developed.

Results

Step 1: Needs Assessment

Literature

The literature search revealed that most women extend their sick leave beyond the recommended period on their own initiative [2]. Patients with delays in RTW reported pain/discomfort, feelings of fear, and infections as delaying factors [1]. Those who reported multiple delaying factors reported a variety of combinations that included feelings of fear, anxiety, depression, and differences in employer expectations [1]. Recovery and RTW time is shorter when the patient receives clear and few restrictions that are not too overly cautious at discharge, when the patient has been provided with RTW advice, or when the patient feels urgency to RTW [1,3,58]. Other important environmental conditions for prolonged sick leave and RTW of patients appeared to be the substantial variation in convalescence recommendations given by gynecologists, family physicians, and occupational physicians [58,59]. Their recommendations are not related to the most successful return to normal and RTW activities or the risks of complications [58]. In addition, a lack of clarity regarding absence duration can provide an obstacle for employers and employees who are keen and willing to establish earlier rehabilitation programs, but would not wish to go against the advice of health care providers [59].

Focus Group Discussions

Out of 105 invited patients, 38 met our inclusion criteria and were willing to participate in the focus group discussions. On the basis of availability on the selected dates for the focus group discussions, 31 patients were assigned to 3 focus groups. Of these patients, 21 were present at the meetings and participated in the focus group discussions (7 patients per meeting). A process evaluation of the focus group discussions will be published in detail elsewhere [Pittens et al, unpublished data, 2012].

Starting with the first aim of the focus group discussions, the most important reported shortcomings and difficulties of currently provided perioperative gynecological and reintegration care were (in random order):

1. Insufficient or no information about the surgical procedure itself, the logistical process in the hospital from admission to discharge, detailed resumption of work activities after the surgical procedure, and the possible consequences of the surgery (physical and mental).

2. Inconsistency of convalescence recommendations given by gynecologists, family physicians, and occupational physicians.

3. Lack of written instructions on resumption of work activities, tailored to individual conditions and work, and consequently insufficient information and instructions given to relatives.

4. Insecurity with respect to physical or mental postoperative symptoms, complications, or delayed recovery. What to do and whom to contact?

5. Poor communication among gynecologists, family physicians, and occupational physicians resulting in inadequate transfer of information about the procedure and one another's treatments.

6. Limited or inadequate guidance by occupational physicians because of a lack of knowledge about different types of surgery and corresponding recovery times. Patients reported experiences of occupational physicians forcing the patient to RTW too early or slowing down the RTW process.

7. Difficulties with work reintegration because of insufficient involvement and understanding of the employee/employer during the perioperative and reintegration period.

8. Inability of patients to discuss the perioperative period and reintegration process at work (with employer and colleagues).

9. Lack of a reintegration plan before the surgery.

10. Few opportunities to contact other patients to exchange experiences.

In general, patients mentioned that when they were unsatisfied with the information or counseling given by their doctors and nurses, they asked family and friends who had undergone surgery about their experiences. However, this led to unrealistic expectations because of different types of surgical procedures and techniques, and the fact that recovery is affected by individual conditions and circumstances.

In the second part of the focus group discussions, the patients brought up many possible solutions and improvements to overcome the mentioned shortcomings and difficulties that were processed into performance objectives during Step 2.

Requirements, content, and specific tools that should be incorporated into an eHealth intervention to improve empowerment during the perioperative period may be summarized as follows:

1. Reliable detailed and personalized information about mentioned shortcomings and difficulties in the information supply. Pictures and videos were considered an accessible supplement to transfer this information.

2. Tools for communication with other patients, employers, gynecologists, occupational physicians, and family physicians.

3. Functionalities to develop a personalized reintegration plan.

With the results of the literature search and focus group discussions, the project group concluded that the main determinants of patients' behavior regarding prolonged sick leave are: (1) inadequate knowledge of important information about the surgery, recovery, and RTW; (2) tendency to extend their sick leave beyond the recommended period; (3) insecurity about postoperative symptoms, complications, and delayed recovery without knowing where to receive appropriate help; (4) lack of skills to compose a work-reintegration plan and to identify possible barriers for RTW; and (5) lack of knowledge about the opportunity to develop and discuss a work reintegration plan before surgery with the employer and an occupational physician. In addition, important environmental conditions of patients' behavior are considered to be: (1) inconsistency and lack of clarity in convalescence recommendations given by gynecologists, family physicians, and occupational physicians; (2) lack of communication among gynecologists, family physicians, and occupational physicians; (3) lack of clarity from health care providers about who to contact in case of postoperative complaints; (4) lack of initiative of the employer and/or occupational physician to develop and discuss a work-reintegration plan before surgery with the employee; and (5) lack of involvement of the employer and occupational physician during the perioperative and reintegration period.

Step 2: Matrices

In total, 12 performance objectives derived from the main behavior objective were formulated (see Table 1).

Table 1. Performance objectives to empower gynecological patients during the perioperative period and return to normal activities and work to obtain timely RTW and prevent work disability.

Who	Performance objectives
Patients	Acquaint themselves with important information including: realistic detailed convalescence recommendations regarding RTW activities; the importance of timely and gradual resumption of work activities after the surgical procedure; the technical aspects of surgical procedures; the admission process at the hospital; the kind of anesthesia that will be used during surgery; main complications that could happen during and after surgery; symptoms that can be expected after surgery (eg, vaginal blood loss and intestinal complaints); the cosmetic consequences of surgery; main psychological consequences of hysterectomy or adnexal surgery; telephone numbers of experts (eg, gynecologist, social workers, and homecare services); what to do and who to contact in case of physical or mental postoperative complaints or delayed recovery; and the risks of work disability after surgery.
	Do not extent their sick leave period beyond recommended period on own initiative.
	Develop a work-reintegration plan.
	Discuss their personalized work-reintegration plan with their employer and/or occupational physician.
	Identify possible barriers for a safe and appropriate RTW.
	Exchange experiences with other patients who underwent the same surgery.
	Receive answers to individual questions and uncertainties about recovery and RTW.
Gynecologists and family physicians	Acquaint themselves with uniform, detailed convalescence recommendations for their patients.
Occupational physicians	Acquaint themselves with detailed convalescence recommendations for their patients.
	Provide the opportunity to develop a work-reintegration plan before surgery.
Employers	Provide the opportunity to develop a work-reintegration plan before surgery.
	Discuss the personalized work-reintegration plan composed by their employees.
	Show involvement with their employee during the perioperative and reintegration period.

In addition to performance objectives for patients, there were also performance objectives formulated for gynecologists, family physicians, occupational physicians, and employers (Table 1). Nevertheless, the primary focus during the next steps of the IM protocol (the development of the eHealth intervention) will be on patients. Ideally, for each group (patients, gynecologists, family physicians, occupational physicians, and employers) an intervention should be developed specific to their needs, wishes, and behavior outcomes to minimize the risks of theory and/or program failure. However, a balance had to be found between the ideal situation and what was within reach of this study. Secondly, the performance objectives of gynecologists, family physicians, occupational physicians, and employers could be considered external determinants of patients' behavior. These determinants can either be influenced by the patients or the patients can learn these skills through the intervention and how to handle them adequately. Finally, the performance objectives of gynecologists, family physicians, occupational physicians, and employers are relatively simple objectives to reach. The researchers are convinced that the main part of these objectives can be reached through making agreements with gynecologists, family physicians, occupational physicians, and employers and by involving them in the evaluation and implementation plan (IM Steps 4-6) without specifying determinants of their behavior and applying specific theoretical methods and strategies for them.

The literature search showed that the main determinants of recovery and return to normal activities and work (in addition to the physical condition of the patient, level of invasiveness of the surgical procedures, and related complications) are the patients' attitude, social influence, and self-efficacy [60-65]. In addition, skills, barriers, and facilitators are important factors that influence RTW [66-68]. For these reasons, the Attitude-Social influence-self-Efficacy (ASE) model [69,70], adapted for recovery and return to normal and work activities, was used to affect behavior of patients (see Appendix 1). The ASE model is comparable to the theory of planned behavior [71], which describes the relation between attitude and behavior. The modified ASE model describes that the behavior of a patient after surgery regarding recovery and return to normal and work activities is determined by attitudinal beliefs, social influence, and self-efficacy beliefs, and is influenced by skills, barriers, and resources. The ASE model was used to create matrices with change objectives. To fill out the matrices, available literature regarding the performance objectives and determinants was studied together with the results of the needs assessment and expertise of the project group. Appendix 2 presents an example of the change objective "Patients develop a work-reintegration plan."

Step 3: Theory-based Methods and Practical Strategies

Numerous practical methods and suitable strategies to affect all formulated determinants were identified and used for the development of tools and materials of the eHealth intervention. Appendix 3 presents some examples of these methods with preconditions for the method necessary for it to succeed [72] and final tool/materials of the eHealth intervention. References and footnotes explain the source and development process of each method, strategy, and tools/materials.

Step 4: Program Plan/Design of the Intervention

With the knowledge obtained in the previous steps of the IM protocol, the project group convened at several meetings to invent various appropriate tools for the eHealth intervention. A website producer specializing in eHealth interventions and a screenwriter were consulted at some of the meetings. In addition, an experienced gynecologist outside the project team was consulted to judge the medical content of one of the tools.

In close collaboration with the website producer, the eHealth intervention was developed with MODX, an open-source hypertext preprocessor (PHP) Web application framework with a capable built-in content management system (CMS). The Internet address of the eHealth intervention is “http://www.ikherstel.nl/www.ikherstel.nl,” which means, “I am recovering” (Appendices 6-10) [73]. The eHealth intervention was developed with special attention to colors, layout, navigation, and readability to create confidence and user-friendliness. For the patient, it consists of two main sections: an Action List to assist in resumption of activities and a central home page. Gynecologists, family physicians, and occupational physicians have access to a different section. Table 2 presents an overview of the tools in the eHealth intervention. For some tools, additional information about the development and functioning is described subsequently.

Table 2. Structure of the eHealth intervention.

Tool	Content	Target
Action list		
Compose a work-reintegration plan	Tool to compose a detailed reintegration plan with adaptations for work if necessary.	Patient, employer, occupational physician
Resume normal activities	Tool to compose detailed advice about when normal (private) activities can be carried out again	Patient, family
Evaluate complications	Estimate severity and consequences of a complication	Patient, gynecologist
Recovery monitor	Monitoring recovery and offering assistance when relevant	Patient
Satisfaction with recommendations	Evaluation and explanation of convalescence recommendations	Patient
Satisfaction with the recovery process	Evaluation of satisfaction with recovery and reintegration process. Provision of advice regarding which care provider(s) to approach to receive appropriate help, when relevant.	Patient
Invite employer	Invite employer for (anonymous) section of the eHealth intervention which includes video and recommendations	Patient, employer
Home page		
Video	Illustrate common pitfalls during the perioperative and reintegration period	Patient, employer, gynecologist
Recommendations for employee	Advice for a successful reintegration	Patient
Recommendations for employer	Advice for appropriate involvement regarding employee during the perioperative and reintegration period	Employer
Frequently asked questions	Extensive list of answers and pictures to most frequently asked questions	Patient
Glossary	Explanation of most frequently used medical terms	Patient
Forum	Ability to interact in public or through private messages with other patients	Patient
Links to other websites	Relevant websites concerning the perioperative and reintegration period	Patient
Section aimed at gynecologists, family physicians, and occupational physicians		
Guidelines	Well-defined convalescence recommendations after hysterectomy and laparoscopic adnexal surgery	Gynecologists, family physicians, occupational physicians
Casuistry	Indications, perioperative course, and recovery regarding hysterectomy or laparoscopic adnexal surgery	Gynecologists, family physicians, occupational physicians
Background information	Specialistic information regarding different kinds of hysterectomy and laparoscopic adnexal surgery	Gynecologists, family physicians, occupational physicians

Action List

When a patient logs onto the eHealth intervention, she will be immediately directed to the Action List. This list consists of different tools developed to target specific determinants, aimed at encouraging return to work activities, coaching patients in case of uncertainties, answering possible questions, prevention of common pitfalls, and improving communication among the patient, care providers, and the employer. An algorithm based on the date of surgery determines the priority in which the different actions should be performed to improve the recovery process. Tools of the action list are:

Composition of a Work-Reintegration Plan

By using this tool, the patient is able to select activities that are required to fulfill her work activities and at what level (eg, lifting

5 kilograms or walking 1 hour.). Consequently, on the basis of the operation date and how the surgery went (input of gynecologist), the eHealth intervention provides the patient with tailored advice about when these activities are thought to be medically safe to resume. The recommendations are based on the results of a modified Delphi study, in which an expert panel of gynecologists, family physicians, and occupational physicians developed detailed multidisciplinary convalescence recommendations for resumption of work activities after hysterectomy and/or laparoscopic adnexal surgery [74]. Moreover, this part of the eHealth intervention provides an overview of potential bottlenecks for reintegration and motivates patients to consider if work adaptations are required temporarily. A printout can be made to discuss the advice with the employer and/or occupational physician to develop an extended reintegration plan.

Resumption of Normal Activities

This functionality guides the patient to compose a detailed tailored plan about the gradual resumption of various daily activities (eg, climbing stairs and vacuum cleaning). Recommendations are based on the results of the modified Delphi study [74]. This tool also evaluates if help is needed for tasks such as housekeeping or taking care of young children. A printout can be made to share with relatives or friends.

Evaluation of Complications

When a complication has occurred, the eHealth intervention carefully determines through a survey which symptoms require additional consultation with care providers or adaptation of the convalescence recommendations. The project group developed the survey and determined which symptoms are severe complications. If the tool is not able to provide recommendations under these circumstances, an email will be sent to inform the gynecologist of the condition of the patient in order to evaluate her symptoms and possible consequences.

Home Page

Video

Because of the influence of modeling behavior on attitude, a video was chosen as the most appropriate medium to deliver an informative message to patients and relevant stakeholders about common pitfalls during the perioperative and reintegration period. The video aims to prevent these problems by stimulating patients and employers to discuss potential problems and to develop a reintegration plan to facilitate and improve reintegration. The experiences of the patients in the focus group discussions were converted into common pitfalls for patients, employers, and health care providers during this period, and a screenwriter processed them into a script for a video showing two cases of a good and bad interaction between a patient and her environment. The screenwriter worked together closely with 3 gynecological patients to make the video geared to the patients' perception of the perioperative and reintegration period.

Recommendations for Employee and Employer

Based on the experiences of the patients in the focus groups, the researchers formulated main recommendations for patients and employers regarding a successful reintegration.

Frequently Asked Questions

Answers to questions brought up during the focus group discussions and those found as main topics in patients' brochures or in discussions of gynecological patients on the Internet were formulated by the researchers (based on the literature and clinical experience) and put into patient leaflets. An experienced gynecologist outside the project team judged all questions and answers on reliability and clarity, and suggested possible adjustments.

Glossary

Based on the literature, an explanation of the most frequently used medical terms was provided by the researchers.

Links to Other Websites

The researchers searched the Internet for the most relevant websites for gynecological patients and made a selection based on relevance, reliability, and clearness of the information.

Section Aimed at Gynecologists, Family Physicians, and Occupational Physicians

Guidelines

Multidisciplinary guidelines with well-defined convalescence recommendations after uncomplicated hysterectomy (laparoscopic supracervical, total laparoscopic/laparoscopic-assisted, vaginal, and abdominal) and laparoscopic adnexal surgery on benign indication are provided. Recommendations are based on a modified Delphi study.

Casuistry

Classic examples of indications for surgery, perioperative course, and recovery after uncomplicated hysterectomy or laparoscopic surgery were developed based on literature and clinical experience of the project group.

Background Information

Elucidation of different types of hysterectomy and laparoscopic adnexal surgery concerning surgical technique, level of invasiveness, and medical consequences were formulated by the researchers.

Test Phase

Fifteen patients, 11 physicians (gynecologists, family physicians, and occupational physicians), 3 eHealth specialists, and 1 representative of a patient organization completed the evaluation form regarding the demo version of the eHealth intervention. Appearance and behavior prescriptions were judged by most as pleasant, conveniently arranged, and helpful. With regard to burdens of using the eHealth intervention, almost all respondents judged the application navigation as clear and the intervention length as appropriate. However, a manual providing an overview of the different tools of the eHealth intervention was found desirable by only one of the respondents. Furthermore, two software incompatibility problems were reported. Concerning the content of the information, the way it was delivered, and the message (source and style), most of the respondents were satisfied and expected that it could empower patients, employers, and physicians. Remarks for improvement were related to supplying more detailed information about the surgery, possible psychological complications after the operation, less complicated sentences, and a more prominent place for the source of the information. Finally, participation of the patient in the treatment and the eHealth intervention's ability to assess and tailor the recommendations to empower patients during the perioperative period and return to work activities were judged as helpful by most of the respondents. There were no suggestions for improvement of these features.

The patients indicated that their input provided during the focus group discussions was recognizably integrated into the intervention. Additionally, almost all patients confirmed that they would recommend the eHealth intervention in the current form to a friend.

Modifications Based on the Test Phase

As described previously, the respondents did not request major revisions of the eHealth intervention and only minor adjustments were proposed. Therefore, none of the original developed tools were removed from the eHealth intervention and no new functionalities were added. Following up on the suggested improvements, a manual with directions for use was added to the eHealth intervention, incompatibility problems with different kinds of software were solved, some information on the eHealth intervention was elaborated on and explained in simplified sentences, and the logo of the university hospital was added in a prominent place on the eHealth intervention. This resulted in the final eHealth intervention that was used to perform a randomized controlled trial (RCT) [75]. Screenshots of the eHealth intervention can be found in Appendices 6-10.

Step 5: Design of an Implementation Plan

In this study, anticipation of adoption and implementation started with the involvement of patients (target group) in all stages of the intervention development and evaluation. Health care providers, occupational physicians, and eHealth specialists participated in the evaluation of the intervention during IM Step 4. In addition, a committee with representatives of the Dutch medical boards of gynecologists, occupational physicians, and family physicians, and a representative of an umbrella patient organization were involved during the development of all steps of the intervention and agreed to stay involved during the final implementation steps of this intervention. Through this committee, a linkage system was created by involving the future users and implementers of the intervention from the start of the intervention development process. Furthermore, an important target of this study was to develop an eHealth intervention that could be used by patients, doctors, and employers without any support to simplify implementation. Evaluation of self-reliant use by patients and important stakeholders was evaluated positively during the test phase of Step 4.

Within the context of a RCT with the eHealth intervention (Step 6), the project group will facilitate its implementation and maintenance. In collaboration with the relevant care providers, the eHealth intervention will be offered as a supplement to standard perioperative care and will involve minimal additional time investment for the care providers. Agreements about usage of the contents of the eHealth intervention will be made with the gynecologists of participating hospitals and the family physicians and occupational physicians of participating patients. Therefore, the main purpose of this step was to create familiarity and support for the eHealth intervention and convalescence recommendations by all prospectively involved users. To reach these purposes for all of the different user groups, information letters will be distributed among patients and care providers. In addition, presentations with background information about the development of the eHealth intervention, its contents, and how to use it will be given to the gynecologists during general teaching meetings at their hospitals. Employers will become familiar with the intervention through invitation for participation by the patients (ie, employees). The eHealth intervention will be primarily used during the period of sick leave after surgery. Therefore, no agreement with the patients' employers to use the eHealth intervention during work hours will be made.

With the information gathered during the process evaluation (Step 6), in collaboration with the committee with representatives of the Dutch medical boards of gynecologists, occupational physicians, and family physicians, and the patient organization, a final implementation plan will be developed. In this plan, medical insurance companies and the Health Care Insurance Board (CVZ) will likely be involved for the final implementation of the eHealth intervention.

Step 6: Evaluation Plan

The evaluation of the eHealth intervention will be performed by a RCT, during which the eHealth intervention will be compared with usual given care at 7 participating medical centers [75]. A power calculation was performed on the primary outcome (sustainable RTW) and showed that a total participation of at least 212 patients, their health care providers, and employers should be the goal. Patients will be recruited to participate in the RCT when they are placed on a waiting list for a hysterectomy or laparoscopic adnexal surgery on benign indication in one of the 7 participating medical centers, are aged 18-65, and they work (either paid or unpaid) for at least 8 hours per week. The main exclusion criteria are malignancy, deep infiltrating endometriosis, concomitant surgical procedures, major comorbidity, sick-listed for more than 2 months, currently in a lawsuit against their employer, and not able to use the Internet or unable to understand the Dutch questionnaires. If a patient participates, the researchers will inform her family physician and occupational physician by letter about the content of the intervention, the group allocation, and what is expected of them regarding the provision of health care. Follow-up will take place approximately 26 weeks after surgery.

Patients willing to participate and who meet the inclusion criteria will be randomized to the intervention or usual care group (control group). Main outcome measures of the RCT are the effectiveness of the eHealth intervention compared to usual care with respect to RTW, general recovery, quality of life, pain intensity, and complications. Part of the RCT will be a process evaluation of the patients, their care providers, and employers in the intervention group. Main outcome measures of the process evaluation are the extent to which the eHealth intervention and convalescence recommendations are used and followed up (compliance); appreciation of the different tools of the eHealth intervention and advice; perceived effectiveness, usage, and implementation barriers; and suggestions for improvement.

The outcome measures will be obtained using questionnaires administered at baseline and at 2, 6, 12, and 26 weeks after surgery. Gynecologists will complete questionnaires 1 day after surgery for each patient and at the end of the study. Employers will be asked to evaluate the eHealth intervention 8 weeks after their employee's surgery.

The study design and procedures of the RCT study were approved by the Medical Ethics Committee of the VU University Medical Center (#2009/218, October 22, 2009).

Discussion

Main Findings

In this study, the IM protocol turned out to be a useful method to develop and tailor an eHealth intervention aimed at the empowerment of gynecological patients during the perioperative period including return to normal activities and work. By using available literature and focus group discussions, it became increasingly clear that to obtain timely RTW and prevent work disability, the intervention should target both behaviors of patients as well as environmental determinants. Performance objectives for obtaining timely RTW and prevention of work disability were formulated and matrices with change objectives, explaining how patients and their environment have to change as a result of the eHealth intervention to reach the performance objectives, were developed. Finally, based on the ASE model [69,70], theoretical methods and practical strategies, suitable tools, and materials for the eHealth intervention were developed. Most of the participating patients and stakeholders judged the intervention to be a promising eHealth tool to empower gynecological patients during the perioperative period to return to their normal activities, including work.

Strengths and Limitations

A primary strength of this study lies in the way the eHealth intervention was developed, tailored, and assessed. Both theory and evidence were combined and patients and most relevant stakeholders were involved, minimizing the risks of theory and/or program failure [72]. The frequent involvement of patients in several steps of the IM process resulted in an eHealth intervention that was specifically tailored to their needs and wishes and therefore more likely to be implemented successfully. In addition to information supply, which is the primary aim of most websites, this eHealth intervention distinguishes itself by monitoring the recovery process, giving tailor-made advice based on patients' workloads, and informing patients when additional consultation of care providers is needed. By linking patients with their gynecologists, convalescence recommendations can be adapted and insecurities regarding consequences of the complications can be solved. Connecting patients and employers facilitates a dialogue and the joint effort to compose a reintegration plan. Furthermore, this eHealth intervention is developed to be used without support and with minimal effort of care providers. Therefore, use of the intervention costs little and implementation is expected to be relatively easy. Moreover, like most eHealth interventions, an important strength is the possibility to use it at the time, place, and pace that fits the patient, care provider, and employer [38]. Finally, the combined approach of encouraging and helping patients to participate in their consultation and empowering clinicians with skills to identify and adapt to the needs of their patients is thought to produce long-term benefits for patients [21].

Main limitations concerning the needs assessment of this study include a possible selection bias; patients assigned to the focus group discussions are a selection of the patients willing to discuss their perioperative problems. Patients less willing to discuss their problems may also experience different perioperative issues. However, through purposeful sampling and by proactively approaching all relevant patients for participation in the focus group discussions, we tried to minimize this selection bias as much as possible. In addition, the influence of dominant patients who might be overly influential cannot be excluded. On the other hand, specific observations on this matter showed that this rarely occurred [Pittens et al, unpublished data, 2012]. Furthermore, these patients already underwent the surgery, whereas the intervention is designed to be used both before and after surgery. It has to be determined whether this intervention is applicable to the entire target population and whether the intervention fits the needs of patients both before and after surgery. Due to practical reasons, not all stakeholders (eg, employers and health care providers) were involved in the needs assessment and development process of this eHealth intervention. As a consequence, the intervention might be less supported by these groups. However, results of prior focus group discussions with supervisors and care providers in another comparable IM study [47] were used and some of those stakeholders were also involved in the test phase. Because this was an exclusively Dutch study directed at the Dutch health care system, a final limitation is that external validity of the eHealth intervention has to be examined before the results may be applied internationally.

Comparison with Other Studies

To our knowledge, this is the first study that tailors an eHealth intervention through the IM protocol to empower gynecological patients during the perioperative period to obtain timely RTW and prevent work disability. Therefore, comparison with other studies is limited. However, previous research showed several developmental and interventional characteristics. For example, it was demonstrated that IM is a successful method to tailor eHealth [45,76] as well as RTW [47,48] interventions. Moreover, Web-based interventions show positive effects on empowerment [25]. Furthermore, it is proven that tailoring an eHealth intervention influences usage positively (eg, time and frequency) and increases the effectiveness of the message [77,78]. In contrast to most eHealth interventions, this intervention aims at secondary and tertiary prevention. Therefore, further research is needed to determine whether the characteristics mentioned previously also apply to the present study.

Although comparable studies are lacking, the approach followed in this study—involving relevant stakeholders in the development of an eHealth intervention—is in line with an observed trend of multi-stakeholder involvement in health care in general [79,80]. Gained experiences in this study might contribute to additional insights for future initiatives on multi-stakeholder involvement in health care.

Interpretation of the Results and Policy Implications

This study shows that the IM protocol can successfully be used for the development and tailoring of an eHealth intervention for gynecological patients. The protocol led to a systematic development of the intervention, it made sure that collaboration with the main target group was realized, and both theory and evidence was used to tailor the intervention.

Furthermore, through the detailed convalescence recommendations provided by the eHealth intervention, patients will be better informed about when it is thought to be medically safe to resume daily and work activities after gynecological surgery and it will give them the possibility to arrange workplace adaptations if necessary [74]. Prospective cohort studies exploring sick leave after general surgical procedures show that return to work is primarily influenced by the expectations of the patient and their supervisors rather than physical factors or the type of surgery [1,10,81]. Therefore, it is assumed that these tailor-made convalescence recommendations will help to accelerate recovery and stimulate patients to resume activities with increasing gradations of strain, which will presumably bring about a quicker recovery and RTW and prevent work disability [82-84]. Therefore, it is expected that this eHealth intervention fulfills patients' needs and is able to empower gynecological patients during the perioperative period and return to normal activities and work [77]. However, its adoption, barriers for usage in daily practice, and implementation possibilities by patients and stakeholders still need to be evaluated more extensively in a process evaluation. Furthermore, a RCT will be needed to assess the effect of empowering gynecological patients during the perioperative period and return to normal activities and work by this eHealth intervention on work disability prevention, resumption of activities, and quality of life [75]. The results are important to assess this intervention's true value and policy implications.

This eHealth intervention is developed for patients who underwent a hysterectomy or laparoscopic adnexal surgery. However, the strategy used to develop the intervention and the final result may also be used as a blueprint for other kinds of surgical procedures.

Conclusion

The development of an eHealth intervention according to the IM protocol to obtain timely RTW and prevent work disability by empowerment and improving communication after gynecological surgery resulted in an intervention based on both theory and evidence and involvement of patients and most stakeholders. This eHealth intervention is well accepted by patients and stakeholders and is considered to be a promising tool to obtain timely RTW and prevent work disability after gynecological surgery. Its effectiveness needs to be proven in a RCT [75].

Acknowledgments

The authors would like to thank all patients in the focus group discussions for their generous contribution of views and time. We also wish to thank all the physicians, eHealth specialists, and the representative of the patient organization who judged the eHealth intervention. In addition, we would like to thank D Stomp for the pleasant collaboration and involvement as website producer of this eHealth intervention, MMA Brood for being panel chair during the focus group discussions, LM Bolten for her dedicated correspondence with all patients, and AJ van der Ham and SC van Veen for their contribution during the focus group discussions and analysis of the transcripts. In addition, our appreciation goes to VU University's production company, Audiovisueel centrum VU-VUmc, and director I Broekhuizen. Finally, we would like to thank Professor G Kok for his valuable recommendations regarding the description of the IM protocol.

This study was carried out with funding from ZonMw (project number 150020037), an organization for health research and development in the Netherlands. The open access publication was supported by The Netherlands Organisation for Scientific Research (NWO).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Attitude Social influence-self-Efficacy model (ASE) [69,70] adapted for recovery and Return to Normal Activities (RNA) and Return to Work (RTW) after gynaecological surgery.

[PDF File (Adobe PDF File), 18KB - jmir_v14i5e124_app1.pdf]

Multimedia Appendix 2

Example of change objectives of patients.

[PDF File (Adobe PDF File), 25KB - jmir_v14i5e124_app2.pdf]

Multimedia Appendix 3

Theoretical methods and practical strategies for recovery and return to normal and work activities.

[PDF File (Adobe PDF File), 42KB - jmir_v14i5e124_app3.pdf]

Multimedia Appendix 4

Evaluation questionaire of the eHealth intervention ikherstel.nl for focus group participants.

[PDF File (Adobe PDF File), 305KB - jmir_v14i5e124_app4.pdf]

Multimedia Appendix 5

Evaluation questionaire of the eHealth intervention ikherstel.nl for professionals.

[PDF File (Adobe PDF File), 303KB - jmir_v14i5e124_app5.pdf]

Multimedia Appendix 6

Screenshot of http://www.ikherstel.nl.

[PDF File (Adobe PDF File), 81KB - jmir_v14i5e124_app6.pdf]

Multimedia Appendix 7

Screenshot of http://www.ikherstel.nl.

[PDF File (Adobe PDF File), 206KB - jmir_v14i5e124_app7.pdf]

Multimedia Appendix 8

Screenshot of http://www.ikherstel.nl.

[PDF File (Adobe PDF File), 752KB - jmir_v14i5e124_app8.pdf]

Multimedia Appendix 9

Screenshot of http://www.ikherstel.nl.

[PDF File (Adobe PDF File), 193KB - jmir_v14i5e124_app9.pdf]

Multimedia Appendix 10

Screenshot of http://www.ikherstel.nl.

[PDF File (Adobe PDF File), 225KB - jmir_v14i5e124_app10.pdf]

References

1. Clayton M, Verow P. A retrospective study of return to work following surgery. Occup Med (Lond) 2007 Oct;57(7):525-531 [FREE Full text] [doi: 10.1093/occmed/kqm082] [Medline: 17906267]
2. Johansen P, Al-Khafagi SK, Thøstesen LM, Lauszus FF, Rasmussen KL. [Analysis of need for sick leave after hysterectomy]. Ugeskr Laeger 2008 Apr 21;170(17):1465-1468. [Medline: 18462627]
3. Kikuchi I, Takeuchi H, Shimanuki H, Kitade M, Kumakiri J, Kuroda K, et al. Questionnaire analysis of recovery of activities of daily living after laparoscopic surgery. J Minim Invasive Gynecol 2008 Jan;15(1):16-19. [doi: 10.1016/j.jmig.2007.08.606] [Medline: 18262138]
4. Brölmann HA, Vonk Noordegraaf A, Bruinvels DJ, de Vet RH, Dirksz AA, Huirne JA. Can prolonged sick leave after gynecologic surgery be predicted? An observational study in The Netherlands. Surg Endosc 2009 Oct;23(10):2237-2241. [doi: 10.1007/s00464-008-0287-0] [Medline: 19118421]
5. Department of Health. Choosing Health: Making Healthy Choices Easier. London: National Health Service (NHS); 2004.
6. van Oostrom SH, Driessen MT, de Vet HC, Franche RL, Schonstein E, Loisel P, et al. Workplace interventions for preventing work disability. Cochrane Database Syst Rev 2009(2):CD006955. [doi: 10.1002/14651858.CD006955.pub2] [Medline: 19370664]
7. Commissie Psychische Arbeidsongeschiktheid. Aanpak verzuim om psychische redenen. Leidraad 2001 Nov 01.
8. Waddell G, Burton AK. Is Work Good for Your Health and Well-being?. London: The Stationery Office (TSO); 2006.

9. Henderson M, Glozier N, Holland Elliott K. Long term sickness absence. BMJ 2005 Apr 9;330(7495):802-803. [doi: 10.1136/bmj.330.7495.802] [Medline: 15817531]
10. Bisgaard T, Klarskov B, Rosenberg J, Kehlet H. Factors determining convalescence after uncomplicated laparoscopic cholecystectomy. Arch Surg 2001 Aug;136(8):917-921. [Medline: 11485527]
11. Warren L, Ladapo JA, Borah BJ, Gunnarsson CL. Open abdominal versus laparoscopic and vaginal hysterectomy: analysis of a large United States payer measuring quality and cost of care. J Minim Invasive Gynecol 2009 Sep;16(5):581-588. [doi: 10.1016/j.jmig.2009.06.018] [Medline: 19835801]
12. Thiel JA, Kamencic H. Assessment of costs associated with outpatient total laparoscopic hysterectom. J Obstet Gynaecol Can 2006 Sep;28(9):794-798. [Medline: 17022920]
13. Everett TR, Crawford RA. Laparoscopic assisted vaginal hysterectomy and bilateral salpingo-oophorectomy as a day surgery procedure: a promising option. J Obstet Gynaecol 2010;30(7):697-700. [doi: 10.3109/01443615.2010.504866] [Medline: 20925613]
14. Dirksz AA, Bruinvels DJ, de Vet HCW, Brolmann HA. Werkhervatting na gynaecologische ingrepen. Scriptie 2008.
15. Löfgren A, Hagberg J, Alexanderson K. What physicians want to learn about sickness certification: analyses of questionnaire data from 4019 physicians. BMC Public Health 2010;10:61 [FREE Full text] [doi: 10.1186/1471-2458-10-61] [Medline: 20144230]
16. Weevers HJ, van der Beek AJ, Anema JR, van der Wal G, van Mechelen W. Work-related disease in general practice: a systematic review. Fam Pract 2005 Apr;22(2):197-204 [FREE Full text] [doi: 10.1093/fampra/cmh727] [Medline: 15710641]
17. Dasinger LK, Krause N, Thompson PJ, Brand RJ, Rudolph L. Doctor proactive communication, return-to-work recommendation, and duration of disability after a workers' compensation low back injury. J Occup Environ Med 2001 Jun;43(6):515-525. [Medline: 11411323]
18. Shulman AG, Amid PK, Lichtenstein IL. Returning to work after herniorrhaphy. BMJ 1994 Jul 23;309(6949):216-217. [Medline: 8069134]
19. Lewin SA, Skea ZC, Entwistle V, Zwarenstein M, Dick J. Interventions for providers to promote a patient-centred approach in clinical consultations. Cochrane Database Syst Rev 2001(4):CD003267. [doi: 10.1002/14651858.CD003267] [Medline: 11687181]
20. Cheraghi-Sohi S, Bower P. Can the feedback of patient assessments, brief training, or their combination, improve the interpersonal skills of primary care physicians? A systematic review. BMC Health Serv Res 2008;8:179 [FREE Full text] [doi: 10.1186/1472-6963-8-179] [Medline: 18715516]
21. Kinnersley P, Edwards A, Hood K, Ryan R, Prout H, Cadbury N, et al. Interventions before consultations to help patients address their information needs by encouraging question asking: systematic review. BMJ 2008;337:a485 [FREE Full text] [Medline: 18632672]
22. Haywood K, Marshall S, Fitzpatrick R. Patient participation in the consultation process: a structured review of intervention strategies. Patient Educ Couns 2006 Oct;63(1-2):12-23. [doi: 10.1016/j.pec.2005.10.005] [Medline: 16406464]
23. Bruegel RB. Patient empowerment--a trend that matters. J AHIMA 1998 Sep;69(8):30-3; quiz 35. [Medline: 10182504]
24. Aujoulat I, d'Hoore W, Deccache A. Patient empowerment in theory and practice: polysemy or cacophony? Patient Educ Couns 2007 Apr;66(1):13-20. [doi: 10.1016/j.pec.2006.09.008] [Medline: 17084059]
25. Samoocha D, Bruinvels DJ, Elbers NA, Anema JR, van der Beek AJ. Effectiveness of web-based interventions on patient empowerment: a systematic review and meta-analysis. J Med Internet Res 2010;12(2):e23 [FREE Full text] [doi: 10.2196/jmir.1286] [Medline: 20581001]
26. Cross M. How the internet is changing health care. BMJ 2008;337:a883. [Medline: 18647776]
27. Randeree E. Exploring technology impacts of Healthcare 2.0 initiatives. Telemed J E Health 2009 Apr;15(3):255-260. [doi: 10.1089/tmj.2008.0093] [Medline: 19382863]
28. de Jong T, Heinrich J, Blatter BM, Anema JR, van der Beek AJ. The feasibility of a web-based counselling program for occupational physicians and employees on sick leave due to back or neck pain. BMC Med Inform Decis Mak 2009;9:46 [FREE Full text] [doi: 10.1186/1472-6947-9-46] [Medline: 19895689]
29. Forkner-Dunn J. Internet-based patient self-care: the next generation of health care delivery. J Med Internet Res 2003 Apr;5(2):e8 [FREE Full text] [doi: 10.2196/jmir.5.2.e8] [Medline: 12857664]
30. Hartmann CW, Sciamanna CN, Blanch DC, Mui S, Lawless H, Manocchia M, et al. A website to improve asthma care by suggesting patient questions for physicians: qualitative analysis of user experiences. J Med Internet Res 2007;9(1):e3 [FREE Full text] [doi: 10.2196/jmir.9.1.e3] [Medline: 17478412]
31. Gerber BS, Eiser AR. The patient physician relationship in the Internet age: future prospects and the research agenda. J Med Internet Res 2001 Apr;3(2):E15 [FREE Full text] [doi: 10.2196/jmir.3.2.e15] [Medline: 11720957]
32. Brug J, Campbell M, van Assema P. The application and impact of computer-generated personalized nutrition education: a review of the literature. Patient Educ Couns 1999 Feb;36(2):145-156. [Medline: 10223019]
33. Kroeze W, Werkman A, Brug J. A systematic review of randomized trials on the effectiveness of computer-tailored education on physical activity and dietary behaviors. Ann Behav Med 2006 Jun;31(3):205-223. [doi: 10.1207/s15324796abm3103_2] [Medline: 16700634]

34. Lustria ML, Cortese J, Noar SM, Glueckauf RL. Computer-tailored health interventions delivered over the Web: review and analysis of key components. Patient Educ Couns 2009 Feb;74(2):156-173. [doi: 10.1016/j.pec.2008.08.023] [Medline: 18947966]
35. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010 Sep;51(3-4):214-221. [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]
36. Neville LM, Milat AJ, O'Hara B. Computer-tailored weight reduction interventions targeting adults: a narrative systematic review. Health Promot J Austr 2009 Apr;20(1):48-57. [Medline: 19402816]
37. Miniwatts Marketing Group. Internet World Stats. 2010. Usage and Population Statistics URL: http://www.internetworldstats.com/stats.htm [accessed 2012-05-29] [WebCite Cache ID 681BI6Gyl]
38. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
39. Alpay LL, Henkemans OB, Otten W, Rövekamp TA, Dumay AC. E-health applications and services for patient empowerment: directions for best practices in The Netherlands. Telemed J E Health 2010 Sep;16(7):787-791. [doi: 10.1089/tmj.2009.0156] [Medline: 20815745]
40. Vitacca M, Mazzù M, Scalvini S. Socio-technical and organizational challenges to wider e-Health implementation. Chron Respir Dis 2009;6(2):91-97. [doi: 10.1177/1479972309102805] [Medline: 19411570]
41. Bartholomew LK, Parcel GS, Kok G. Intervention mapping: a process for developing theory- and evidence-based health education programs. Health Educ Behav 1998 Oct;25(5):545-563. [Medline: 9768376]
42. Bartholomew LK, Parcel GS, Kok G, Gottlieb NH. Planning Health Promotion Programs: An Intervention Mapping Approach. San Francisco, CA: Jossey-Bass; 2006.
43. Kok G, van Essen GA, Wicker S, Llupià A, Mena G, Correia R, et al. Planning for influenza vaccination in health care workers: an Intervention Mapping approach. Vaccine 2011 Nov 3;29(47):8512-8519. [doi: 10.1016/j.vaccine.2011.09.008] [Medline: 21939722]
44. Samoocha D, Bruinvels DJ, Anema JR, Steenbeek R, van der Beek AJ. Empowerment of disability benefit claimants through an interactive website: design of a randomized controlled trial. BMC Med Inform Decis Mak 2009;9:23 [FREE Full text] [doi: 10.1186/1472-6947-9-23] [Medline: 19426557]
45. Albada A, van Dulmen S, Otten R, Bensing JM, Ausems MG. Development of E-info gene(ca): a website providing computer-tailored information and question prompt prior to breast cancer genetic counseling. J Genet Couns 2009 Aug;18(4):326-338. [doi: 10.1007/s10897-009-9221-4] [Medline: 19440661]
46. Prins RG, van Empelen P, Beenackers MA, Brug J, Oenema A. Systematic Development of the YouRAction program, a computer-tailored physical activity promotion intervention for Dutch adolescents, targeting personal motivations and environmental opportunities. BMC Public Health 2010;10:474 [FREE Full text] [doi: 10.1186/1471-2458-10-474] [Medline: 20701782]
47. van Oostrom SH, Anema JR, Terluin B, Venema A, de Vet HC, van Mechelen W. Development of a workplace intervention for sick-listed employees with stress-related mental disorders: Intervention Mapping as a useful tool. BMC Health Serv Res 2007;7:127 [FREE Full text] [doi: 10.1186/1472-6963-7-127] [Medline: 17697364]
48. Vermeulen SJ, Anema JR, Schellart AJ, van Mechelen W, van der Beek AJ. Intervention mapping for development of a participatory return-to-work intervention for temporary agency workers and unemployed workers sick-listed due to musculoskeletal disorders. BMC Public Health 2009;9:216 [FREE Full text] [doi: 10.1186/1471-2458-9-216] [Medline: 19573229]
49. Kristensen TS. Intervention studies in occupational epidemiology. Occup Environ Med 2005 Mar;62(3):205-210 [FREE Full text] [doi: 10.1136/oem.2004.016097] [Medline: 15723887]
50. Green LW, Kreuter MW. Health Program Planning: An Educational and Ecological Approach. 4th edition. New York: McGraw Hill Professional; 2005.
51. Stichting Informatie Gezondheidszorg. Landelijke Medische Registratie LMR - Verrichtingen. Prismant 2009.
52. Centraal Bureau voor de Statistiek. Women with paid work. Statline databank 2009.
53. Morgan D. Focus Groups as Qualitative Research. Thousand Oaks, CA: Sage; 1997.
54. Barbour RS, Kitzinger J. Developing Focus Group Research: Politics Theory and Practice. London: Sage; 1999.
55. Scientific Software Development. ATLAS.ti version 6.2. Berlin: Scientific Software Development; 2010.
56. van Achterberg T, Huisman-de Waal GG, Ketelaar NA, Oostendorp RA, Jacobs JE, Wollersheim HC. How to promote healthy behaviours in patients? An overview of evidence for behaviour change techniques. Health Promot Int 2011 Jun;26(2):148-162 [FREE Full text] [doi: 10.1093/heapro/daq050] [Medline: 20739325]
57. 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. [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
58. Møller C, Ottesen M, Kehlet H, Ottesen BS. [Convalescence recommendations after hysterectomy. A study of opinions among Danish physicians]. Ugeskr Laeger 2001 Dec 10;163(50):7043-7047. [Medline: 11794035]
59. Clayton M, Verow P. Advice given to patients about return to work and driving following surgery. Occup Med (Lond) 2007 Oct;57(7):488-491 [FREE Full text] [doi: 10.1093/occmed/kqm063] [Medline: 17906266]

60. Shaw L, Segal R, Polatajko H, Harburn K. Understanding return to work behaviours: promoting the importance of individual perceptions in the study of return to work. Disabil Rehabil 2002 Mar 10;24(4):185-195. [Medline: 11926259]
61. Rosenberger PH, Jokl P, Ickovics J. Psychosocial factors and surgical outcomes: an evidence-based literature review. J Am Acad Orthop Surg 2006 Jul;14(7):397-405. [Medline: 16822887]
62. Celestin J, Edwards RR, Jamison RN. Pretreatment psychosocial variables as predictors of outcomes following lumbar surgery and spinal cord stimulation: a systematic review and literature synthesis. Pain Med 2009 May;10(4):639-653. [doi: 10.1111/j.1526-4637.2009.00632.x] [Medline: 19638142]
63. Shehmar M, Gupta JK. The influence of psychological factors on recovery from hysterectomy. J R Soc Med 2010 Feb;103(2):56-59. [doi: 10.1258/jrsm.2009.090198] [Medline: 20118335]
64. Brouwer S, Krol B, Reneman MF, Bültmann U, Franche RL, van der Klink JJ, et al. Behavioral determinants as predictors of return to work after long-term sickness absence: an application of the theory of planned behavior. J Occup Rehabil 2009 Jun;19(2):166-174. [doi: 10.1007/s10926-009-9172-5] [Medline: 19333738]
65. Oyeflaten I, Hysing M, Eriksen HR. Prognostic factors associated with return to work following multidisciplinary vocational rehabilitation. J Rehabil Med 2008 Jul;40(7):548-554 [FREE Full text] [doi: 10.2340/16501977-0202] [Medline: 18758672]
66. Svajger A, Winding K. Perceptions of possibilities of returning to work with chronic musculoskeletal disorders. Work 2009;32(4):443-454. [doi: 10.3233/WOR-2009-0855] [Medline: 19478415]
67. Lambeek LC, van Mechelen W, Knol DL, Loisel P, Anema JR. Randomised controlled trial of integrated care to reduce disability from chronic low back pain in working and private life. BMJ 2010;340:c1035 [FREE Full text] [Medline: 20234040]
68. Spelten ER, Sprangers MA, Verbeek JH. Factors reported to influence the return to work of cancer survivors: a literature review. Psychooncology 2002 Mar;11(2):124-131. [Medline: 11921328]
69. de Vries H, Dijkstra M, Kuhlman P. Self-efficacy: The third factor besides attitude and subjective norm as a predictor of behavioural intentions. Health Educ Res 1988;3:273-282.
70. de Vries H. Determinanten van gedrag. In: Gezondheidsvoorlichting en gedragsverandering [Health Education and Behavioral Change]. Assen: Van Gorcum; 1993:109-132.
71. Ajzen I. From intentions to action: a theory of planned behaviour. In: Action-Control: From Cognition to Behaviour. Heidelberg: Springer; 1985:11-39.
72. Schaalma H, Kok G. Decoding health education interventions: the times are a-changin'. Psychol Health 2009 Jan;24(1):5-9. [doi: 10.1080/08870440903126348] [Medline: 20186636]
73. Vonk Noordegraaf A, Huirne JAF, Van Mechelen W, Brolmann HAM, Anema JR. Ikherstel. URL: http://dev.ikherstel.nl/ [accessed 2012-05-29] [WebCite Cache ID 6818criCG]
74. Vonk Noordegraaf A, Huirne JA, Brölmann HA, van Mechelen W, Anema JR. Multidisciplinary convalescence recommendations after gynaecological surgery: a modified Delphi method among experts. BJOG 2011 Dec;118(13):1557-1567. [doi: 10.1111/j.1471-0528.2011.03091.x] [Medline: 21895950]
75. Vonk Noordegraaf A, Huirne JA, Brölmann HA, Emanuel MH, van Kesteren PJ, Kleiverda G, et al. Effectiveness of a multidisciplinary care program on recovery and return to work of patients after gynaecological surgery; design of a randomized controlled trial. BMC Health Serv Res 2012;12:29 [FREE Full text] [doi: 10.1186/1472-6963-12-29] [Medline: 22296950]
76. Schmid AA, Andersen J, Kent T, Williams LS, Damush TM. Using intervention mapping to develop and adapt a secondary stroke prevention program in Veterans Health Administration medical centers. Implement Sci 2010;5:97 [FREE Full text] [doi: 10.1186/1748-5908-5-97] [Medline: 21159171]
77. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of Web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]
78. Quintiliani LM, Campbell MK, Bowling JM, Steck S, Haines PS, DeVellis BM. Results of a randomized trial testing messages tailored to participant-selected topics among female college students: physical activity outcomes. J Phys Act Health 2010 Jul;7(4):517-526. [Medline: 20683094]
79. van de Bovenkamp HM, Trappenburg MJ, Grit KJ. Patient participation in collective healthcare decision making: the Dutch model. Health Expect 2010 Mar;13(1):73-85. [doi: 10.1111/j.1369-7625.2009.00567.x] [Medline: 19719537]
80. Williamson C. The patient movement as an emancipation movement. Health Expect 2008 Jun;11(2):102-112. [doi: 10.1111/j.1369-7625.2007.00475.x] [Medline: 18494955]
81. Jones KR, Burney RE, Peterson M, Christy B. Return to work after inguinal hernia repair. Surgery 2001 Feb;129(2):128-135. [Medline: 11174702]
82. Callesen T, Klarskov B, Bech K, Kehlet H. Short convalescence after inguinal herniorrhaphy with standardised recommendations: duration and reasons for delayed return to work. Eur J Surg 1999 Mar;165(3):236-241. [doi: 10.1080/110241599750007108] [Medline: 10231657]
83. Ottesen M, Sørensen M, Kehlet H, Ottesen B. Short convalescence after vaginal prolapse surgery. Acta Obstet Gynecol Scand 2003 Apr;82(4):359-366. [Medline: 12716321]

84. Bay-Nielsen M, Thomsen H, Andersen FH, Bendix JH, Sørensen OK, Skovgaard N, et al. Convalescence after inguinal herniorrhaphy. Br J Surg 2004 Mar;91(3):362-367. [doi: 10.1002/bjs.4437] [Medline: 14991640]

Abbreviations

ASE: Attitude–Social influence–self-Efficacy
CMS: content management system
IM: intervention mapping
RCT: randomized controlled trial
RTW: return to work

Edited by R Crutzen; submitted 24.08.11; peer-reviewed by G Kok, H van Beekhuizen, S Detaille; comments to author 11.10.11; revised version received 31.01.12; accepted 29.05.12; published 19.10.12.

Please cite as:
Vonk Noordegraaf A, Huirne JA, Pittens CA, van Mechelen W, Broerse JE, Brölmann HA, Anema JR
eHealth Program to Empower Patients in Returning to Normal Activities and Work After Gynecological Surgery: Intervention Mapping as a Useful Method for Development
J Med Internet Res 2012;14(5):e124
URL: http://www.jmir.org/2012/5/e124/
doi:10.2196/jmir.1915
PMID:23086834

Original Paper

Effect of a Web-Based Intervention to Promote Physical Activity and Improve Health Among Physically Inactive Adults: A Population-Based Randomized Controlled Trial

Andreas Wolff Hansen[1], MSc; Morten Grønbæk[1], MD, PhD, Dr. med.; Jørn Wulff Helge[2], MSc, PhD; Maria Severin[1], MSc; Tine Curtis[1], mag.scient.soc., PhD; Janne Schurmann Tolstrup[1], MSc, PhD, Dr. med.

[1]National Institute of Public Health, University of Southern Denmark, Copenhagen, Denmark
[2]Centre of Healthy Aging, Department of Biomedical Sciences, University of Copenhagen, Copenhagen, Denmark

Corresponding Author:
Andreas Wolff Hansen, MSc
National Institute of Public Health
University of Southern Denmark
Oester Farimagsgade 5A, 2.
Copenhagen, DK-1353
Denmark
Phone: 45 65507730
Fax: 45 39208010
Email: awh@niph.dk

Abstract

Background: Many people in Western countries do not follow public health physical activity (PA) recommendations. Web-based interventions provide cost- and time-efficient means of delivering individually targeted lifestyle modification at a population level.

Objective: To examine whether access to a website with individually tailored feedback and suggestions on how to increase PA led to improved PA, anthropometrics, and health measurements.

Methods: Physically inactive adults (n = 12,287) participating in a nationwide eHealth survey and health examination in Denmark were randomly assigned to either an intervention (website) (n = 6055) or a no-intervention control group (n = 6232) in 2008. The intervention website was founded on the theories of stages of change and of planned behavior and, apart from a forum page where a physiotherapist answered questions about PA and training, was fully automated. After 3 and again after 6 months we emailed participants invitations to answer a Web-based follow-up questionnaire, which included the long version of the International Physical Activity Questionnaire. A subgroup of participants (n = 1190) were invited to a follow-up health examination at 3 months.

Results: Less than 22.0% (694/3156) of the participants logged on to the website once and only 7.0% (222/3159) logged on frequently. We found no difference in PA level between the website and control groups at 3- and 6-month follow-ups. By dividing participants into three groups according to use of the intervention website, we found a significant difference in total and leisure-time PA in the website group. The follow-up health examination showed no significant reductions in body mass index, waist circumference, body fat percentage, and blood pressure, or improvements in arm strength and aerobic fitness in the website group.

Conclusions: Based on our findings, we suggest that active users of a Web-based PA intervention can improve their level of PA. However, for unmotivated users, single-tailored feedback may be too brief. Future research should focus on developing more sophisticated interventions with the potential to reach both motivated and unmotivated sedentary individuals.

Trial Registration: Clinicaltrials.gov NCT01295203; http://clinicaltrials.gov/ct2/show/NCT01295203 (Archived by WebCite at http://www.webcitation.org/6B7HDMqiQ)

(J Med Internet Res 2012;14(5):e145) doi:10.2196/jmir.2109

KEYWORDS

Intervention study: computer intervention; health behavior; primary prevention; adults

Introduction

Physical activity (PA) is associated with lower mortality and morbidity rates from cardiovascular disease, diabetes mellitus, cancer, and osteoporosis [1-6]. Despite the health benefits of PA, many people in Western countries do not follow public health PA recommendations [7,8].

Web-based interventions have been successfully applied to improve lifestyle and change health behavior targeting weight loss, stress management, fall-related injuries, smoking cessation, and heavy drinking [9-12]. The Internet has several advantages in delivering health promotion—for example, cost and time effectiveness, 24-hour accessibility, and generation of instant personalized or individually tailored feedback [13-15]. Individually tailored feedback, rather than a more general prevention message, is likely to be more effective, as users can identify with relevant personal information rather than general information [15-19]. Overall, Web-based interventions provide cost- and time-efficient means of delivering individually targeted lifestyle modification at a population level. Based on the existing Web-based randomized controlled trials in relation to PA, it is unclear whether the Internet can effectively deliver PA interventions [20-22]. A review from 2009 [23] identified Web-based PA interventions in primary prevention; only 4 of 16 studies had above-average external validity and reported a positive between-group effect in PA. Basically, two approaches exist in PA Web-based interventions. PA is either the only target in the intervention or included in a general lifestyle intervention targeting several behaviors simultaneously. However, evidence supporting the one or the other is inconclusive [22,23]. Many Web-based interventions conducted to date have limitations. Some studies used an inadequately powered sample, lacked a no-information control group, had a large dropout rate, or relied only on self-reported outcome measures.

The aim of this study was to examine whether an automated Web-based intervention would lead to increased PA among inactive persons in a large population. More specifically, we aimed to determine whether access to a website with individually tailored feedback on PA level and suggestions to increase PA would result in improvements in self-reported PA, anthropometrics, and physiological measurements in an intervention group compared with a no-information control group.

Methods

Setting

The intervention study was nested in the Danish Health Examination Survey 2007-2008 (DANHES) [24], a nationwide health study in Denmark. DANHES was carried out in 13 municipalities in 2007-2008 and comprised a Web-based questionnaire and a health examination test. All adults (18+ years of age) in 13 municipalities (N = 538,497) were invited by letter to take part in a Web-based questionnaire. The questionnaire assessed lifestyle-related health aspects with a focus on diet, smoking, alcohol use, and exercise. A representative sample in each of the 13 municipalities were also invited to a health examination test (n = 180,103). For details see Eriksen et al [24].

Recruitment of Participants

The DANHES was used to recruit participants and as a baseline assessment. The intervention study was conducted in 11 of the 13 municipalities participating in DANHES during May 2008–May 2009. We excluded 2 municipalities from the intervention study, as one served as a pilot study for DANHES and paper questionnaires were used in the other.

The main inclusion criterion for participating in the intervention study was being physically inactive during leisure time. This was defined by the participants' answer to a 4-category question describing PA level in leisure time. We included participants in the lowest 2 categories, mostly sedentary or light activities, in the study and excluded participants in the highest categories, moderate and vigorous PA. Further exclusion criteria were presence of serious heart problems, not being able to perform everyday activities, or missing values in the International Physical Activity Questionnaire (IPAQ) and the leisure-time PA question.

We identified participants who met the inclusion criteria by a screening program and invited them to join the intervention study at the end of the questionnaire in DANHES. If willing to participate, each participant was randomly assigned by the registration program to either an intervention (website) or a no-intervention control group. The only incentive given to participants was the possibility of being assigned to the intervention group. Blinding was not feasible.

The participants in the website group received an email with a link to a PA website immediately after allocation to the website group. In addition, relevant data from the health survey were automatically transferred to the intervention website. To access the website, the participants were required to log on to the website, using the same personal username and password given in the health survey.

All participants gave informed consent before being enrolled in the study. The study was approved by the Danish National Committee on Biomedical Research Ethics (H-D-2008-035).

The Intervention Website

The intervention website was founded on the theories of stages of change [25] and of planned behavior [26]. First, we identified determinants in the theory: intentions, attitudes, self-efficacy, skills, social support, and knowledge. Second, we specified key objectives for each determinant and decided where in the intervention the objectives would be implemented (Table 1). The key objectives were used in the wider planning of the intervention to specify how objectives could be translated into action in the real-life intervention.

Table 1. Key determinants and objectives, and how they were incorporated in the Web-based physical activity (PA) intervention.

Determinant	Objective	Used in the intervention
Intentions	Have intention to increase PA	Tailored feedback or emails sent to users
	Maintain motivation and intention	Personal data (biofeedback)
Attitudes	Experience that PA is important and requires an extraordinary effort	Tailored feedback or emails sent to users
Self-efficacy	Be confident that PA can be increased	Tailored feedback or emails sent to users
Skills	Demonstrate skills to set goals	Goal setting
	Can identify situations where the new behavior is being challenged	General recommendations
Social and external support	Know that help can be provided by an expert	The forum for users
	Know that equals can be found in the forum	The forum for users
	Receive positive feedback from other users of the forum	The forum for users
	Know that family and friends in the near social environment can support and participate in the new behavior	General recommendations
Knowledge	Gain knowledge of how PA can be increased and which kind of activities are best suited	Tailored feedback, general training programs
	Know how increased PA can benefit everyday life	Tailored feedback, general recommendations

The website was structured as three major parts: (1) a personal page, which included individually tailored PA advice and a personal profile, (2) a page with training programs and general recommendations, and (3) a forum and discussion page for questions from participants.

The individually tailored PA advice consisted of three parts: (1) a general introduction, (2) normative feedback, which related the participant's PA to the current PA recommendations and (3) general advice about using the tools on the website. The normative feedback was based on the summarized PA time from the participant's answers in the IPAQ. Feedback was given in the domains of everyday activity, fitness training, and strength training. In each domain we defined categories in which the participants received tailored feedback according to their level of PA. The categories were partly based on PA recommendations from the Danish National Board of Health translated into minutes per week [27], together with analyses from answers in the IPAQ from 2 municipalities. The analyses showed that the categories needed to be wide because of overreporting in the IPAQ. The categories were everyday activity, being the total activity time from questions in the IPAQ summarized into low (<1200 min/week), moderate (1200-3500 min/week), and high levels (>3500 min/week); fitness training, being summarized time from vigorous and moderate-intensity activity from the transport and the leisure-time domain and moderate-intensity activity from the domestic domain in the IPAQ, at low (<40 min/week), moderate (40-350 min/week), and high (>350 min/week) levels; and strength training, being summarized time from the highest intensity level in the 4 domains in the IPAQ, at low (<100 min/week) and moderate to high (>100 min/week) levels. Participants who were 60 years of age or older were given extra advice regarding the importance of strength training. To monitor their progress during the study, the participants could register personal data, such as waist circumference and the result of a short fitness and strength test in the personal profile. We used this kind of biofeedback to keep the participants motivated during the intervention. Furthermore, current activity per day could be calculated with a short activity calculator, and it was possible to set goals for the following 4 weeks. We decided on 4 weeks for goal setting as a balance between time for seeing results and maintaining motivation. All participants were encouraged to make a personal profile to set their goals, monitor their progress, and implement their goals.

The part of the website that included the page with the training programs and general recommendations was structured in the same way as the PA advice in the 3 domains: everyday activity, fitness training, and strength training. The participants were encouraged to go through the different suggestions and programs and pick the ones that suited them best based on the individually tailored advice and goals set by each participant. General information about motivation and relevant links were also presented on the website.

On the forum and discussion page, a physiotherapist experienced in PA counseling answered all questions about PA and training from participants. In addition, participants could share experiences and give each other tips or search for training partners in a second forum.

We kept the tailored PA advice short so as not to overload the participant with information and, apart from training programs and general recommendations, as a means of following the theory of stages of change [25]. The theory describes how change is a process of progressing through a series of stages: precontemplation, contemplation, preparation, action, maintenance, and termination. Hence, participants ready for change and already in the action stage could make a personal profile, set goals, and find training programs. Participants in the precontemplation to preparation stage who did not log on to the website or did not make a personal profile were sent two email reminders to encourage them to become involved with the intervention. Participants in the maintenance stage who made a personal profile were sent reminders and encouraging emails to keep the profile updated after 4, 8, 12, and 16 weeks.

The content of the website was developed by the research team. Two professional Web companies did the graphic design and implementation. The intervention was pretested among experts and representatives of the target population. We tested screening of participants, invitation to the intervention, automatic generation of individually tailored advice, email generation, and general usability of the website. Furthermore, we used the first municipality participating in the study (n = 1298) for a pilot study. Comments and suggestions from these participants were used to fine-tune the website. Multimedia Appendix 1 shows recruitment of participants and the intervention website.

Baseline and Follow-Up Health Examination

The baseline health examination included measurements of blood pressure, height, weight, body fat percentage, and grip strength. Aerobic fitness was estimated from a watt-max [28] or a sub-max test [29] on an ergometer bike. Body mass index (BMI) was calculated from measurements of height and body weight. For further details about methods and test procedures, see Eriksen et al [24].

From 3 selected municipalities in the intervention study, we invited 1190 participants to a follow-up health examination after 12 weeks. The participants were invited by email, which was sent 3 to 4 weeks prior to the examination. If the participant did not respond, a reminder email was sent 1 week after the first. The follow-up examination included the same measurements as the baseline health examination, and the same test procedures were followed. The follow-up examination was blinded to the examiners. After the health examination, the participants received their results to track developments from baseline to follow-up. We expected to see improvements in measurements included in the health examination, as participants in the intervention group were encouraged to engage in moderate- and high-intensity PA on a weekly basis. Improvements have been seen after 3 months in other studies depending on diet [30-32].

Baseline and Follow-Up Questionnaire

The baseline questionnaire included demographics, health, lifestyle, and health behavior. We used the long version of the IPAQ, which is known to be a valid and reliable instrument for assessing PA [33], both at baseline and at follow-up. It consists of 31 items that collect information on PA in the 4 domains work, transport, housework and gardening, and leisure time. Motivation for changing PA behavior was assessed by the question “Do you wish to be more physically active than you already are?” (yes, yes/maybe, or no).

After 3 months and after 6 months, we invited all participants by email to answer a follow-up questionnaire. The follow-up questionnaire included questions about use of the website for the website group. Due to a technical error, only half the participants were invited to answer the 3-month follow-up questionnaire.

Outcome Measures

The primary analysis in this intervention was overall level of PA based on self-reported PA from the IPAQ. The secondary outcome measures were blood pressure, height, weight, body fat percentage, and grip strength. The outcome measures were specified a priori.

As a post hoc outcome measure, a secondary analysis was carried out among active users of the intervention. Here we divided the participants of the website group into three groups according to user activity (no log-on, log-on once, and log-on more than once) and assessed level of PA. Furthermore, we calculated the odds ratios of being an active user of the website.

Power Estimates

We assumed that a reasonable effect of the intervention on total PA time estimated by the IPAQ would be around 12% for the intervention group and 5% for the control group. With a power of 80% probability of detecting a 12% versus 5% difference as statistically significant at the 5% level, we calculated the minimum sample size to be 250 in each group. We expected that approximately 50% of the participants in DANHES were sedentary. Assuming that 80% accepted participation and 25% were lost to follow-up, this would still give us a large population and hence ensure sufficient power.

Statistical Analyses

We analyzed the IPAQ results according to the *Guidelines for Data Processing and Analysis of the International Physical Activity Questionnaire* [34] with the exception that we included participants with a missing value in day or time in the follow-up analysis. Nonparametric analyses were used, since IPAQ data were not distributed normally (Wilcoxon rank sum test between two groups and Kruskal-Wallis test between more than two groups).

Results were primarily analyzed as intention-to-treat analyses with the use of the last observation carried forward to account for missing data at follow-up. We analyzed completer data including only participants who completed the follow-up health examination or questionnaire.

Website use was assessed by the follow-up questionnaire and combined with information provided by the company that was responsible for the website, which recorded whether a participant logged on.

Odds ratios of being an active user (several log-ons versus one or none) were calculated in relation to sex, highest education level (<10, 10–12, 13–14, or 15+ years), age group (18–44, 45–64, or 65+ years), and motivation to be more active (yes, yes/maybe, no) by the use of logistic regression.

For all statistical calculations and analyses, we used Stata version 11.2 (StataCorp LP, College Station, TX, USA). We performed chi-square tests to examine differences in proportions between the groups. We considered $P < .05$ as statistically significant.

Results

Participation

In total, 53,956 persons participated in DANHES in the 11 municipalities. Of these, 28,054 participants met the inclusion criterion and were invited to participate in the intervention study (Figure 1). Among these, reasons for nonparticipation were refusal (8593) and not answering the invitation (6242). A technical error gave some participants in the control group access to the website and resulted in exclusion of 895 participants. A total of 12,287 participants were enrolled in the study, resulting in a 43.80% participation rate. The response rates in the 3-month questionnaire were 57.55% (2375/4127) in the intervention group and 66.41% (2175/3257) in the website group.

In the health examination, 32 participants were excluded due to participation in another health intervention, leaving 583 in the website group and 585 in the control group. In total, 434 (37.2%) participated in the follow-up health examination, with 215 in the intervention groupand 219 in the control group. Participants who were lost to follow-up in the health examination and questionnaire were not significantly different from those who completed follow-up in respect to baseline characteristics (data not shown).

Baseline

Baseline characteristics did not differ significantly between the website and control groups as shown in Table 2. Mean age of the participants was 50 (SD 13.6) years and 64.82% (3925/6055) were women; overall, 18.95% (2329/12,287) were mostly sedentary and the rest reported light PA in their leisure time.

No significant differences were found in baseline characteristics between the website and control groups in the health examination (Table 3). The mean BMI among participants was 25 (SD 3.8) kg/m^{2}and mean body fat percentage was 30.5% (SD 8.1). Aerobic fitness was somewhat low at 32 (SD 7.8) mL/min/kg.

Follow-Up

Participants in the website group did not report a significantly different level of PA compared with the control group at the 6-month follow-up (Table 4).

The result was the same in the completer analysis and analysis at 3-month follow-up (data not shown). Analyzing participants who stated that they wanted to change their PA separately gave the same result (data not shown). In relation to other health and baseline characteristics, no significant differences were found at follow-up at 3 and 6 months (data not shown). Furthermore, no significant changes were found in either the website or the control group from baseline to follow-up (data not shown).

The results from the follow-up health examination showed no significant differences between the website and control groups (Table 5). From baseline to follow-up, we found no significant changes in either the website or the control group (data not shown).

When we divided the participants into three groups according to use of the intervention website (no log-on, log-on once, and log-on more than once), we found a significant difference between the groups in leisure-time PA and total PA (Figure 2). Dividing participants from the website group who participated in the follow-up health examination into the same three subgroups of website use did not show significant differences between the groups (data not shown).

Figure 1. Flow diagram of the Web-based intervention (Denmark, 2008). DANHES = Danish Health Examination Survey 2007–2008.

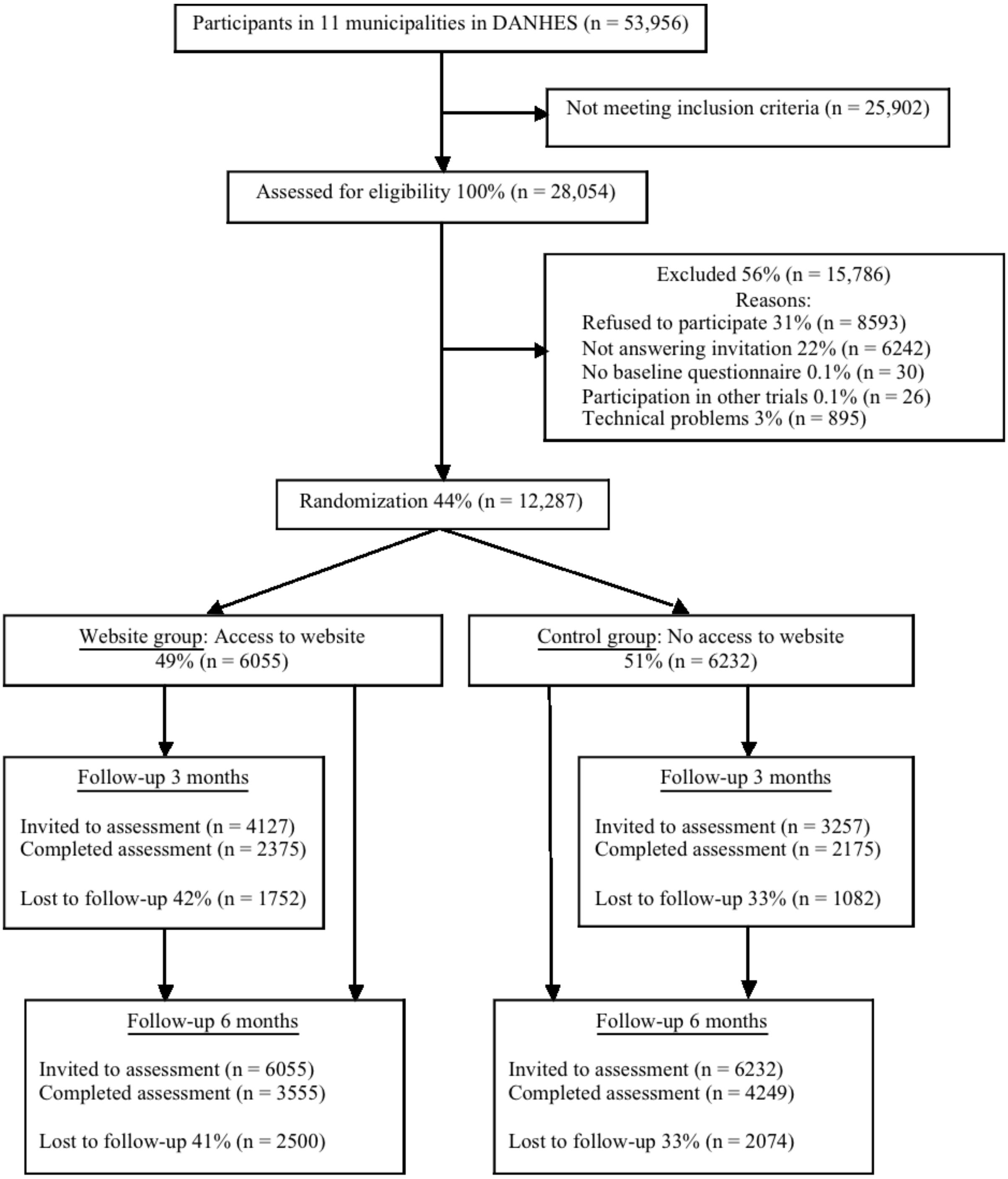


Table 2. Baseline characteristics of the participants by website and control group (Denmark, 2008).

Characteristic	Website group (n = 6055)		Control group (n = 6232)		*P* value[a]
Age (years), mean (SD)	50.7	(13.6)	50.4	(13.7)	.31
Sex (women), n (%)	3924	(64.8%)	4043	(64.9%)	.96
Education (years), n (%)					.18
<10	461	(7.6%)	430	(6.9%)	
10–12	1212	(20.02%)	1259	(20.20%)	
12–14	1491	(24.62%)	1470	(23.59%)	
15+	2891	(47.75%)	3073	(49.31%)	
Physical activity in leisure time, n (%)					.55
Sedentary	1157	(19.11%)	1172	(18.81%)	
Low	4898	(80.89%)	5060	(81.19%)	
Vigorous or moderate	0	0	0	0	
Physical activity by IPAQ [b] (min/week), median (25th–75th percentile) [c]					
Work	60	(0–780)	60	(0–780)	.66
Transportation	160	(40–390)	180	(45–390)	.37
Household	500	(180–1110)	480	(180–1080)	.04
Leisure time	200	(60–465)	195	(60–420)	.15
Sitting	2310	(1650–3180)	2340	(1680–3300)	.06
Total PA	1600	(840–2640)	1560	(840–2485)	.11
Wish to be more physically active (yes), n (%)	3197	(52.80%)	3372	(54.11%)	.32
Self-rated health good or very good, n (%)	4323	(71.40%)	4482	(71.92%)	.53

[a] Independent *t* test for difference between groups for age and chi-square test for differences between groups in categorical characteristics.

[b] International Physical Activity Questionnaire.

[c] Website group n = 4435 and control group n = 4509. Wilcoxon rank sum test for difference between groups.

Table 3. Baseline characteristics of the subsample of participants in the health examination by website and control group (Denmark, 2008).

Characteristic	Website group (n = 583)		Control group (n = 585)		*P* value[a]
Sex (women), n (%)	345	(59.2%)	332	(56.8%)	
Age (years), mean (SD)	51.2	(13.9)	50.5	(13.2)	.39
BMI[b] (kg/m^2), mean (SD)	25.4	(3.8)	25.0	(3.8)	.12
Waist circumference (cm), mean (SD)	90.1	(12.0)	89.6	(11.8)	.43
Body fat (%), mean (SD)	30.4	(8.2)	30.5	(8.0)	.87
Systolic blood pressure (mmHg), mean (SD)	125.0	(16.8)	123.1	(16.3)	.05
Diastolic blood pressure (mmHg), mean (SD)	79.7	(10.0)	78.9	(10.6)	.15
Arm strength (kg), mean (SD)	26.9	(9.4)	29.3	(9.6)	.37
Aerobic fitness[c] (mL/min/kg), mean (SD)	32.0	(7.9)	31.5	(7.7)	.85

[a] Independent *t* test for difference between groups and chi-square test for differences between groups in categorical characteristics.

[b] Body mass index.

[c] Aerobic fitness total either from watt-max or 1-point test.

Table 4. Physical activity assessed by International Physical Activity Questionnaire (min/week) at 6-month follow-up by website and control group (intention-to-treat analysis) (Denmark, 2008).

Type of physical activity[a]	Website group: (n = 4435)		Control group (n = 4509)		P value[b]
Work	60	(0–800)	60	(0–825)	.62
Transportation	180	(45–400)	200	(60–420)	.62
Household	480	(180–1080)	480	(180–1080)	.17
Leisure time	200	(60–450)	200	(60–420)	.25
Sitting	2220	(1500–3060)	2220	(1500–3150)	.52
Total physical activity	1575	(845–2580)	1560	(840–2520)	.25

[a] Variables are shown as median (25th–75th percentile).

[b] Wilcoxon rank sum for difference between groups.

Table 5. Health examination measurements of the subsample of participants at 3-month follow-up by website and control group (intention-to-treat analysis) (Denmark, 2008).

Measurement[a]	Website group (n = 583)		Control group (n = 585)		P value[b]
BMI[c] (kg/m^2)	25.3	(0.2)	25.0	(0.2)	.12
Waist circumference (cm)	90.0	(0.5)	89.1	(0.5)	.34
Body fat (%)	30.4	(0.3)	30.5	(0.3)	.87
Systolic blood pressure (mmHg)	125.1	(0.7)	123.1	(0.7)	.04
Diastolic blood pressure (mmHg)	79.4	(0.4)	78.5	(0.4)	.12
Arm strength (kg)	27.5	(0.5)	26.9	(0.4)	.32
Aerobic fitness[d] (mL/min/kg)	31.6	(0.4)	31.8	(0.3)	.70

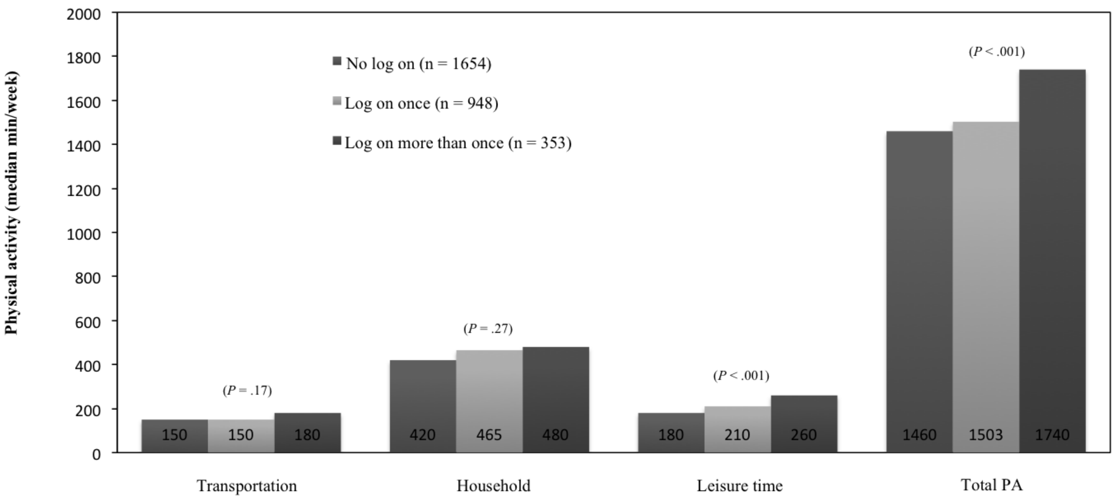
[a] Variables are shown as mean (SE).

[b] Independent *t* test for difference between groups.

[c] Body mass index.

[d] Aerobic fitness total either from watt-max or 1-point test.

Figure 2. Physical activity (PA) in three domains and total PA at 6 months by frequency of log-ons to intervention website in the website group. Kruskal-Wallis test for differences between the groups (Denmark, 2008).



Website Users

We found that 22.0% (694/3156) of the website group logged on to the website once and only 7.0% (222/3159) logged on more than once (Table 6). The question about website use was answered by 3245 participants, of whom 73 (2%) reported that they had logged on to the website, even though reports from the website company responsible for the website showed that they did not log on to the website. We excluded these participants in the analyses of website users. According to 70.0% (710/1014) of the respondents in the website group, forgetfulness was the main reason for not logging on to the website (data not shown).

In the group of participants who did log on to the website, 31.4% (318/1014) believed that the intervention helped them to increase PA. When we analyzed this group separately, we did not find significant changes in PA from baseline to follow-up (data not shown). The website's PA adviser received few questions and the participant forum was not used at all.

Analysis of the active users of the intervention in relation to age group, educational level, motivation, and sex showed that participants in the age group 45–64 years (odds ratio 1.6, 95% confidence interval 1.2–2.1) and 65+ years (odds ratio 2.0, 95% confidence interval 1.4–2.8) were more likely to log on more than once than were those in the age group 18–44 years. Motivated participants (odds ratio 1.5, 95% confidence interval 1.0–2.1) were more likely than nonmotivated participants to log on more than once.

Table 6. Use of the intervention website at 6-month follow-up in the website group (Denmark, 2008)[a].

Website use	n	%
How often did you use the website during the last 6 months? (n = 3159)		
I have not logged on to the website	2243	71.00%
I have logged on to the website once	694	22.0%
I have logged on to the website several times	159	5.0%
I have logged on to the website several times and made a personal profile	63	2%
Has the website helped you to increase your physical activity level? (n = 1014)		
Not at all	509	50.2%
Yes, a little	246	24.3%
Yes, a lot or some	72	7%
Don't know	187	18.4%

[a]Percentages may not sum to 100% due to rounding.

Discussion

The present study evaluated the effectiveness of a Web-based intervention to increase PA and improve health among physically inactive persons in a real-life setting. At follow-up we did not find any significant differences in PA and health measurements between the website and control groups. Nevertheless, participants in the website group increased their leisure-time PA and total PA in minutes/week according to active use of the intervention website.

One of the fundamental methodological problems in eHealth trials is that a proportion of people in the intervention group will not use the intervention or will use it only sparingly [35]. In this study, 7% logged on more than once and only 2% used the website as intended in the website group. With such a low use of the intervention website, a null finding in PA and health in the website group is not surprising. We recruited the participants from a large health examination survey. In the planning phase this seemed an ideal and easy way to enroll participants, but the percentage of nonusers in the study proved this wrong. Embedding the intervention in a large survey probably influenced the usage rate considerably. As stated, the main reason for nonuse in the intervention group was that participants simply forgot about the intervention. In a review of 15 Web-based PA interventions [22], the attrition rate ranged between 7% and 69% with an overall attrition of 27%, which is much lower than in the present study. A key feature of successful Web-based PA interventions is that studies enrolled self-selected volunteers [36-42]. It is likely that the participation of these highly motivated individuals resulted in higher retention and usage rates. Surprisingly, some participants in our study reported that they had logged on to the intervention website, even though reports showed that they never logged on. We believe that this was because the intervention was embedded in the survey, which could have led to confusion in some of the participants. The mean age was 50 years and the low computer skills among older participants might also be a reason for the above.

Other reasons that may explain the high number of nonusers in the study could be that Web-based interventions require more active participation than other types of mass media interventions, such as print material. To read the suggestions and use the website tools, participants have to sit at a computer and log on to the intervention. This kind of active participation requires time and effort from the participants, which may attract only highly motivated individuals [17]. If discontinuing the intervention, users are not obliged to explain why. In addition, a behavioral gap between positive intentions and behavior change has been demonstrated [43], which means that, despite

53% of the participants in this study intending to be more physically active, many did not act on this desire.

In this study we found an increase of leisure-time PA and total PA per week for the active users of the intervention. For these active and motivated participants, single-time feedback was sufficient to change PA level. The results are very much in line with a study by Smeets et al, who found that single-tailored feedback did not have an impact on the study population in a Dutch Web-based intervention [44]. However, when analyzing motivated users separately, Smeets and colleagues found an increase in PA level. Johnson and Wardle [37] examined the effect of a commercial Web-based weight loss program, using self-monitoring and feedback, on diet and PA and found that participants in the study lost a significant amount of weight. However, the study included only individuals who paid at least a month's subscription and recorded weight spanning at least 28 days, supporting that self-engagement and true users can achieve a positive effect using a Web-based intervention.

A review of PA interventions found that studies with more supervision and contact through texting and email with participants were more successful and more often reported positive outcomes on PA than did studies with few contacts [22]. Research has shown that forming a new habit (ie, automaticity for the behavior) requires a large number of repetitions [45]. A single-time intervention may raise awareness and interest but, to progress further, additional intervention interactions are needed. An approach differing from the single-feedback and static website was seen in a successful Web-based tobacco cessation intervention [46]. There, the authors organized the program content into multiple pieces that were made available to the participants sequentially and for a restricted period. In this way, participants progressed through a predetermined sequence of modules with restricted freedom, referred to as tunneling. Furthermore, the program combined email, Web, interactive voice response, and texting, and for 54 weeks comprised more than 400 contacts. Because a change in behavior and increase in PA require great effort from unmotivated participants, it is possible that single-tailored feedback is insufficient to change behavior and is simply too brief an intervention. One Web-based PA intervention [47] examined the relationship between the individual and his or her environment and found a positive effect by including environmental information in the feedback, especially among overweight participants. This approach requires more detailed tailoring but should be considered in future interventions, as the outdoor environment and facilities nearby set the scene for PA. The results of this intervention show that the step from intention to action (ie, signing up to an intervention and using it in practice) should receive considerably more attention to improve the exposure or use of the intervention.

We did not find any improvements in health measurements at 3-month follow-up in the website group compared with the control group, not even when dividing participants into three subgroups according to their use of the intervention website. The possible increase in PA might not have been sufficient to improve health measurements in already-healthy adults if the increase was not in moderate to vigorous activity and for a considerable amount of time. Further, more than 3 months may be necessary to allow for physiological changes to occur [48].

Some methodological limitations must be considered in this study. The IPAQ was developed to estimate the PA of individuals in different domains. The validity of the IPAQ as a tool to measure changes in PA behavior may be questioned. Nonetheless, it has been used in several PA Web-based interventions [36,41,42,49-51]. Pedersen et al [52] also used the IPAQ in a workplace intervention study and could not detect a 1 hour/week change in PA at work, even though participants were being monitored and the PA 1 hour/week actually took place. Their use of the IPAQ may therefore have led to an underestimation of the effect of the intervention.

We found that active users of the intervention achieved a positive effect on PA. However, conclusions in randomized controlled trials should be drawn from consideration of differences between groups rather than in a subsample of the intervention group [53], since the validity of between-group comparisons established by randomization is not preserved in this analysis.

The high number of participants is a strength of this study, as subgroup analyses were possible. We found that the age groups 44–65 and 65+ years and motivated users were more likely to log on to the intervention website. Nevertheless, we found no change in PA in these groups despite their greater likelihood of logging on to the intervention.

We rate the external validity of this study high, as we used intention-to-treat analyses and recruited participants from a generalizable population, which enables assessment of the preventive potential if translated into practice.

The relatively high percentage of participants who were lost to follow-up is a limitation, although dropout was equally distributed between the website group and the control group and is thus not expected to influence the results.

Conclusion

Web-based research is still in an early stage, with many lessons to be learned. The question is how we can move forward and develop interventions that can change PA behavior. The finding in this study suggests that active users of a Web-based intervention can achieve a positive effect. However, for unmotivated users, single-tailored feedback may be too brief. Future research should focus on the step from intention to action and on developing more sophisticated interventions. This is seen in Web-based smoking cession interventions, which combine different types of media and have many contacts, and thereby have the potential to reach both motivated and unmotivated sedentary individuals.

Acknowledgments

The authors are grateful to the participants and the participating municipalities. The authors would also like to thank Morten Zacho for help with development of the intervention website.

This study was funded by TrygFonden, Denmark.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screen shoots recruitment of participants and intervention website.

[PDF File (Adobe PDF File), 355B - jmir_v14i5e145_app1.pdf]

Multimedia Appendix 2

CONSORT E-Health Checklist V1.6.1 [54].

[PDF File (Adobe PDF File), 587KB - jmir_v14i5e145_app2.pdf]

References

1. Bauman AE. Updating the evidence that physical activity is good for health: an epidemiological review 2000-2003. J Sci Med Sport 2004 Apr;7(1 Suppl):6-19. [Medline: 15214597]
2. Sesso HD, Paffenbarger RS, Lee IM. Physical activity and coronary heart disease in men: The Harvard Alumni Health Study. Circulation 2000 Aug 29;102(9):975-980 [FREE Full text] [Medline: 10961960]
3. Sofi F, Capalbo A, Cesari F, Abbate R, Gensini GF. Physical activity during leisure time and primary prevention of coronary heart disease: an updated meta-analysis of cohort studies. Eur J Cardiovasc Prev Rehabil 2008 Jun;15(3):247-257. [doi: 10.1097/HJR.0b013e3282f232ac] [Medline: 18525378]
4. Physical Activity Guidelines Advisory Committee report, 2008. To the Secretary of Health and Human Services. Part A: executive summary. Nutr Rev 2009 Feb;67(2):114-120. [doi: 10.1111/j.1753-4887.2008.00136.x] [Medline: 19178654]
5. Kesäniemi A, Riddoch CJ, Reeder B, Blair SN, Sørensen T. Advancing the future of physical activity guidelines in Canada: an independent expert panel interpretation of the evidence. Int J Behav Nutr Phys Act 2010;7:41 [FREE Full text] [doi: 10.1186/1479-5868-7-41] [Medline: 20459785]
6. Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, Franklin BA, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Med Sci Sports Exerc 2007 Aug;39(8):1423-1434. [doi: 10.1249/mss.0b013e3180616b27] [Medline: 17762377]
7. National Institute of Public Health. Danish Health Interview Survey 2005: Questionnaire for Personal Interviews With Response Frequencies. Copenhagen: National Institute of Public Health; 2006.
8. Jones DA, Ainsworth BE, Croft JB, Macera CA, Lloyd EE, Yusuf HR. Moderate leisure-time physical activity: who is meeting the public health recommendations? A national cross-sectional study. Arch Fam Med 1998;7(3):285-289. [Medline: 9596466]
9. Rooke S, Thorsteinsson E, Karpin A, Copeland J, Allsop D. Computer-delivered interventions for alcohol and tobacco use: a meta-analysis. Addiction 2010 Aug;105(8):1381-1390. [doi: 10.1111/j.1360-0443.2010.02975.x] [Medline: 20528806]
10. 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. [doi: 10.1016/j.ypmed.2008.02.014] [Medline: 18403003]
11. 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]
12. Kirsch SE, Lewis FM. Using the World Wide Web in health-related intervention research. A review of controlled trials. Comput Inform Nurs 2004;22(1):8-18. [Medline: 15069844]
13. Eysenbach G. Issues in evaluating health websites in an Internet-based randomized controlled trial. J Med Internet Res 2002 Dec;4(3):E17 [FREE Full text] [doi: 10.2196/jmir.4.3.e17] [Medline: 12554548]
14. Cummins CO, Evers KE, Johnson JL, Paiva A, Prochaska JO, Prochaska JM. Assessing stage of change and informed decision making for Internet participation in health promotion and disease management. Manag Care Interface 2004 Aug;17(8):27-32. [Medline: 15471108]
15. Brug J, Oenema A, Kroeze W, Raat H. The internet and nutrition education: challenges and opportunities. Eur J Clin Nutr 2005 Aug;59 Suppl 1:S130-7; discussion S138. [doi: 10.1038/sj.ejcn.1602186] [Medline: 16052181]

16. Skinner CS, Campbell MK, Rimer BK, Curry S, Prochaska JO. How effective is tailored print communication? Ann Behav Med 1999;21(4):290-298. [Medline: 10721435]
17. Oenema A, Tan F, Brug J. Short-term efficacy of a web-based computer-tailored nutrition intervention: main effects and mediators. Ann Behav Med 2005 Feb;29(1):54-63. [doi: 10.1207/s15324796abm2901_8] [Medline: 15677301]
18. Kreuter MW, Wray RJ. Tailored and targeted health communication: strategies for enhancing information relevance. Am J Health Behav 2003;27 Suppl 3:S227-S232. [Medline: 14672383]
19. de Vries H, Brug J. Computer-tailored interventions motivating people to adopt health promoting behaviours: introduction to a new approach. Patient Educ Couns 1999 Feb;36(2):99-105. [Medline: 10223015]
20. Enwald HP, Huotari ML. Preventing the obesity epidemic by second generation tailored health communication: an interdisciplinary review. J Med Internet Res 2010;12(2):e24 [FREE Full text] [doi: 10.2196/jmir.1409] [Medline: 20584698]
21. Marcus BH, Ciccolo JT, Sciamanna CN. Using electronic/computer interventions to promote physical activity. Br J Sports Med 2009 Feb;43(2):102-105. [doi: 10.1136/bjsm.2008.053744] [Medline: 19052143]
22. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
23. Neville LM, O'Hara B, Milat A. Computer-tailored physical activity behavior change interventions targeting adults: a systematic review. Int J Behav Nutr Phys Act 2009;6:30 [FREE Full text] [doi: 10.1186/1479-5868-6-30] [Medline: 19490649]
24. Eriksen L, Grønbæk M, Helge JW, Tolstrup JS, Curtis T. The Danish Health Examination Survey 2007-2008 (DANHES 2007-2008). Scand J Public Health 2011 Mar;39(2):203-211. [doi: 10.1177/1403494810393557] [Medline: 21257645]
25. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol 1983 Jun;51(3):390-395. [Medline: 6863699]
26. Ajzen I. Action-control: from cognition to behavior. In: Kuhl J, Beckman J, editors. In: Kuhl J, Beckman J, editors. From Intentions to Actions: A Theory of Planned Behavior. Heidelberg: Springer; 1985:11-39.
27. National Board of Health D. Physical Activity: A Handbook on Pevention and Treatment [in Danish]. Copenhagen: National Board of Health D; 2003.
28. Andersen LB. A maximal cycle exercise protocol to predict maximal oxygen uptake. Scand J Med Sci Sports 1995 Jun;5(3):143-146. [Medline: 7552756]
29. Åstrans PO, Rodahl K. Evaluation of physical performance on the basis of tests. In: Provenzano MD, editor. Provenzano MD, editor. Textbook of Work Physiology: Physiological Bases of Exercise. New York, NY: McGraw-Hill; 1986:354-390.
30. Ross R, Dagnone D, Jones PJ, Smith H, Paddags A, Hudson R, et al. Reduction in obesity and related comorbid conditions after diet-induced weight loss or exercise-induced weight loss in men. A randomized, controlled trial. Ann Intern Med 2000 Jul 18;133(2):92-103. [Medline: 10896648]
31. Ross R, Janssen I, Dawson J, Kungl AM, Kuk JL, Wong SL, et al. Exercise-induced reduction in obesity and insulin resistance in women: a randomized controlled trial. Obes Res 2004 May;12(5):789-798. [doi: 10.1038/oby.2004.95] [Medline: 15166299]
32. Nordby P, Auerbach PL, Rosenkilde M, Kristiansen L, Thomasen JR, Rygaard L, et al. Endurance Training Per Se Increases Metabolic Health in Young, Moderately Overweight Men. Obesity (Silver Spring) 2012 Mar 22. [doi: 10.1038/oby.2012.70] [Medline: 22436841]
33. 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]
34. The IPAQ Group. Karolinska Institute. 2005 Nov. Guidelines for Data Processing and Analysis of the International Physical Activity Questionnaire (IPAQ): Short and Long Forms URL: http://www.ipaq.ki.se./scoring.pdf [accessed 2012-10-02] [WebCite Cache ID 6B7PZhzue]
35. 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]
36. Hurling R, Catt M, Boni MD, Fairley BW, Hurst T, Murray P, et al. Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. J Med Internet Res 2007;9(2):e7 [FREE Full text] [doi: 10.2196/jmir.9.2.e7] [Medline: 17478409]
37. Johnson F, Wardle J. The association between weight loss and engagement with a web-based food and exercise diary in a commercial weight loss programme: a retrospective analysis. Int J Behav Nutr Phys Act 2011;8:83 [FREE Full text] [doi: 10.1186/1479-5868-8-83] [Medline: 21810222]
38. Kypri K, McAnally HM. Randomized controlled trial of a web-based primary care intervention for multiple health risk behaviors. Prev Med 2005;41(3-4):761-766. [doi: 10.1016/j.ypmed.2005.07.010] [Medline: 16120456]
39. Marcus BH, Napolitano MA, King AC, Lewis BA, Whiteley JA, Albrecht A, et al. Telephone versus print delivery of an individualized motivationally tailored physical activity intervention: Project STRIDE. Health Psychol 2007 Jul;26(4):401-409. [doi: 10.1037/0278-6133.26.4.401] [Medline: 17605559]
40. Napolitano MA, Fotheringham M, Tate D, Sciamanna C, Leslie E, Owen N, et al. Evaluation of an internet-based physical activity intervention: a preliminary investigation. Ann Behav Med 2003;25(2):92-99. [Medline: 12704010]

41. Spittaels H, De Bourdeaudhuij I, Vandelanotte C. Evaluation of a website-delivered computer-tailored intervention for increasing physical activity in the general population. Prev Med 2007 Mar;44(3):209-217. [doi: 10.1016/j.ypmed.2006.11.010] [Medline: 17197015]
42. Vandelanotte C, De Bourdeaudhuij I, Sallis JF, Spittaels H, Brug J. Efficacy of sequential or simultaneous interactive computer-tailored interventions for increasing physical activity and decreasing fat intake. Ann Behav Med 2005 Apr;29(2):138-146. [doi: 10.1207/s15324796abm2902_8] [Medline: 15823787]
43. Sheeran P, Trafimow D, Finlay KA, Norman P. Evidence that the type of person affects the strength of the perceived behavioural control-intention relationship. Br J Soc Psychol 2002 Jun;41(Pt 2):253-270. [Medline: 12133227]
44. Smeets T, Brug J, de Vries H. Effects of tailoring health messages on physical activity. Health Educ Res 2008 Jun;23(3):402-413 [FREE Full text] [doi: 10.1093/her/cyl101] [Medline: 17032705]
45. Lally P, van Jaarsveld CHM, Potts HWW, Wardle J. How are habits formed: modelling habit formation in the real world. Eur J Soc Psychol 2010;40(6):998-1009. [doi: 10.1002/ejsp.674]
46. Brendryen H, Kraft P. Happy ending: a randomized controlled trial of a digital multi-media smoking cessation intervention. Addiction 2008 Mar;103(3):478-84; discussion 485. [doi: 10.1111/j.1360-0443.2007.02119.x] [Medline: 18269367]
47. van Stralen MM, de Vries H, Mudde AN, Bolman C, Lechner L. The long-term efficacy of two computer-tailored physical activity interventions for older adults: main effects and mediators. Health Psychol 2011 Jul;30(4):442-452. [doi: 10.1037/a0023579] [Medline: 21639638]
48. Murphy MH, Nevill AM, Murtagh EM, Holder RL. The effect of walking on fitness, fatness and resting blood pressure: a meta-analysis of randomised, controlled trials. Prev Med 2007 May;44(5):377-385. [doi: 10.1016/j.ypmed.2006.12.008] [Medline: 17275896]
49. Vandelanotte C, De Bourdeaudhuij I, Brug J. Two-year follow-up of sequential and simultaneous interactive computer-tailored interventions for increasing physical activity and decreasing fat intake. Ann Behav Med 2007 Apr;33(2):213-219. [doi: 10.1080/08836610701310086] [Medline: 17447874]
50. Spittaels H, De Bourdeaudhuij I, Brug J, Vandelanotte C. Effectiveness of an online computer-tailored physical activity intervention in a real-life setting. Health Educ Res 2007 Jun;22(3):385-396 [FREE Full text] [doi: 10.1093/her/cyl096] [Medline: 16971674]
51. Marshall AL, Leslie ER, Bauman AE, Marcus BH, Owen N. Print versus website physical activity programs: a randomized trial. Am J Prev Med 2003 Aug;25(2):88-94. [Medline: 12880874]
52. Pedersen MT, Blangsted AK, Andersen LL, Jørgensen MB, Hansen EA, Sjøgaard G. The effect of worksite physical activity intervention on physical capacity, health, and productivity: a 1-year randomized controlled trial. J Occup Environ Med 2009 Jul;51(7):759-770. [doi: 10.1097/JOM.0b013e3181a8663a] [Medline: 19528834]
53. Ware JH. Interpreting incomplete data in studies of diet and weight loss. N Engl J Med 2003 May 22;348(21):2136-2137. [doi: 10.1056/NEJMe030054] [Medline: 12761370]
54. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

BMI: body mass index
DANHES: Danish Health Examination Survey 2007–2008
IPAQ: International physical activity questionnaire
PA: physical activity

Edited by G Eysenbach; submitted 15.03.12; peer-reviewed by M Wanner; comments to author 08.05.12; revised version received 15.06.12; accepted 11.07.12; published 30.10.12.

Please cite as:
Hansen AW, Grønbæk M, Helge JW, Severin M, Curtis T, Tolstrup JS
Effect of a Web-Based Intervention to Promote Physical Activity and Improve Health Among Physically Inactive Adults: A Population-Based Randomized Controlled Trial
J Med Internet Res 2012;14(5):e145
URL: http://www.jmir.org/2012/5/e145/
doi:10.2196/jmir.2109
PMID:23111127

Original Paper

Effectiveness of a Proactive Mail-Based Alcohol Internet Intervention for University Students: Dismantling the Assessment and Feedback Components in a Randomized Controlled Trial

Preben Bendtsen[1], PhD; Jim McCambridge[2], PhD; Marcus Bendtsen[1,3], BSC; Nadine Karlsson[1], PhD; Per Nilsen[1], PhD

[1]Department of Medical and Health Sciences, Faculty of Health Sciences, Linköping University, Linköping, Sweden
[2]Faculty of Public Health & Policy, London School of Hygiene and Tropical Medicine, London, United Kingdom
[3]Department of Computer and Information Science, Institute of Technology, Linköping University, Linköping, Sweden

Corresponding Author:
Preben Bendtsen, PhD
Department of Medical and Health Sciences
Faculty of Health Sciences
Linköping University
Hälsouniversitetet
Kansliet
Linköping, SE-581 83
Sweden
Phone: 46 702324615
Fax: 46 13 14 94 03
Email: preben.bendtsen@liu.se

Abstract

Background: University students in Sweden routinely receive proactive mail-based alcohol Internet interventions sent from student health services. This intervention provides personalized normative feedback on alcohol consumption with suggestions on how to decrease drinking. Earlier feasibility trials by our group and others have examined effectiveness in simple parallel-groups designs.

Objective: To evaluate the effectiveness of electronic screening and brief intervention, using a randomized controlled trial design that takes account of baseline assessment reactivity (and other possible effects of the research process) due to the similarity between the intervention and assessment content. The design of the study allowed for exploration of the magnitude of the assessment effects per se.

Methods: This trial used a dismantling design and randomly assigned 5227 students to 3 groups: (1) routine practice assessment and feedback, (2) assessment-only without feedback, and (3) neither assessment nor feedback. At baseline all participants were blinded to study participation, with no contact being made with group 3. We approached students 2 months later to participate in a cross-sectional alcohol survey. All interventions were fully automated and did not have any human involvement. All data used in the analysis were based on self-assessment using questionnaires. The participants were unaware that they were participating in a trial and thus were also blinded to which group they were randomly assigned.

Results: Overall, 44.69% (n = 2336) of those targeted for study completed follow-up. Attrition was similar in groups 1 (697/1742, 40.01%) and 2 (737/1742, 42.31% retained) and lower in group 3 (902/1743, 51.75% retained). Intention-to-treat analyses among all participants regardless of their baseline drinking status revealed no differences between groups in all alcohol parameters at the 2-month follow-up. Per-protocol analyses of groups 1 and 2 among those who accepted the email intervention (36.2% of the students who were offered the intervention in group 1 and 37.3% of the students in group2) and who were risky drinkers at baseline (60.7% follow-up rate in group 1 and 63.5% in group 2) suggested possible small beneficial effects on weekly consumption attributable to feedback.

Conclusions: This approach to outcome evaluation is highly conservative, and small benefits may follow the actual uptake of feedback intervention in students who are risky drinkers, the precise target group.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN): 24735383; http://www.controlled-trials.com/ISRCTN24735383 (Archived by WebCite at http://www.webcitation.org/6Awq7gjXG)

(J Med Internet Res 2012;14(5):e142) doi:10.2196/jmir.2062

KEYWORDS

Alcohol drinking; Web-based intervention; proactive intervention; university students

Introduction

University entrance is associated with increased levels of heavy drinking. Alcohol is often used to promote social integration despite the potential negative consequences of excessive drinking among students. A survey of more than 18,000 students aged 17 to 30 years in 21 countries found that substantial numbers of them used alcohol, ranging from 29% of men and 6% of women in South Africa to 95% of men and 93% of women in Ireland [1]. Young adult students in Sweden also show evidence of high alcohol consumption. A survey conducted with 4575 undergraduate university students on four campuses in Sweden found that 96% of students had consumed alcohol in the preceding 12 months [2]. A recent study surveyed 1585 first-year students in Linköping University and found that heavy episodic (binge) drinking was reported by 51% of the women and 71% of the men [3].

Several studies have assessed the effectiveness of various forms of Internet-based screening, combined with brief interventions, to change drinking behaviors, as shown for example by Boon et al [4] and Saitz et al [5]. A recent Cochrane review by Moreira and colleagues [6] compared the effectiveness of various forms of social norms interventions in university and college settings involving electronic screening and brief intervention (e-SBI), mailed feedback on pen-and-paper screening, individual face-to-face feedback, and group face-to-face feedback. Evidence indicates that e-SBI methods influence a wider range of factors related to alcohol habits, including peak blood alcohol concentration, drinking frequency, and drinking quantity, than other delivery methods and are less expensive to use [6].

The use of e-SBI has to a great extent emerged as an efficient approach to reaching large numbers of adolescents as a result of high levels of Internet use among young people. Advantages of the use of Internet-delivered interventions compared with face-to-face consultations include greater reach and implementation, higher consistency of intervention content, and closer matching of intervention to patient characteristics and recommended guidelines [7,8]. The psychometric properties of existing screening instruments when administered online have been found to be reliable [9].

Many earlier studies of e-SBI among university students have focused on the effectiveness of a normative feedback on students' drinking behaviors [10-15]. A meta-analysis conducted by Carey et al [16] found that students receiving personalized normative feedback demonstrated significant reductions in harmful alcohol-related behaviors. Overall, personalized normative interventions appear to be effective in reducing alcohol use and related problems among university and college students when the student's own drinking is compared with that of other students, of the same sex and age, at the same university or college.

More research is needed on many aspects of e-SBI in student populations. The majority of the research has been conducted in the United States. Many previous studies have required respondents to participate in e-SBIs taking place in controlled settings, rather than allowing students to access e-SBIs using their own computers [17,18]. Only a few published studies have described projects that made more comprehensive use of electronic media, by recruiting large numbers of participants via email, and having participants complete e-SBIs at their own convenience, using their own or others' computers. A substantial number of e-SBI projects, including our own studies [3,19], have been more feasibility studies and were not performed as full-scale randomized controlled trials. One exception is several randomized controlled trials performed in Australia by Kypri et al, including a large-scale study of 13,000 university students [20].

Another area where more knowledge is needed concerns whether the assessment or screening of alcohol consumption per se reduces drinking [21-26]. Several reviews of brief alcohol interventions have noted unexpected levels of reductions in drinking among control groups [27,28]. Indeed it is not uncommon to find a 20% alcohol consumption reduction in control groups who do not receive an evaluated intervention. Such reductions have usually been explained by regression to the mean, the effects of natural variation in people's drinking patterns, and by a possible inadvertent intervention effect of the assessment and research procedures [21]. If students can be influenced to reduce their drinking at hazardous or harmful levels with simple screening, large-scale implementation of screening surveys among university students might have a considerable public health impact, with no need for a more elaborate normative feedback intervention.

Alcohol screening involves answering a series of questions and is thus a form of brief assessment. A recently published systematic review of randomized evaluations of assessment effects in brief intervention found somewhat equivocal evidence of assessment reactivity when studies with adults, some of which identified no brief intervention effects, were included [26]. When attention was restricted to university students, however, stronger evidence was obtained, confirming small assessment effects as a result of answering questions in both interviews and self-completion questionnaires, including online questionnaires. These findings are relevant to the possibility of developing brief assessment-based interventions and also suggest that previous studies may have been biased by contamination of control groups. University students may possibly be more receptive to assessment effects than other populations.

This study addressed several of the identified gaps in previous research. The overall aim of this study was to evaluate the effectiveness of e-SBI, using an randomized controlled trial design that takes account of baseline assessment reactivity (and other possible effects of the research process) due to the similarity between the intervention and assessment content. The design of the study allowed exploration of the magnitude of the assessment effects per se. More specifically, the study tested 4 main hypotheses as follows:

1. Alcohol Use Disorders Identification Test (AUDIT) scores in the group that was assessed at baseline and received feedback (group 1) will be lower than in the group that was not contacted at baseline (group 3), providing a test of the effects of universal e-SBI provision in an unselected population of university students.

2. AUDIT scores in the baseline assessment-only group (group 2) will be lower than in the no-contact baseline group (group 3), providing a test of the effects of assessment-only in an unselected population of university students.

3. AUDIT scores among risky drinkers in group 1 will be lower than in group 2, providing a test of the effects of adding feedback to assessment-only among risky drinkers who participated at baseline.

4. AUDIT scores in group 1 as a whole will be lower than in group 2 as a whole, providing a test of the effects of adding feedback to assessment-only in an unselected population of university students

Methods

Study Setting and Population

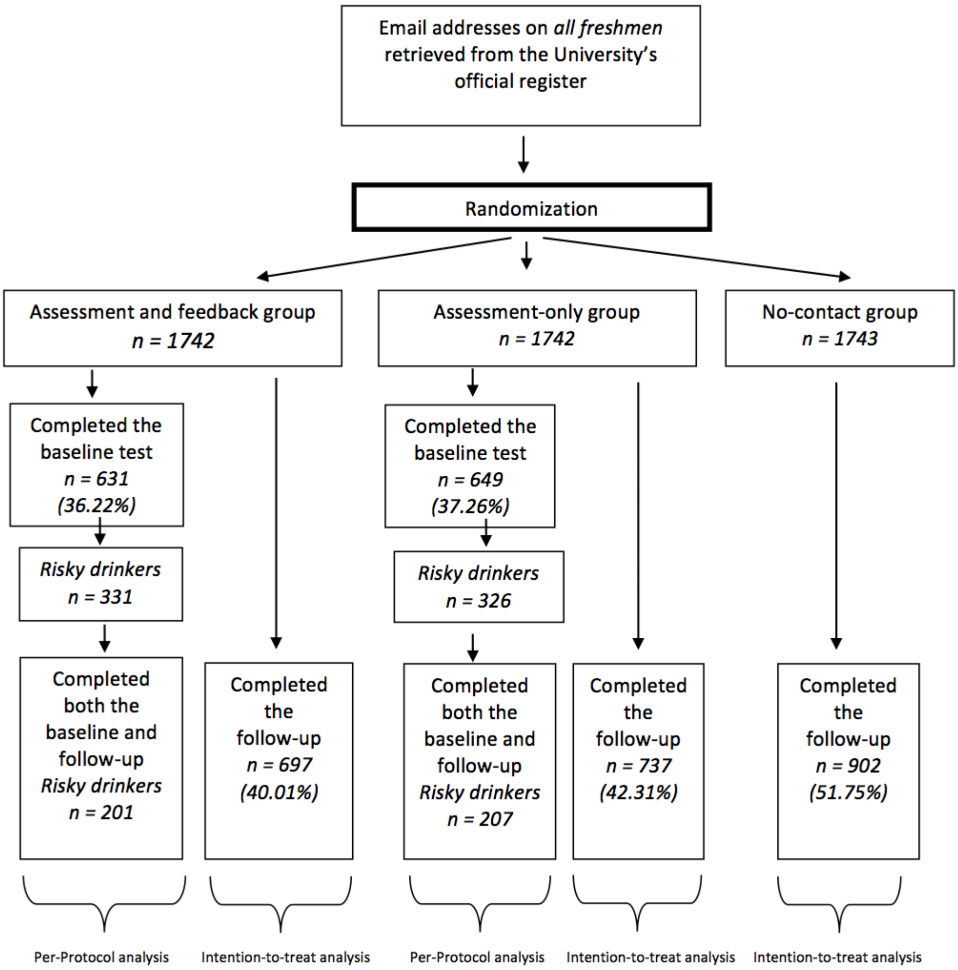
The study was performed at Linköping University in the mid to southern part of Sweden among all 5227 freshmen entering the university. Every student at the university has a personal official university email address that is obligatory for the student to use, with all official mail delivered through this address. The e-SBI was performed by the individual students on personal computers at a location of their preference, usually at home.

Design of the Study

The study was performed as a three-arm parallel-groups trial in which routine provision of e-SBI (group 1) was compared with the same without feedback (ie, assessment-only; group 2) and no-contact control (group 3) study conditions. Groups 1 and 2 completed identical assessments, the sole difference between them being that group 1 was provided with normative feedback as usual, whereas participants in group 2 were simply thanked for their participation and given a link to a commonly used website concerning alcohol information with no normative feedback. Group 3 was contacted only after 2 months, at which time both groups 1 and 2 also completed outcome data collection (Figure 1).

The study was designed to protect the control group as much as possible from the possible effects of research participation [29]. Thus, participants in the control group were not aware that they were part of a trial (nor were groups 1 and 2) and were contacted only at the time of follow-up.

Figure 1. CONSORT flowchart for the study.



Randomization and Other Study Procedures

We collected email addresses from the official university register in three separate data files. All participants had a 1 in 3 probability of allocation to any particular study condition. Randomization was computerized and did not use any strata or blocks. Each participant was given a random number between 0.0 and 1.0 with 5 decimals. For instance, participant A might be given number 0.12345 and participant B might get 0.54321. The list of participants was then sorted in descending order by this number, and the list was cut into 3 equal parts (or as equal as possible depending on the number of participants).

At baseline the students received an email from the student health care service with a short greeting welcoming them to the university followed by an invitation to perform the Internet alcohol intervention. All three groups were sent an email 2 months later by the Swedish principal investigator (PB). This mail made no reference to the previous email from the student health care services and comprised an invitation to participate in an online alcohol research survey, with inclusion in a lottery draw for 100 cinema tickets as a small incentive to increase participation. The alcohol study outcomes were derived from the 10-item AUDIT questionnaire [30] in this survey, with two reminders sent 1 week apart, also containing a link to the questionnaire.

Blinding

Groups 1 and 2 were unaware that they were participating in a research study when they responded to the initial emails. Both groups were led to expect that these emails were provided as routine practice by the student health care center to help students think about their drinking. All three groups were unaware they were participating in an intervention study and that they had been randomly assigned to a study group. At follow-up, no explanation of the true nature of the study was given to students. Instead they were invited to participate in a seemingly unrelated student alcohol survey. As all study procedures were automated, the research team had no direct contact with study participants. The use of blinding and deception in this trial raised ethical issues. All students were subsequently offered the opportunity to provide feedback on their alcohol consumption at the time of the follow-up. Ethical approval for the study was given by the regional ethics committee in Linköping, Sweden (reference number: 2010/291-31).

Description of the Computerized Alcohol Intervention

The study was based on an email-based Internet alcohol intervention (e-SBI) that has been developed by the Lifestyle Intervention Research group at Linköping University [3,19]. The computerized intervention is a fully automated single-session intervention and has been used for some years as part of routine practice in the great majority of Swedish universities. Consequently the system is stable, and no bug fixes or content changes were necessary during the trial period, and we observed no downtime of the system. The participants were given an opportunity to comment on the intervention by sending an email to the company that operates the intervention. During the trial period, fewer than 10 participants commented on the content of the intervention, giving mostly positive comments.

After students received the welcoming email from the student health care service, they were sent an invitation to perform the Internet alcohol intervention by clicking on a hyperlink to the test. Two reminders were sent 1 week apart to those who had not answered and thereafter the link was closed. The link could be used only once in order to ensure that each student performed the test only once. The test included questions about average consumption day by day in a typical week during the last 3 month, frequency of binge drinking, highest blood alcohol concentration in the last 3 months (calculated by the computer based on the student's input), negative consequences related to alcohol, views on how much other students and peers drink, and questions about motivation to change. A demo version of the computerized test can be viewed at demo.livsstilstest.nu.

Group 1 students then got immediate feedback consisting of 3 statements summarizing their weekly consumption, their frequency of heavy episodic drinking, and their highest blood alcohol concentration during the last 3 months, comparing the respondent's drinking patterns against the safe drinking limits established by the Swedish National Institute of Public Health. Immediately after this followed comprehensive normative feedback with information describing the participant's alcohol use compared with that of Swedish university students and, if applicable, personalized advice concerning the need for reducing any unhealthy levels or pattern of consumption. The student viewed the feedback on screen and could print it out. In addition the student received an email with a pdf file of the feedback.

Measurements

In the baseline assessment, used in the per-protocol analysis, we calculated the average weekly consumption, in a typical week, as the number of standard glasses day by day. Frequency of heavy episodic drinking was assessed as at most once a month, 2–3 times a month, 1–2 times a week, or 3 or more times a week. Risky drinking at baseline was defined as either having risky weekly consumption (more than 14 standard glasses a week for men or more than 9 glasses a week for women, with 1 standard glass constituting 12 g of pure alcohol) or engaged in heavy episodic drinking (5 or more standard glasses for men, or 4 or more for women) 2 times a month or more often.

In the intention-to-treat analysis we assessed risky drinking at follow-up with the AUDIT questionnaire [30], asking the students about their drinking habits during the last 4 weeks. Students with an AUDIT score ≥8 for men and ≥6 for women were considered risky drinkers. AUDIT problem score was calculated as the sum of questions 7-10 of the AUDIT questionnaire, and AUDIT dependence score was calculated as the sum of questions 4-6. We used the third question in the AUDIT questionnaire concerning heavy episodic drinking separately in the analysis.

Statistical Methods

We examined differences in proportions between groups with chi-squared tests. Weekly consumption of alcohol, total AUDIT score, AUDIT problem and dependence scores were logarithmically transformed (adding a constant of unit to handle zero values) and compared between groups by Student *t* test. The absolute change in weekly alcohol consumption (in the log-transformed scale) between baseline and follow-up was compared between groups with multilevel linear regression (where occasions were nested within individuals). The relative change in weekly alcohol consumption between baseline and follow-up was compared between groups by Mann-Whitney *U* test. All tests were performed 2-sided at $P < .05$. The statistical analyses were performed using SPSS version 19 (IBM Corporation, Somers, NY, USA) and Stata version 12 (StataCorp LP, College Station, TX, USA).

Results

Overall Response Rate

The response rate to the baseline survey was similar (approximately 37%) in groups 1 and 2. At follow-up, however, attrition was higher in groups 1 and 2 than in group 3 (Table 1). Thus, the response rate for the follow-up survey was approximately 52% in group 3 compared with around 41% in group 1 and 2 ($P < .001$, $\chi^2_2 = 54.6$). The proportion who were risky drinkers at baseline was similar in groups 1 and 2 (around 56%) as well as in all groups at follow-up (around 50%) (Table 1).

Table 1. Attrition rates of all groups: group 1 (assessment and feedback), group 2 (assessment-only), and group 3 (no contact).

Timeline	Group 1	Group 2	Group 3	Total
Baseline				
No. randomly assigned	1742	1742	1743	5227
Completed baseline test, n (%)	631 (36.2%)	649 (37.3%)	NA[a]	1280 (36.7%)
Risky drinkers, n (%)[b]	331 (57.2%)	326 (55.6%)	NA	657 (56.4%)
Follow-up (for intention-to-treat analyses)				
Completed follow-up, n (%)	697 (40.0%)	737 (42.3%)	902 (51.7%)	2336 (44.7%)
Risky drinkers, n (%)[c]	354 (50.8%)	364 (49.4%)	454 (50.3%)	1172 (50.2%)
Follow-up (for per-protocol analyses)				
Risky drinkers, n (%)[c]	201 (57.4%)	207 (54.9%)	NA	408 (56.1%)

[a] Not applicable.

[b] Risky drinking was defined as having either a risky weekly consumption (>14 standard drinks/week for men or >9/week for women, with 1 standard drink constituting 12 g of pure alcohol) or engaged in a heavy episodic drinking (≥5 standard drinks for men or ≥4 for women) ≥2 times a month.

[c] Risky drinking was defined as having an Alcohol Use Disorders Identification Test (AUDIT) score of ≥8 for men and ≥6 for women.

Intention-to-Treat Analyses of All Groups

The sociodemographic characteristics of the participants in the follow-up from all three groups did not differ with regard to sex, age, and university section (see Table 2). The intention-to-treat analyses included all participants from all three groups who answered the follow-up survey. No alcohol parameters differed significantly between the three groups (see Table 3).

Table 2. Intention-to-treat analyses[a] of groups 1, 2, and 3[b] (total n = 2336): sample characteristics.

Characteristic	Group 1 (n = 697)	Group 2 (n = 737)	Group 3 (n = 902)	χ^2 (*df*)	*P* value
Gender, n (%)					
Male	324 (46.5%)	336 (45.6%)	446 (49.4%)	2.71 (2)	.26
Female	373 (53.5%)	401 (54.4%)	456 (50.6%)		
Age (years), n (%)					
18–20	308 (44.2%)	352 (47.8%)	419 (46.5%)	3.75 (4)	.44
21–25	292 (41.9%)	303 (41.1%)	363 (40.2%)		
≥26	97 (13.9%)	82 (11.1%)	120 (13.3%)		
University faculty, n (%)					
Arts and sciences	209 (30.0%)	197 (26.7%)	250 (27.7%)	4.66 (6)	.59
Technology	299 (42.9%)	344 (46.7%)	411 (45.6%)		
Education	91 (13.1%)	82 (11.1%)	115 (12.7%)		
Health sciences	98 (14.1%)	114 (15.5%)	126 (14.0%)		

[a] Including all who responded at follow-up and irrespective of drinking status and baseline participation.

[b] Group 1 (assessment and feedback), group 2 (assessment-only), and group 3 (no-contact control).

Table 3. Intention-to-treat analyses of groups 1, 2, and 3 (total n = 2336): sample characteristics at follow-up and pairwise comparisons between groups 1 and 3.

Alcohol parameter	Group 1 (n = 697)	Group 2 (n = 737)	Group 3 (n = 902)	1 vs 2 Statistic (*df*) *P* value	1 vs 3 Statistic (*df*) *P* value	2 vs 3 Statistic (*df*) *P* value
AUDIT[a] score						
Total, mean (SD)	7.3 (5.9)	6.9 (5.5)	7.3 (5.9)	1.31[b] (1432) .19	0.29[b] (1597) .77	–1.09[b] (1637) .28
≥8 (men)/6 (women) (n, %)	354 (50.8%)	364 (49.4%)	454 (50.3%)	0.28[c] (1) .6	0.03[c] (1) .9	0.14[c] (1) .7
Problem score (mean, SD)	1.8 (2.7)	1.6 (2.4)	1.8 (2.6)	1.81[b] (1432) .07	0.50[b] (1597) .62	–1.41[c] (1637) .2
Dependence score (mean, SD)	0.8 (1.4)	0.7 (1.2)	0.8 (1.4)	1.21[b] (1432) .23	–0.16[b] (1597) .87	–1.44[c] (1637) .2
Weekly consumption (g), mean (median)	79.8 (48.0)	79.7 (48.0)	86.0 (48.0)	0.55[b] (1432) .58	–0.25[b] (1597) .80	–0.83[b] (1637) .41
Frequency of monthly heavy episodic drinking						
Never	171 (24.5%)	189 (25.6%)	244 (27.1%)	2.61[c] (4) .6	4.70[c] (4) .3	0.78[c] (4) .9
Less than monthly	158 (22.7%)	171 (23.2%)	196 (21.7%)			
Monthly	249 (35.7%)	238 (32.3%)	288 (31.9%)			
Weekly	117 (16.8%)	138 (18.7%)	173 (19.2%)			
Daily/almost daily	2 (0.3%)	1 (0.1%)	1 (0.1%)			

[a] Alcohol Use Disorders Identification Test.

[b] *t* test.

[c] χ^2.

Per-Protocol Analyses of Groups 1 and 2

The per-protocol analyses of groups 1 and 2 assessed outcomes among those participants with risky drinking at baseline who also answered the follow-up survey. Thus, 201 (57.4%) participants of the 377 who provided follow-up data from group 1 also provided data from the baseline survey, and similarly 207 (54.9%) of 421 from group 2 were included in the per-protocol analysis. There were no significant differences between groups 1 and 2 in the sociodemographic characteristics of the participants in the baseline survey. Weekly alcohol consumption at baseline was similar in both group with a mean consumption of 135.9 g/week for group 1 and 133.4 g/week for group 2 (median 120 g/week in both groups). Frequency of heavy episodic drinking did not differ significantly between the two groups at baseline.

Table 4 displays the results of the per-protocol analyses, showing one statistically significant difference in the relative change in weekly consumption between baseline and follow-up (*P* = .03); the absolute change in this outcome was not statistically significant. There were no other statistically significant differences between groups 1 and 2. The majority of risky drinkers in both groups reported drinking in a nonrisky way at follow-up.

Table 4. Per-protocol analyses of group 1 (assessment and feedback) and group 2 (assessment-only) (n = 408).

	Group 1 (n = 201)	Group 2 (n = 207)	Statistic (*df*)	*P* value
Weekly alcohol consumption				
Weekly consumption[a] (g) at follow-up (g), mean (median)	131.4 (108.0)	143.3 (108.0)	–1.12 (406)[b]	.26
Absolute change in average weekly consumption (g) between baseline and follow-up, mean (median)	–4.5 (–12.0)	9.9 (0.0)	1.48[c]	.14
Relative change (%) in average weekly consumption between baseline and follow-up, mean (median)	8.3 (–14.3)	20.8 (0.0)	38487[d]	.03
Distribution of heavy episodic drinking occasions at follow-up				
Never	2 (1.0%)	3 (1.4%)	5.0 (4)[e]	.29
Less than monthly	21 (10.4%)	22 (10.6%)		
Monthly	109 (54.2%)	97 (46.9%)		
Weekly	67 (33.3%)	85 (41.1%)		
Daily or almost daily	2 (1.0%)	0 (0.0%)		
AUDIT[f] score at follow-up				
Total score, mean (SD)	11.6 (5.8)	11.0 (4.9)	0.65 (406)[b]	.52
Score ≥8 (men)/6 (women), n (%)	168 (83.6%)	177 (85.5%)	0.3 (1)[e]	.59
Problem score, mean (SD)	3.3 (3.3)	2.8 (2.9)	1.38 (406)[b]	.17
Dependence score, mean (SD)	1.5 (1.9)	1.2 (1.4)	0.75 (406)[b]	.45
Changed from risk to no risk, n (%)	118 (58.7%)	108 (52.2%)	1.8 (1)[e]	.18

[a] Calculated from Alcohol Use Disorders Identification Test (AUDIT) questionnaire questions 1 and 2.

[b] *t* test.

[c] *z* test.

[d] *U* test.

[e] χ^2.

[f] Alcohol Use Disorders Identification Test.

Discussion

This study found no between-group differences, indicating an absence of evidence of assessment or feedback benefit in the intention-to-treat analyses among unselected university students. There was one between-group difference among those who were risky drinkers at study entry in the per-protocol analysis. This potentially provides some very modest evidence of benefit attributable to receiving feedback in addition to assessment, although this should be interpreted in the context of the wider study finding of no effects.

There are three principal obstacles to making more definitive statements attesting to evidence of intervention ineffectiveness in the present study. First, the differential attrition between group 3 and groups 1 and 2 suggests problems of nonequivalence between the groups and thus bias in direct comparisons. Second, the present study was undertaken to prepare for a subsequent large trial and was thus not originally designed to produce such conclusions. Lastly, the approach taken to outcomes evaluation in which populations are randomized and compared regardless of their need for intervention should be carefully considered. Each of these issues shall be addressed in some detail.

Because of the randomized nature of this study, it can be inferred that differential attrition was caused by the earlier involvement of groups 1 and 2 than of group3 with the study. The earlier invitation to participate in the alcohol e-SBI was apparently not sufficiently different from the later alcohol survey, or the mere fact of a second alcohol-related email invitation may have interfered with the likelihood of accepting the invitation. Whatever the precise mechanism, the main implication is clear: selection bias is possible, if not likely, and outcomes for group 3 cannot be validly assumed to be directly comparable with those for groups 1 and 2 in relation to intervention effectiveness. Because of the lack of contact with group 3, the basis for making comparisons to evaluate this possibility is limited. This is restricted to the unlikelihood of data being collected at follow-up having been altered differentially between groups by the passage of time. Here there is evidence of a slight tendency for a greater proportion of male participants, and the presence of other unmeasured confounders cannot be ruled out. It should be noted that this issue of differential attrition particularly complicates randomized comparisons involving group 3 and is less of a problem for comparisons restricted to groups 1 and 2 only. A stronger conclusion can thus be drawn in relation to whether feedback added to the effects of assessment-only, and there is little evidence here that it did, notwithstanding one statistically significant difference among the per-protocol analysis outcomes.

The second reason for caution relates to the highly naturalistic Internet study context and the need to confront the difficult methodological challenges that this implies during the course of a long-term research program. Attrition has been a major source of difficulty in previous work in developing e-SBI in Linköping and other Swedish universities [31].

It is also a significant problem in the conduct of online trials in other populations as underscored by Eysenbach [32] and others [33]. The initial take-up of the routine service provision of e-SBI has been similar to that observed here and is likely to be influenced by factors such as patterns of email use, rates of risky drinking, the salience of alcohol, and interest in intervention. In Sweden, as elsewhere, there are also seasonal influences on drinking, including proximity to exams, and these complicate any analyses of change over time. Randomization, it should be noted, safeguards the validity of between-group comparisons, if attrition and other similar sources of bias are equivalently distributed between groups [34]. In the previous follow-up studies undertaken at Linköping University, less than half of those who participated at baseline did so at first follow-up, and approximately one-quarter participated in second follow-ups [19,34]. Here, rather than follow-up emails being sent by the student health care service as was done previously, blinding of participants to trial conduct was implemented. This involved an explicit attempt to separate the experience of follow-up from earlier e-SBI delivery. An email was sent by the first author (PB) requesting participation in a survey of student alcohol consumption, partially following the approach of Kypri and colleagues, who invited participation in a series of surveys at the outset and who obtained high follow-up rates [20]. As has been seen, this innovation, along with the use of cinema ticket incentives, was partially successful in restricting attrition at follow-up. It also introduced differential attrition as has been discussed. To rectify this, we have decided that in the next trial we will abbreviate the alcohol outcome measures and conceal them within a lifestyle questionnaire in the follow-up study. The overall attrition rate could be further improved with the use of stronger incentives, though this would potentially compromise the pragmatic nature of the study [35].

The third main reason for caution in drawing conclusions from the present study relates to our intention-to-treat approach to outcomes evaluation, which was highly conservative. The intervention comprised an automated email providing a means of accessing a website in an unselected population with an elevated prevalence of hazardous and harmful drinking. Thus, the intervention was delivered more widely than was necessary, as we only wished to intervene with risky drinkers. The intervention could be defined more narrowly as being delivered to those who accessed the website, with the email merely being the means of recruitment. Even if this definition is applied, the intervention would still have been accessed by students whose drinking was not risky and who would thus not have been deemed to merit individual targeting for intervention. More narrowly still, outcome evaluation could have been restricted to those whose drinking was found to be risky. The overarching problem is that a greater number of people were randomly assigned than would have been targeted for intervention. The primary rationale for proceeding in this way was that assessing eligibility, baseline data collection, and intervention delivery were all quickly integrated in 1 brief online session. We shaped our research design pragmatically around the real-world intervention opportunity, matching the research study to routine practice as it is delivered, rather than interfering with it for research purposes, which would have introduced external validity problems. This approach takes advantage of an opportunity to avoid any research participation effects that may be associated with screening and other aspects of study entry, in much the same way as cluster randomized trials can be used for this purpose. The obvious major disadvantage of this approach is that it biases hypothesis testing toward the null and thus is highly conservative. Thinking about outcomes evaluation needs to take account of these issues.

There then arises the question of the consistency of study findings with the existing literature. Put simply, there are no existing studies against which to compare our intention-to-treat findings, as none have used no-contact control groups. The per-protocol comparisons more closely reflect existing studies, and smaller between-group differences are observed. Thus, for both internal and external validity purposes, our test of the third hypothesis is particularly important. Comparisons with the existing literature also need to take account of the highly naturalistic study context. If our results are confirmed in further studies, they have important implications when considering the effectiveness of online alcohol interventions.

Our unusual study design confers many limitations, as well as strengths, some of which have already been considered. We used the AUDIT as an efficient summary measure of alcohol consumption and whether it may be hazardous or harmful. Although the AUDIT has been validated in online student contexts [9], this does not extend to use as an outcome measure in a trial. As well as uncertainty about such use, more direct behavioral measures of drinking may be better suited to universal prevention contexts. As the study was completely automated, there was no potential for subversion of randomization, nor of observer bias in ascertainment of study outcomes. The initial take-up or reach of the intervention is neither a simple strength nor limitation of this study, being part of the object of evaluation. Necessarily, the outcomes were self-reported and, although computerized data collection may minimize social desirability bias, the validity of self-reported outcome data in brief alcohol intervention trials needs to be studied. The approach used here involves deception, and therefore it is appropriate to consider whether less-ethically problematic methods could be used. For example, if we were only concerned with constraining assessment reactivity, would it not have been possible to adopt informed consent procedures and simply withhold assessment? This would indeed have been possible had we been interested only in exploring assessment reactivity effects. We are aware, however, of the potential for other research participation effects [29] and specifically wished to control for this possibility here. This need requires novel or underused approaches to research design, for example, and studies may involve avoiding informed consent [36]. As well as developing research methods in this program of study, we are very conscious of the need both to undertake ethical analyses in parallel and to undertake dedicated empirical studies to assist ethical evaluations. Further trials that provide access to large samples are likely to be useful for further substantive effectiveness trials, along with dedicated methodological and ethical studies of the issues contended with here.

Acknowledgments

The study was funded by The Swedish Council For Working Life and Social Research (FAS, in Swedish), grant number 2010-0024. JM is supported by a Wellcome Trust Research Career Development fellowship in Basic Biomedical Science (WT086516MA).

Authors' Contributions

JM and PB had the original idea for the study, obtained funding, and led its design, supported by MB and PN. PB had overall responsibility for study implementation. MB did all computer programming associated both with intervention delivery and with study data collection. PB and JM wrote the first draft of the study protocol, to which all authors contributed.

Conflicts of Interest

PB and MB own a company that has developed the e-SBI used in this study.

Multimedia Appendix 1

CONSORT ehealth checklist V1.6.1 [37].

[PDF File (Adobe PDF File), 576KB - jmir_v14i5e142_app1.pdf]

References

1. Dantzer C, Wardle J, Fuller R, Pampalone SZ, Steptoe A. International study of heavy drinking: attitudes and sociodemographic factors in university students. J Am Coll Health 2006;55(2):83-89. [doi: 10.3200/JACH.55.2.83-90] [Medline: 17017304]
2. Bullock S. Alcohol, Drugs and Student Lifestyle! A Study of the Attitudes, Beliefs and Use of Alcohol and Drugs Among Swedish University Students. Stockholm University SoRAD Research Report no 21. Stockholm: Stockholm University; 2004.
3. Andersson A, Wiréhn AB, Olvander C, Ekman DS, Bendtsen P. Alcohol use among university students in Sweden measured by an electronic screening instrument. BMC Public Health 2009;9:229 [FREE Full text] [doi: 10.1186/1471-2458-9-229] [Medline: 19594906]
4. Boon B, Risselada A, Huiberts A, Riper H, Smit F. Curbing alcohol use in male adults through computer generated personalized advice: randomized controlled trial. J Med Internet Res 2011;13(2):e43 [FREE Full text] [doi: 10.2196/jmir.1695] [Medline: 21719412]
5. Saitz R, Palfai TP, Freedner N, Winter MR, Macdonald A, Lu J, et al. Screening and brief intervention online for college students: the ihealth study. Alcohol Alcohol 2007 Feb;42(1):28-36 [FREE Full text] [doi: 10.1093/alcalc/agl092] [Medline: 17130139]
6. Moreira MT, Smith LA, Foxcroft D. Social norms interventions to reduce alcohol misuse in university or college students. Cochrane Database Syst Rev 2009(3):CD006748. [doi: 10.1002/14651858.CD006748.pub2] [Medline: 19588402]
7. Noell J, Glasgow RE. Interactive technology applications for behavioral counseling: issues and opportunities for health care settings. Am J Prev Med 1999 Nov;17(4):269-274. [Medline: 10606195]
8. Elliott JC, Carey KB, Bolles JR. Computer-based interventions for college drinking: a qualitative review. Addict Behav 2008 Aug;33(8):994-1005. [doi: 10.1016/j.addbeh.2008.03.006] [Medline: 18538484]
9. Thomas BA, McCambridge J. Comparative psychometric study of a range of hazardous drinking measures administered online in a youth population. Drug Alcohol Depend 2008 Jul 1;96(1-2):121-127. [doi: 10.1016/j.drugalcdep.2008.02.010] [Medline: 18406079]
10. Neal DJ, Carey KB. Developing discrepancy within self-regulation theory: use of personalized normative feedback and personal strivings with heavy-drinking college students. Addict Behav 2004 Feb;29(2):281-297. [Medline: 14732417]
11. Kypri K, Saunders JB, Williams SM, McGee RO, Langley JD, Cashell-Smith ML, et al. Web-based screening and brief intervention for hazardous drinking: a double-blind randomized controlled trial. Addiction 2004 Nov;99(11):1410-1417. [doi: 10.1111/j.1360-0443.2004.00847.x] [Medline: 15500594]

12. Walters ST, Vader AM, Harris TR. A controlled trial of web-based feedback for heavy drinking college students. Prev Sci 2007 Mar;8(1):83-88. [doi: 10.1007/s11121-006-0059-9] [Medline: 17136461]
13. McNally AM, Palfai TP. Brief group alcohol interventions with college students: examining motivational components. J Drug Educ 2003;33(2):159-176. [Medline: 12929707]
14. Murphy JG, Duchnick JJ, Vuchinich RE, Davison JW, Karg RS, Olson AM, et al. Relative efficacy of a brief motivational intervention for college student drinkers. Psychol Addict Behav 2001 Dec;15(4):373-379. [Medline: 11767271]
15. Spijkerman R, Roek MA, Vermulst A, Lemmers L, Huiberts A, Engels RC. Effectiveness of a web-based brief alcohol intervention and added value of normative feedback in reducing underage drinking: a randomized controlled trial. J Med Internet Res 2010;12(5):e65 [FREE Full text] [doi: 10.2196/jmir.1465] [Medline: 21169172]
16. Carey KB, Scott-Sheldon LA, Carey MP, DeMartini KS. Individual-level interventions to reduce college student drinking: a meta-analytic review. Addict Behav 2007 Nov;32(11):2469-2494. [doi: 10.1016/j.addbeh.2007.05.004] [Medline: 17590277]
17. Neighbors C, Lewis MA, Bergstrom RL, Larimer ME. Being controlled by normative influences: self-determination as a moderator of a normative feedback alcohol intervention. Health Psychol 2006 Sep;25(5):571-579. [doi: 10.1037/0278-6133.25.5.571] [Medline: 17014274]
18. Lewis MA, Neighbors C, Oster-Aaland L, Kirkeby BS, Larimer ME. Indicated prevention for incoming freshmen: personalized normative feedback and high-risk drinking. Addict Behav 2007 Nov;32(11):2495-2508. [doi: 10.1016/j.addbeh.2007.06.019] [Medline: 17658695]
19. Bendtsen P, Johansson K, Akerlind I. Feasibility of an email-based electronic screening and brief intervention (e-SBI) to college students in Sweden. Addict Behav 2006 May;31(5):777-787. [doi: 10.1016/j.addbeh.2005.06.002] [Medline: 15996827]
20. Kypri K, Hallett J, Howat P, McManus A, Maycock B, Bowe S, et al. Randomized controlled trial of proactive web-based alcohol screening and brief intervention for university students. Arch Intern Med 2009 Sep 14;169(16):1508-1514. [doi: 10.1001/archinternmed.2009.249] [Medline: 19752409]
21. Jenkins RJ, McAlaney J, McCambridge J. Change over time in alcohol consumption in control groups in brief intervention studies: systematic review and meta-regression study. Drug Alcohol Depend 2009 Feb 1;100(1-2):107-114. [doi: 10.1016/j.drugalcdep.2008.09.016] [Medline: 19041196]
22. McCambridge J, Day M. Randomized controlled trial of the effects of completing the Alcohol Use Disorders Identification Test questionnaire on self-reported hazardous drinking. Addiction 2008 Feb;103(2):241-248. [doi: 10.1111/j.1360-0443.2007.02080.x] [Medline: 18199302]
23. Walters ST, Vader AM, Harris TR, Jouriles EN. Reactivity to alcohol assessment measures: an experimental test. Addiction 2009 Aug;104(8):1305-1310. [doi: 10.1111/j.1360-0443.2009.02632.x] [Medline: 19624323]
24. Del Boca FK, Darkes J. The validity of self-reports of alcohol consumption: state of the science and challenges for research. Addiction 2003 Dec;98 Suppl 2:1-12. [Medline: 14984237]
25. McCambridge J. [Commentary] Research assessments: instruments of bias and brief interventions of the future? Addiction 2009 Aug;104(8):1311-1312. [doi: 10.1111/j.1360-0443.2009.02684.x] [Medline: 19624324]
26. McCambridge J, Kypri K. Can simply answering research questions change behaviour? Systematic review and meta analyses of brief alcohol intervention trials. PLoS One 2011;6(10):e23748 [FREE Full text] [doi: 10.1371/journal.pone.0023748] [Medline: 21998626]
27. Bien TH, Miller WR, Tonigan JS. Brief interventions for alcohol problems: a review. Addiction 1993 Mar;88(3):315-335. [Medline: 8461850]
28. Kaner EF, Beyer F, Dickinson HO, Pienaar E, Campbell F, Schlesinger C, et al. Effectiveness of brief alcohol interventions in primary care populations. Cochrane Database Syst Rev 2007(2):CD004148. [doi: 10.1002/14651858.CD004148.pub3] [Medline: 17443541]
29. Kypri K, McCambridge J, Wilson A, Attia J, Sheeran P, Bowe S, et al. Effects of Study Design and Allocation on participant behaviour--ESDA: study protocol for a randomized controlled trial. Trials 2011;12(1):42 [FREE Full text] [doi: 10.1186/1745-6215-12-42] [Medline: 21320316]
30. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. Addiction 1993 Jun;88(6):791-804. [Medline: 8329970]
31. Ekman DS, Andersson A, Nilsen P, Ståhlbrandt H, Johansson AL, Bendtsen P. Electronic screening and brief intervention for risky drinking in Swedish university students--a randomized controlled trial. Addict Behav 2011 Jun;36(6):654-659. [doi: 10.1016/j.addbeh.2011.01.015] [Medline: 21316157]
32. 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]
33. 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]

34. Ståhlbrandt H, Leifman A, Johnsson KO, Berglund M. Alcohol trajectories over three years in a Swedish residence hall student population. Int J Environ Res Public Health 2010 Apr;7(4):1432-1447. [doi: 10.3390/ijerph7041432] [Medline: 20617038]
35. Khadjesari Z, Murray E, Kalaitzaki E, White IR, McCambridge J, Thompson SG, et al. Impact and costs of incentives to reduce attrition in online trials: two randomized controlled trials. J Med Internet Res 2011;13(1):e26 [FREE Full text] [doi: 10.2196/jmir.1523] [Medline: 21371988]
36. McCambridge J, Butor-Bhavsar K, Witton J, Elbourne D. Can research assessments themselves cause bias in behaviour change trials? A systematic review of evidence from solomon 4-group studies. PLoS One 2011;6(10):e25223 [FREE Full text] [doi: 10.1371/journal.pone.0025223] [Medline: 22039407]
37. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

AUDIT: Alcohol Use Disorders Identification Test
e-SBI: electronic screening and brief intervention

Edited by G Eysenbach; submitted 20.01.12; peer-reviewed by D Rosenbloom, T Palfai; comments to author 29.04.12; revised version received 29.05.12; accepted 11.07.12; published 31.10.12.

Please cite as:
Bendtsen P, McCambridge J, Bendtsen M, Karlsson N, Nilsen P
Effectiveness of a Proactive Mail-Based Alcohol Internet Intervention for University Students: Dismantling the Assessment and Feedback Components in a Randomized Controlled Trial
J Med Internet Res 2012;14(5):e142
URL: http://www.jmir.org/2012/5/e142/
doi:10.2196/jmir.2062
PMID:23113955

Review

The Smartphone in Medicine: A Review of Current and Potential Use Among Physicians and Students

Errol Ozdalga[1*], MD; Ark Ozdalga[2*], BS; Neera Ahuja[1*], MD

[1]Stanford University Hospital, Stanford, CA, United States
[2](not affiliated with an institution), San Jose, CA, United States
[*]all authors contributed equally

Corresponding Author:
Errol Ozdalga, MD
Stanford University Hospital
MC 5209
300 Pasteur Dr
Stanford, CA, 94305
United States
Phone: 1 4084992117
Fax: 1 8885617051
Email: eozdalga@stanford.edu

Related Article:

Letter: http://www.jmir.org/2013/4/e89/

Abstract

Background: Advancements in technology have always had major impacts in medicine. The smartphone is one of the most ubiquitous and dynamic trends in communication, in which one's mobile phone can also be used for communicating via email, performing Internet searches, and using specific applications. The smartphone is one of the fastest growing sectors in the technology industry, and its impact in medicine has already been significant.

Objective: To provide a comprehensive and up-to-date summary of the role of the smartphone in medicine by highlighting the ways in which it can enhance continuing medical education, patient care, and communication. We also examine the evidence base for this technology.

Methods: We conducted a review of all published uses of the smartphone that could be applicable to the field of medicine and medical education with the exclusion of only surgical-related uses.

Results: In the 60 studies that were identified, we found many uses for the smartphone in medicine; however, we also found that very few high-quality studies exist to help us understand how best to use this technology.

Conclusions: While the smartphone's role in medicine and education appears promising and exciting, more high-quality studies are needed to better understand the role it will have in this field. We recommend popular smartphone applications for physicians that are lacking in evidence and discuss future studies to support their use.

(J Med Internet Res 2012;14(5):e128) doi:10.2196/jmir.1994

KEYWORDS

Smartphone; technology; education; medicine; telemedicine; health care; Android; iPhone; BlackBerry; mobile phone

Introduction

On April 3, 1973 the first cellular phone call was placed by a general manager at Motorola [1]. Ever since, mobile communication has drastically changed the way we work and live our lives. More recently, another technology is again driving such change: the smartphone. Faster processors, improved memory, and smaller batteries in concert with highly efficient operating systems capable of advanced functions have paved the way for applications (commonly referred to as apps) that are affecting our personal and work environments. Like other industries, the field of medicine experienced the resounding effects of the smartphone. In fact, it may be among those industries where the impact has been most profound. One market research firm estimates that 72% of US physicians use a smartphone, and the research firm expects this number to rise to 81% in 2012 [2]. In another study, 85% of medical providers working in Accreditation Council for Graduate Medical Education training programs reported use of the smartphone [3].

Today, smartphones are being manufactured by numerous companies and are one of the fastest growing sectors in the technology industry. Operating systems include Google's Android, Apple's iOS, Research in Motion's BlackBerry, Nokia's Symbian, and the Windows Phone 7 platform. From patient monitoring and diagnostics to more efficient medical education and communication, smartphones serve a vital role in the practice of medicine today. In this review, we will the available literature to understand how the smartphone has changed the field of internal medicine and medical education. We also survey the ways in which the smartphone is used to better understand how that impact might be achieved. We conclude this review with suggested apps for physicians based on anecdotal experience and suggest studies that can better answer these questions.

Methods

Database Search

We searched both Medline via PubMed and Scopus databases for the literature review. Using PubMed we searched a combination of the Medical Subject Heading *cellular phone* with the related key words *smartphone**, *smart phone**, *mobile phone*, *iPhone*, *android*, *blackberry*, and *windows mobile*. All terms were combined using the Boolean operator OR. Our Scopus search used the following keywords: *smartphone**, *smart phone**, *iPhone*, *android*, *blackberry*, and *windows mobile*, and was combined with *medicine* to narrow our results. We limited our search to journals written in English in both databases. Both searches were also limited to the last 5 years. The citations of the search results were then manually queried for eligible articles and reviews. These searches were conducted over a period of 10 months, from August 2011 to May 2012.

Search Criteria

Given the broad nature of our review, we attempted to include all articles that either exemplified particular roles for the smartphone in health care or discussed its implications. For the purposes of this article, we defined the smartphone as any cellular device that has additional functions including a camera, global positioning system (GPS), and Wi-Fi capabilities and is running one of the following operating systems: iPhone, Android, BlackBerry, or Windows Mobile. Our search criteria included any primary article or review that discussed innovative roles of the smartphone in the field of internal medicine. These roles included patient care, medical reference, and continuing education. We also sought uses of the smartphone in medical education, communication, and research. We excluded any articles that were limited to mobile phones (considered the precursor to the smartphone) or personal digital assistants. Additionally, we did not include papers within the field of surgery or the surgical subspecialties given the already broad scope of this topic. Moreover, as internal medicine physicians, we felt it would be inappropriate to comment on the uses of the smartphone in this setting.

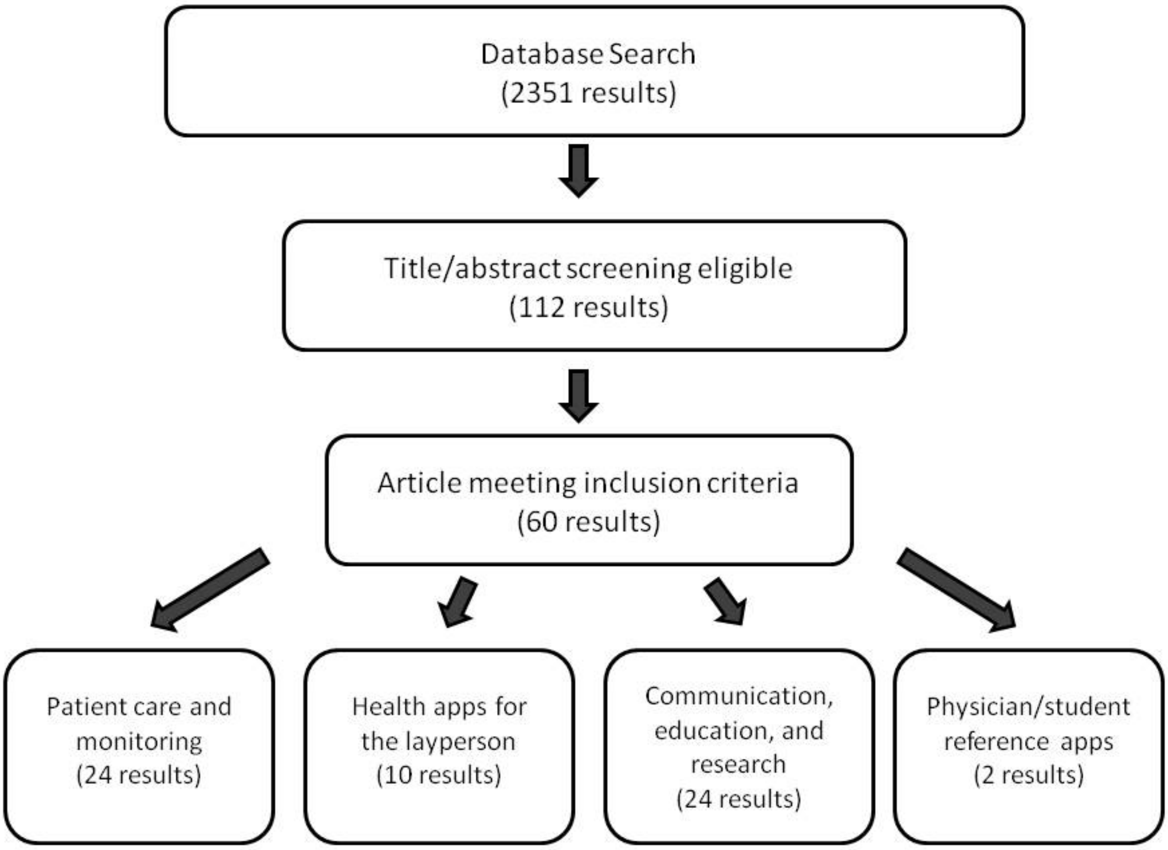
Research on the use of smartphones in medicine and medical education is limited, which hindered our development of a systematic article selection process. We did give priority to primary sources that were controlled, multicentered reports providing outcome data.

Results

Search Results

From our initial combined database search, we retrieved 2351 articles (Figure 1). A title and abstract review was conducted, from which we identified 112 articles for detailed review and added 1 article from the citation review. Of these, 60 articles met the eligibility criteria. Given the large sample size, we further analyzed these articles to subdivide them into the following categories: (1) patient care and monitoring, (2) health apps for the layperson, (3) communication, education, and research, and (4) physician or student reference apps. Of note, some papers involved both patient care and communication, but we categorized them based on the smartphone's primary purpose. For example, if the smartphone was aiding patient care via telediagnosis, then we placed the article in the category of communication.

Figure 1. Flow diagram of the process of article selection and results.



Patient Care and Monitoring

Our literature search found several examples of the use of the smartphone's features for patient monitoring. One such example involved patients with Alzheimer disease. An attempt to deal with the risk of wandering was proposed with the use of the Android app iWander [4]. The app works by using the smartphone's GPS to track the patient at all times. The patient's age, level of dementia, and home location on the GPS are input into the software. If the GPS detects that the patient is away from his or her home (for example, uncharacteristically late in the day or during inclement weather), the algorithm may predict that the patient has become confused. The app then requests that the patient manually confirm his or her status. Not providing confirmation triggers an alarm that notifies the patient's family and primary care doctor or contacts emergency personnel. It has also been suggested that identifying Alzheimer patients with depression might be possible simply by monitoring behaviors via the smartphone's functions, such as their movements using Bluetooth technology, their communication patterns, and their level of activity from the GPS. It is important to note that this app is limited by factors such as GPS and Internet reliability. Also, an older patient with mild dementia may have little ability to use such modern devices.

The smartphone has also been used in rehabilitation [5]. Using smartphones connected via Bluetooth to a single-lead electrocardiograph (ECG) device, patients who were unable to attend traditional hospital-based rehabilitation were monitored in real time through their smartphones while they exercised in their own neighborhoods [6]. This small study followed 6 patients who recently had a coronary event or angioplasty for 116 exercise sessions. Information obtained from the smartphones allowed researchers to track their patients' heart rates, single-lead rhythms, locations, altitudes, and walking speeds. This information was then used to create custom exercise regimens, leading to improved postintervention 6-minute walk tests. Patients also reported reduced depression and improved quality of life on questionnaires.

Other studies demonstrated the smartphone's potential in patient monitoring. Shoes fitted with sensors that communicate with the smartphone were used to follow the activity level of patients who have recently had a stroke [7]. The smartphone's accelerometer can be used to interpret gait and balance of patients [8-11]. Another study entailed connecting a single-lead ECG to a smartphone to diagnose and follow treatment with sleep apnea [12], providing a possible alternative to costly and labor-intensive polysomnography. One study used smartphones to promote physical activity by asking participants to routinely log their results [13].

Recognizing the challenges of a growing elderly population, one group worked on the European Union-funded project Enhanced Complete Ambient Assisted Living Experiment [14]. This project intended to build a comprehensive remote monitoring system targeted at older people with chronic diseases. Using sensors attached to garments, continuous information was monitored and collected. Data from this 3-year project (June 2009 to May 2012) intended to show how smartphone technology provides an environment where older people can maintain their independence. At the same time, the researchers hoped to provide a way to minimize health care costs through early detection of acute illnesses and a decreased need for skilled nursing homes.

Patients with type 1 diabetes are also among those who could benefit from smartphone technology, by using Diabeo [15]. Diabeo is an app that collects information such as self-measured plasma glucose, carbohydrate counts, and planned physical activity prior to making insulin dosing recommendations. Researchers in France conducted a 6-month multicenter study of 180 adult patients with type 1 diabetes with glycated hemoglobin above 8%. They found that patients using Diabeo together with telephone conversations had lower glycated hemoglobin levels than those with clinic visits. The app was used safely with no differences in hypoglycemic events.

The ability to automatically monitor patients with diabetic and heart conditions from their smartphones is being developed [16]. This technology extends to other conditions such as movement disorders or bipolar disorder [17-19]. Additionally, engineers are testing the smartphone to be used as a device for monitoring patients' balance using the phone's accelerometer [20].

The use of the smartphone as a patient-monitoring device has also been described in resource-poor countries. Smartphones used by health care workers treating malaria in rural Thailand allowed for better follow-up, medication adherence, and collection of information [21]. A similar study in Kenya allowed workers to collect data during home visits [22].

With a hands-free microphone, the smartphone has been used to record heart sounds for tracking heart rate and heart rate variability. The phone's camera along with its light-emitting diode light source has been shown to measure heart rate accurately [23]. Recently, teams have begun working on ECG recording devices that work with smartphones [24,25]. Moreover, the smartphone is being used for echocardiography [26]. MobiSante (MobiSante, Inc, Redmond, WA, USA) became the first company to design and build a US Food and Drug Administration (FDA)-approved cell phone-based medical diagnostic tool with an ultrasound probe in January 2011. A smartphone connected to a Doppler device has been used for blood flow measurement [27].

Health Apps for the Layperson

In our review we did not find clear data describing usage trends of apps for patients. Additionally, we did not find any evidence that these apps lead to wellness. Here we briefly review some apps mentioned in our literature search for laypersons.

Weight loss and fitness apps are among the most used. The apps Lose It! and Calorie Counter provide a way for people to keep track of how many calories they consume and burn for better control of their weight loss goals [28]. Based on the input information, such as the type and quantity of food consumed, these apps calculate the user's total daily caloric expenditure.

Other apps help track the amount of exercise an individual does. Using the GPS and accelerometer, phones can be turned into and navigators and pedometers [28].

Wellness apps that teach yoga are available, as are apps that focus on other forms of relaxation such as breathing [28]. Women can input the dates of their periods and body temperature to help predict ovulation. Some apps remind a patient to take his or her medication. Other apps contain an individual's important medical information such as allergies, medications, and contact phone numbers in the event of an emergency. There is also an iPhone app that offers free hearing tests [29].

Of the available data concerning the validity of apps to promote wellness, a review of obesity-related apps for diet and exercise showed that a vast majority of them rated low on a custom scoring system based on topics covered, accuracy, and other parameters [30]. Similar findings were noted with reviews of apps for alcohol abuse and smoking cessation [31,32].

One report described a method of surveying participants to create a framework from which to create an app promoting physical activity [33].

iTriage (iTriage, LLC, Denver, CO, USA) is an app that provides patients with information such as the locations of nearby emergency rooms, doctors by specialty, and other practical information [34]. It provides emergency room wait times and allows for registration via the app at participating locations. Another similar app was designed to improve diagnosis and treatment times of stroke patients [35]. ZocDoc (ZocDoc, Inc, New York, NY, USA) allows patients to conveniently make appointments with physicians who choose to use this system. Patients can view open slots and other information about participating doctors.

Communication, Education, and Research

The smartphone has been used for years in hospitals with limited network capability [36]. It also has been shown to improve communication among doctors and nurses on inpatient wards. Timely communication within hospitals remains a fundamental means by which to reduce medical errors [37]. The internal medicine program at Toronto General Hospital conducted a study using dedicated BlackBerrys for each medical team [38]. Nurses could call the team or use a Web program to send emails to these phones for less-urgent issues. Overall, surveys from residents reported improvements in communication and decreased disruption of workflow. Nurses reported decreased time spent attempting to contact physicians; however, there was no change in response time for urgent issues. Another study by this group also illustrated the efficiency of smartphones over pagers but noted a perceived increase in interruptions and weakened interprofessional relationships [39]. They also reported value in the ability to receive nonurgent messages via email; however, there has been disagreement as to what types of messages are appropriate for various communication methods [40].

Communication is also affected by integrating with electronic medical records. One company, Epic Systems (Epic Systems Corporation, Verona, WI, USA) has partnered with Apple (Apple Inc, Cupertino, CA, USA), releasing versions of Epic for both the iPhone and iPad [41]. Another app specifically designed for one particular hospital is being created to provide doctors with access to patient records from smartphones [42].

Interesting and educational patient physical findings are better documented with the use of the phone's camera. A group demonstrated that they could accurately diagnose acute stroke on brain computed tomography scans through the use of iPhones with identical accuracy to standard workstations [43]. Another study of stroke patients found comparable examinations of patients in person and via iPhone [44].

Several examples demonstrating the smartphone's role in communication can be found in developing countries with scarce resources [45]. In Africa, the amount of network coverage to send text messages with pictures ranges from 1.5% to 92.2% [46], providing an opportunity to send pictures of physical findings to aid in telediagnosis. Pictures from phone cameras of Gram stains have been sent via text messaged for remote diagnosis [47]. Video clips of limited echocardiographic studies were taken in remote Honduran villages sent via iPhones to experts for interpretation [48]. This has been reproduced with lung ultrasound [49]. Engineers have created various microscopes that attach to smartphones, providing a cost-effective and mobile way to bring more technology to poor and rural regions [50-52]. Development of point-of-care apps for human immunodeficiency virus (HIV) infection treatment to support physicians with limited HIV training in undeveloped regions is expected to minimize errors and improve outcomes [53].

Outbreaks Near Me (HealthMap, Boston Children's Hospital, Boston, MA, USA) provides information on disease outbreaks by geography [28,54]. This project, funded by Google and done in collaboration with organizations including the US Centers for Disease Control and Prevention, obtains information from multiple resources, including online news, eyewitness accounts, and official reports.

In one example of the uses of the smartphone in medical education, doctors who were trained to use a smartphone app for teaching advanced life support had significantly improved scores during cardiac arrest simulation testing [55]. A survey among medical residents in Botswana showed how a smartphone preinstalled with medical apps can be an effective way to obtain information in a resource-poor region [56].

We found two articles describing the use of the smartphone in medical research. In one example, the smartphone was used to improve data collection during trials [57]. In another, a study of falls risk in the elderly, the smartphone's accelerometer was used to help detect those at higher risk [58].

Physician and Student Reference Apps

During our literature review, we found very limited data regarding the use of reference apps by physicians or medical students. A study in 2010 claimed that over 60% of physicians surveyed felt that Epocrates (Epocrates, Inc, San Mateo, CA, USA) helped to reduce medical errors [59]. Epocrates claims that their app can help save 20 minutes of time each day for many of their users [60], but this is not supported by the evidence base.

Discussion

Here we summarize the database search identifying how the smartphone is being used in medicine. We focused on data that would either support or negate the impact of the smartphone and then surveyed the range of uses to better understand the forms in which that impact might occur.

With respect to patient care and monitoring, we found various ways of using the smartphone to monitor patients. We identified research attempting to provide evidence that the smartphone has advantages in this area; however, much of this is still in the preliminary phase. Apps such as iWander for people with dementia (see Patient Care and Monitoring above) could improve quality of life and decrease financial burden. Approximately 5.3 million Americans have Alzheimer disease, and it has been estimated that this number could quadruple by the year 2050 [61]. As we enter a new era of rising medical costs exacerbated by a growing elderly population, our health care system is looking for ways to meet the rising demand. It remains to be seen whether the smartphone can help.

We found a wide range of apps for the layperson, from wellness apps to apps that allow improved communication with health care providers. The greatest concern is the general lack of regulation and an evidence base for many of these wellness apps. Much like the general information available on websites, the content of many health-related apps is poorly scrutinized for accuracy. The FDA does not control the content of most apps; only when apps cross the line of providing direct medical advice does the FDA make approval mandatory. An example of this is the diabetes app WellDoc DiabetesManager System (WellDoc, Inc, Baltimore, MD, USA), which required FDA approval when it started providing medical advice based on input blood glucose levels. Other apps available for free download include symptom checkers, where people can input basic symptoms such as abdominal pain and get a whole list of possible causes, prompting inappropriate self-diagnosis and unneeded anxiety.

Not surprisingly, we found a larger number of articles that discussed ways in which the smartphone is improving communication on internal medicine wards. These results suggest that there may be a role for better communication between doctors and nurses; however, drawbacks such as weakened interprofessional relationships may produce new issues. We found it interesting that many people are looking into using the smartphone for remote diagnosis. It is easy to imagine the huge benefits that could be reaped in resource-poor regions of the world, but this may also lead to a change in insurance companies' reimbursement methods. In the future, patients may not need to see their physicians in person as often to get the same quality of care. However, we have not found any evidence specific to the smartphone to support this.

We also find some examples of the smartphone's use in education, such as a program that teaches cardiopulmonary resuscitation, and the use of smartphones by students in resource-poor countries. Medical schools in the United States are also beginning to use technology more in their curriculum. For example, at Stanford University's medical school all students are given an iPad to use in place of text books. Online resources are easily accessed.

As the role of the smartphone continues to grow, we can only expect that its role in medical education will expand with more institutions incorporating it into their curriculums. However, we need to have better evidence to support both its use and methods of how best to use it.

Among our categories, we found that the area most lacking in evidence is the use of smartphones for physician or student reference apps. We found only one older study looking into the effectiveness of apps to aid in the practice of medicine, with Epocrates. There is no clear reported data on usage statistics, but we presume that given the high number physicians with smartphones and the large availability of apps, many physicians are using these reference apps. A survey of health care providers showed that attitudes toward using smartphones are in general very positive [62].

Medical Reference Apps

Given the importance of medical reference apps and the paucity of published data regarding available apps and the evidence for their use, we present a list of commonly used apps and make suggestions for future research toward better understanding their utility. This list of apps derives from our anecdotal experience, for which we have given preference to apps known to have a vast database, to have reliable content, to be well respected (or contain information adapted from well-respected resources), and to have been available for many years. There are no conflicts of interest. We review some of the most popular and important apps being used to enhance continuing medical education, improve patient care, and promote communication (Table 1).

Table 1. List and description of popular medical applications for physicians.

Smartphone application	Description
Epocrates	Free up-to-date pharmacologic reference and paid medical reference.
DynaMed	Medical and pharmacologic reference
Johns Hopkins' Antibiotics Guide	Antimicrobial reference
Sanford Guide to Antimicrobial Therapy	Antimicrobial reference
Diagnosaurus	Differential diagnosis
Taber's Medical Dictionary, Stedman's Medical Dictionary, and Dorland's Medical Dictionary	Medical dictionaries and reference
Archimedes (Archimedes 360°)	Free medical calculator (Archimedes 360° is available for a fee)
AHRQ ePSS[a]	Primary care prevention
Medscape	Medical reference, news, and education
Massachusetts General Hospital's Pocket Medicine	Medical reference
Washington Manual of Medical Therapeutics	Medical reference
QuantiaMD	Medical education with interactive cases
MedPage Today	Medical news
Doximity	Social networking for physicians and physician communication

[a] Agency for Healthcare Research and Quality Electronic Preventive Services Selector.

Epocrates is well known for offering a free, up-to-date pharmacologic reference that is available for all smartphone platforms. Strengths of this program include drug dosage guidelines, adverse reactions, mechanism of action, and a drug interaction checker. Epocrates also offers an upgrade to the full version that includes more comprehensive disease and laboratory information. Another resource for pharmacologic reference is mobilePDR (PDR Network, LLC, Montvale, NJ, USA), available for free to doctors after validation of credentials.

Another resource in medical-related apps for doctors is Skyscape (Skyscape.com, Inc., Marlborough, MA, USA). Skyscape focuses on creating mobile phone apps for all health care professions and boasts over 600 apps spanning 35 specialties. They formulate popular textbooks into searchable programs on the smartphone. There are many notable apps, including Massachusetts General Hospital's Pocket Medicine and The Washington Manual of Medical Therapeutics. Unbound Medicine (Unbound Medicine, Inc, Charlottesville, VA, USA) produces a similar product offering. Both companies have apps available for many categories, including pharmacology, medical references, and medical dictionaries.

DynaMed (Figure 2) is a full medical reference app from EBSCO Publishing (EBSCO Publishing Inc, Ipswich, MA, USA). DynaMed houses a large repository of disease, syndrome, and drug information. It differs from other resources such as UpToDate (UpToDate, Inc, Waltham, MA, USA) in that it optimizes its content for use and display on a mobile platform. Once the app is fully downloaded, an Internet connection is no longer needed to access this program. 5-Minute Clinical Consult (Lippincott Williams & Wilkins, Philadelphia, PA, USA) is another medical reference app organized similarly to DynaMed. UpToDate offers a mobile Web version of their ubiquitous Web platform that is similar in organization but does not offer a smartphone app.

A popular infectious diseases resource is Johns Hopkins' Antibiotic Guide (Unbound Medicine, Inc), providing detailed information regarding antibiotics and pathogens. Another well-known antibiotic resource is the Sanford Guide to Antimicrobial Therapy (Antimicrobial Therapy, Inc, Sperryville, VA, USA).

Mobile differential diagnosis programs can help ensure that common diagnoses are not overlooked or discovered too late. One such well-known program is Diagnosaurus (The McGraw-Hill Companies, Inc; New York, NY, USA).

Other apps, such as medical calculators, are very prevalent in app stores and can help quickly calculate risk scores or other common calculations, such as water deficit in hypernatremia. Skyscape offers a free medical calculator called Archimedes. Lastly, medical dictionaries such as Taber's, Stedman's, and Dorland's are invaluable resources to have readily available in one's pocket.

An app for primary prevention is offered for free by the US Department of Health and Human Services, Agency for Healthcare Research and Quality (AHRQ). The AHRQ Electronic Preventive Services Selector is an app designed to assist primary care physicians in identifying screening, counseling, and preventive measures based on their patient's age, sex, and other risk factors.

The aforementioned programs are just a few of the large number of evolving resources on the mobile phone. Companies such as Medscape (Medscape, LLC, New York, NY, USA) offer a mobile resource for medical and drug information. They also provide medical news and case studies for continuing education. QxMD (QxMD Software Inc, Vancouver, BC, Canada) specializes in mobile medical programs such as their popular ECG Guide. Information from the *International Classification of Diseases*, 9th revision (ICD-9) can be quickly obtained from QxMD's ICD-9 app. DynaMed's app also contains ICD-9 information. QuantiaMD (Quantia Communications, Inc, Waltham, MA, USA) has a mobile app that specializes in continuing medical education by providing well-scripted interactive case studies that can be shared with colleagues (Figure 3). Continuing medical education credits can also be earned. MedPage Today (MedPage Today, LLC, Little Falls, NJ, USA) allows physicians to stay on top of the latest medical news, organize news by interest, and earn continuing medical education credits.

Doximity (Doximity Inc, San Mateo, CA, USA) has been likened to a Facebook for doctors. It allows physicians, once registered, to network and even communicate patient-related information in a Health Insurance Portability and Accountability Act-compliant text messaging environment (Figure 4). A credential check of a potential user's medical license (which is already in their database) is required to sign up.

While many of these apps have been available for years and are very popular, there are still no data to both support their use and help us understand how best to use them. We believe that studies surveying doctors on the perceived impact that specific apps create, as well as examining patient care outcomes, can help us understand how powerful these apps can be. The use of these apps by students while on clinical rotations can also support education at the bedside. This could translate into an improved quality of education that could be a focus of examination.

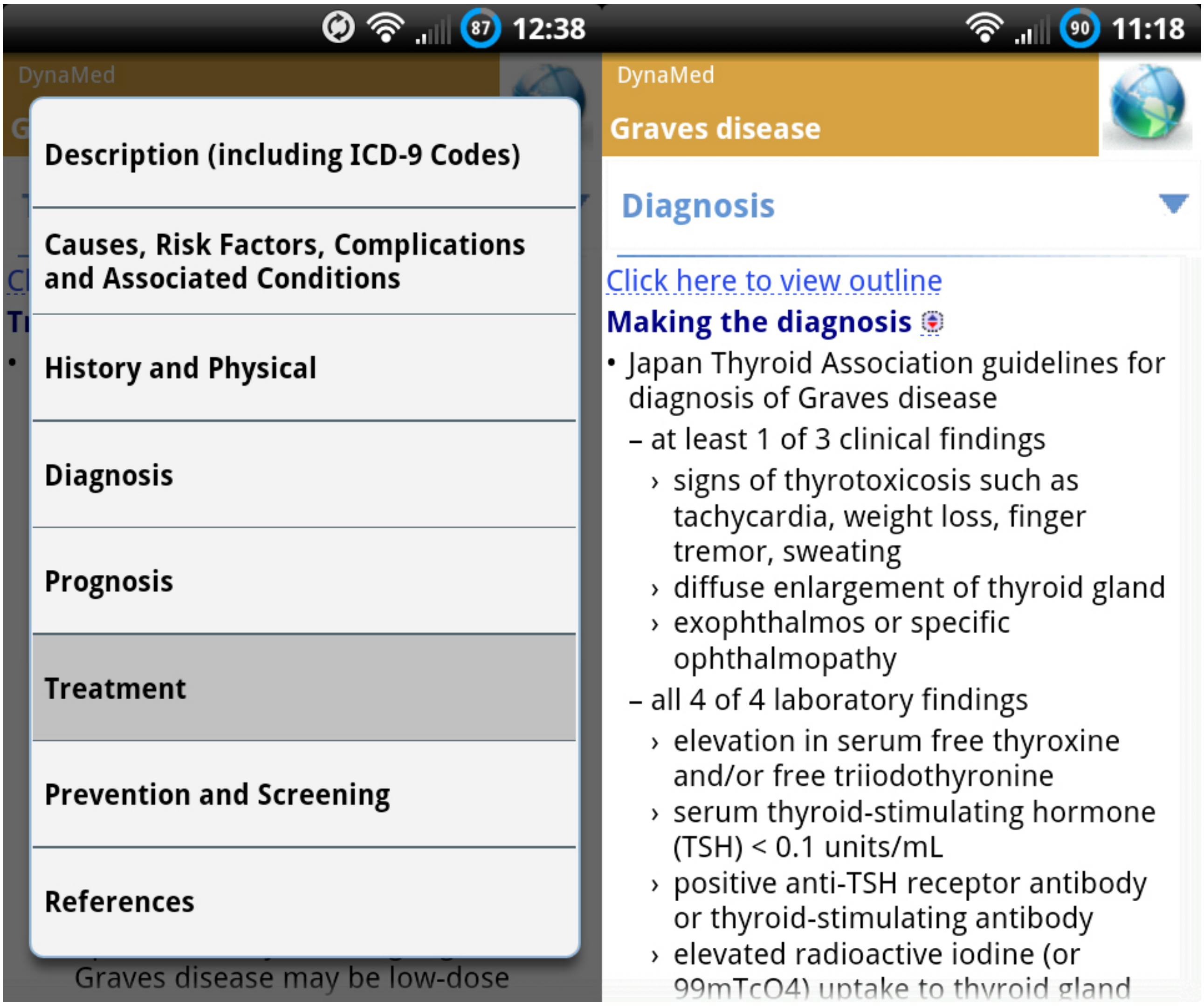
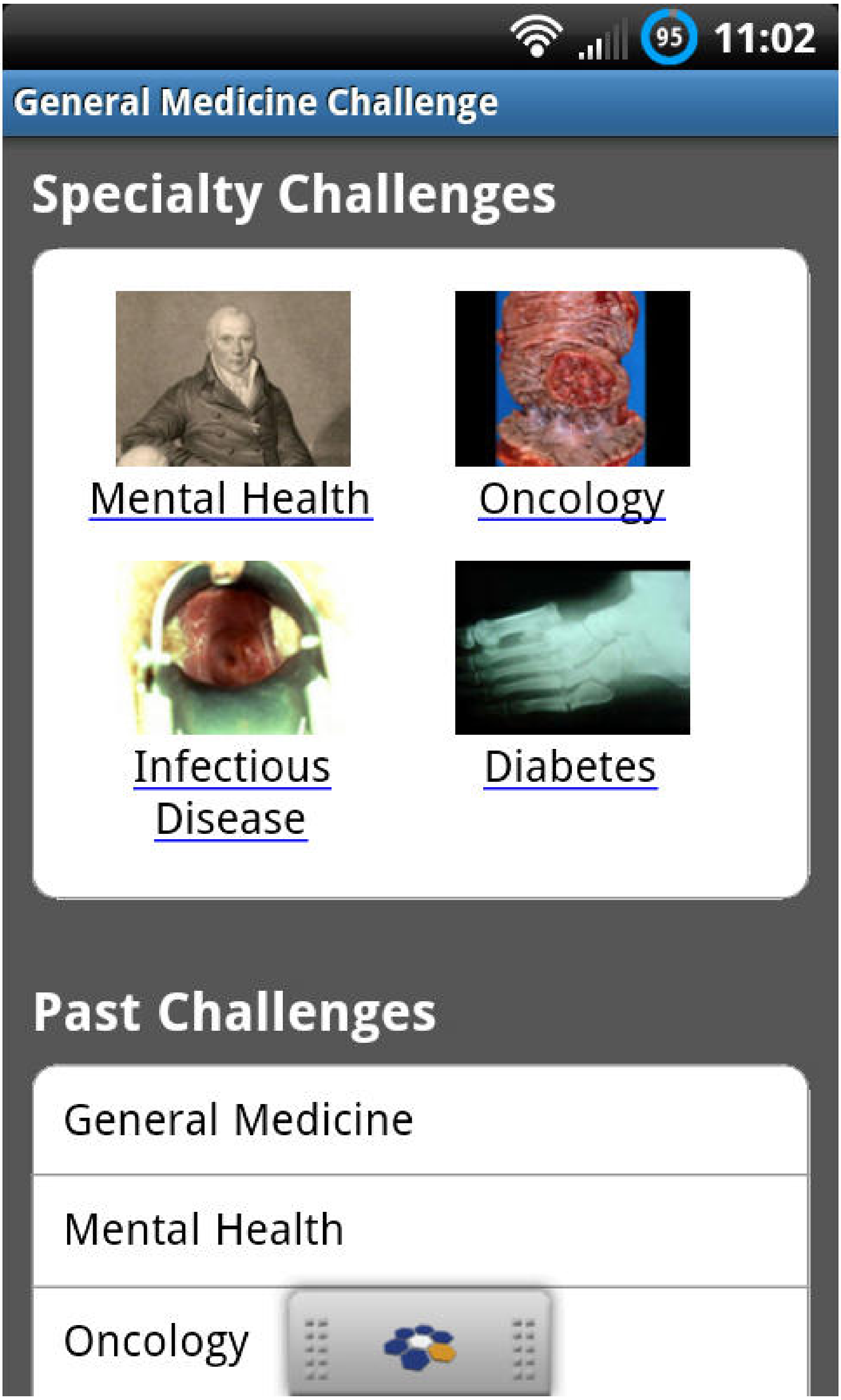


Figure 2. DynaMed’s medical reference program with organization of topics shown at left.

Making the diagnosis

- Japan Thyroid Association guidelines for diagnosis of Graves disease
  - at least 1 of 3 clinical findings
    - signs of thyrotoxicosis such as tachycardia, weight loss, finger tremor, sweating
    - diffuse enlargement of thyroid gland
    - exophthalmos or specific ophthalmopathy
  - all 4 of 4 laboratory findings
    - elevation in serum free thyroxine and/or free triiodothyronine
    - serum thyroid-stimulating hormone (TSH) < 0.1 units/mL
    - positive anti-TSH receptor antibody or thyroid-stimulating antibody
    - elevated radioactive iodine (or 99mTcO4) uptake to thyroid gland

Figure 3. QuantiaMD allows for interactive case challenges within multiple specialties, features guest lecturers, and allows physicians to earn continuing medical education points, all from the smartphone.



General Medicine Challenge

Specialty Challenges

Mental Health

Oncology

Infectious Disease

Diabetes

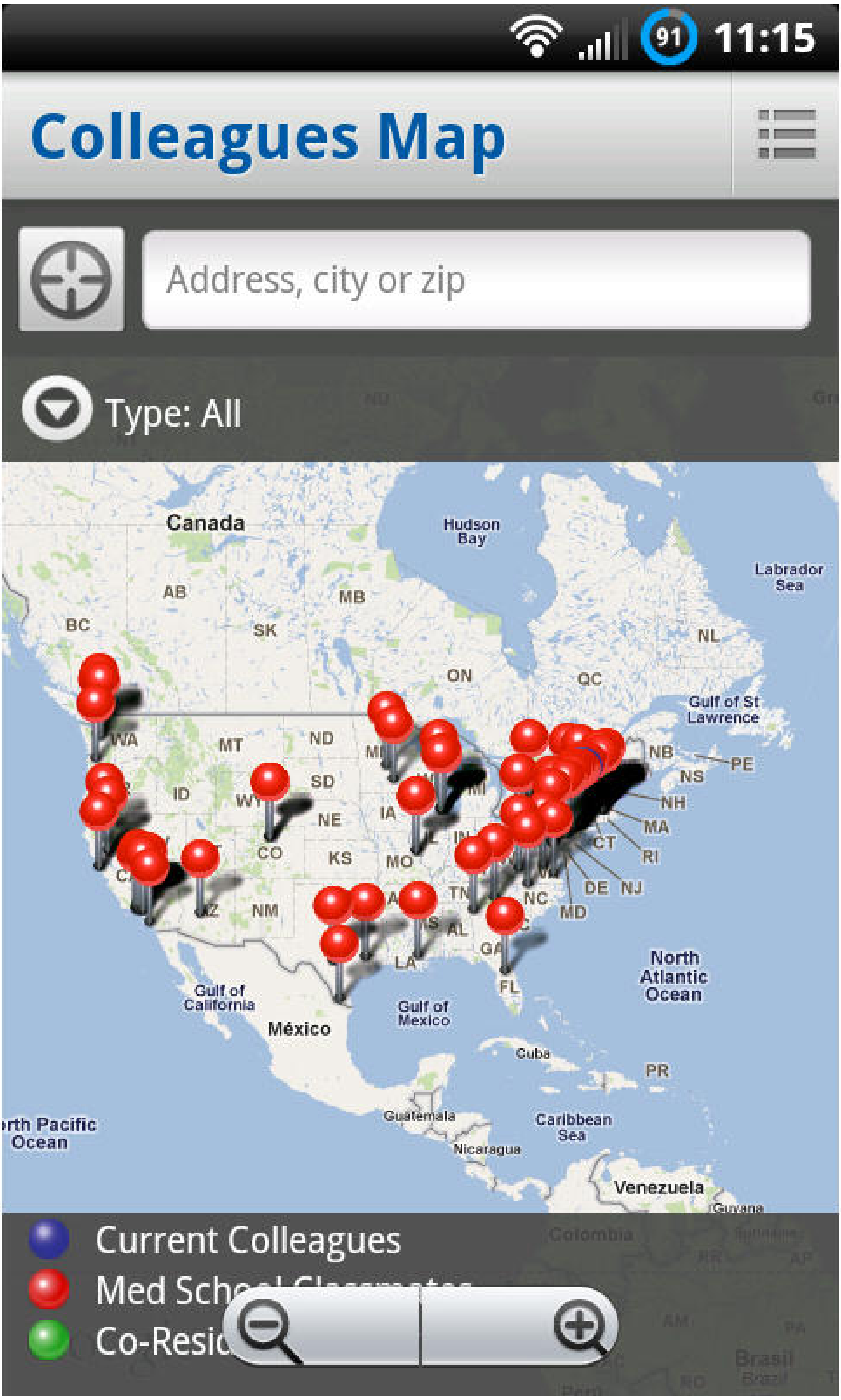
Past Challenges

General Medicine

Mental Health

Oncology

Figure 4. Doximity connects colleagues from around the United States and allows for secure communication.



Drawbacks and Obstacles

Although there are numerous benefits to integrating smartphones into the practice of medicine and one’s personal life, there are noteworthy limitations. We again emphasize the ramifications mentioned above of patients self-diagnosing using apps that are not regulated. Moreover, the major technological improvements of both hardware and software are still relatively new and, thus, sometimes unreliable. Furthermore, older physicians and others less inclined to use or intimidated by new technologies may be at a disadvantage if the use of smartphones becomes more requisite within medicine. Similarly, elderly patients may find it difficult to use and interpret the information provided to them by their smartphone, possibly putting them at greater risk than those who are more technologically savvy. And finally, as we become more dependent on technology, we become more dependent on it working flawlessly, with catastrophic implications when it fails.

Doctors and patients are not able to take full advantage of smartphone technology in areas such as teleconferencing, sending pictures, and emailing, due to health care system reimbursement processes in the United States. These systems usually reimburse only the time spent with patients face-to-face. As the smartphone integrates its way even more permanently into our medical practices, a greater question arises: will this mobile technology improve communication between doctors and patients or detract from it by limiting the personalized interactions that occur best at the bedside or in the office?

Limitations of This Review

The major limitation of the review stems from the overall paucity of high-quality studies such as multicentered or controlled trials using the smartphone in medicine. While we did find some studies of patient monitoring and communication, even these categories leave many questions to be answered, and future studies are either planned or underway.

Additionally, we again note that this review did not include papers that demonstrated novel uses of the smartphone in the field of surgery and its surgical subspecialties. As internal medicine physicians, we felt that our analysis on this subject may be inaccurate and thus chose not to include this.

Another limitation of this study is the rapid and evolving nature of this technology. We intended to make this review as up-to-date as possible, including the addition of new reports just prior to publication of this paper; however, this topic is evolving as rapidly as advancements in the industry are made, outpacing our ability to provide the most current study possible.

Conclusion

The amount of research in the use of the smartphone in medicine is rapidly growing, but there are very few good-quality studies to answer many questions about its use and the impact it may have. Apps for pharmacology, medical references, and a myriad of other categories are providing physicians with quick and practical medical information that will aid in education and patient care. Communication within hospitals and between patients is improving. Additionally, developing countries now have the potential to access better diagnostic tools in resource-poor regions. However, many obstacles still stand in the way of this progress. The question regarding whether smartphones in medicine will lead to a healthier population with better patient–doctor relationships remains to be answered. Nevertheless, the smartphone has a very bright future in the world of medicine, while doctors, engineers, and others alike continue to contribute more ingenuity to this dynamic field. It is our hope that by informing the medical community of the numerous ways in which the smartphone can be used to benefit health care providers, patients, and their families, the smartphone may one day be recognized as a diagnostic and therapeutic tool that is as irreplaceable as the stethoscope has been in the practice of medicine.

Acknowledgments

We would like to thank Lauren Maggio, MS (LIS), MA, Director of Research and Instruction at Stanford University’s Lane Library, for her invaluable support with this review. We would also like to thank Kelley Skeff, MD, for his advice. We finally thank the editor, reviewers, and our colleagues for their useful comments, which greatly helped to improve this paper.

Conflicts of Interest

None declared.

References

1. Lippi G, Plebani M. Laboratory applications for smartphones: risk or opportunity? Clin Biochem 2011 Mar;44(4):273-274. [doi: 10.1016/j.clinbiochem.2010.12.016] [Medline: 21223958]
2. Kiser K. 25 ways to use your smartphone. Physicians share their favorite uses and apps. Minn Med 2011 Apr;94(4):22-29. [Medline: 21560878]
3. Franko OI, Tirrell TF. Smartphone App Use Among Medical Providers in ACGME Training Programs. J Med Syst 2012 Oct;36(5):3135-3139. [doi: 10.1007/s10916-011-9798-7] [Medline: 22052129]
4. Sposaro F, Danielson J, Tyson G. iWander: An Android application for dementia patients. Conf Proc IEEE Eng Med Biol Soc 2010;2010:3875-3878. [doi: 10.1109/IEMBS.2010.5627669] [Medline: 21097072]
5. Wu HH, Lemaire ED, Baddour N. Change-of-state determination to recognize mobility activities using a BlackBerry smartphone. Conf Proc IEEE Eng Med Biol Soc 2011;2011:5252-5255. [doi: 10.1109/IEMBS.2011.6091299] [Medline: 22255522]
6. Worringham C, Rojek A, Stewart I. Development and feasibility of a smartphone, ECG and GPS based system for remotely monitoring exercise in cardiac rehabilitation. PLoS One 2011;6(2):e14669 [FREE Full text] [doi: 10.1371/journal.pone.0014669] [Medline: 21347403]

7. Edgar S, Swyka T, Fulk G, Sazonov ES. Wearable shoe-based device for rehabilitation of stroke patients. Conf Proc IEEE Eng Med Biol Soc 2010;2010:3772-3775. [doi: 10.1109/IEMBS.2010.5627577] [Medline: 21097053]
8. Yamada M, Aoyama T, Mori S, Nishiguchi S, Okamoto K, Ito T, et al. Objective assessment of abnormal gait in patients with rheumatoid arthritis using a smartphone. Rheumatol Int 2011 Dec 23. [doi: 10.1007/s00296-011-2283-2] [Medline: 22193221]
9. Nishiguchi S, Yamada M, Nagai K, Mori S, Kajiwara Y, Sonoda T, et al. Reliability and validity of gait analysis by android-based smartphone. Telemed J E Health 2012 May;18(4):292-296. [doi: 10.1089/tmj.2011.0132] [Medline: 22400972]
10. Mellone S, Tacconi C, Chiari L. Validity of a Smartphone-based instrumented Timed Up and Go. Gait Posture 2012 May;36(1):163-165. [doi: 10.1016/j.gaitpost.2012.02.006] [Medline: 22421189]
11. Lee BC, Kim J, Chen S, Sienko KH. Cell phone based balance trainer. J Neuroeng Rehabil 2012;9:10 [FREE Full text] [doi: 10.1186/1743-0003-9-10] [Medline: 22316167]
12. Bsoul M, Minn H, Tamil L. Apnea MedAssist: real-time sleep apnea monitor using single-lead ECG. IEEE Trans Inf Technol Biomed 2011 May;15(3):416-427. [doi: 10.1109/TITB.2010.2087386] [Medline: 20952340]
13. Kirwan M, Duncan MJ, Vandelanotte C, Mummery WK. Using smartphone technology to monitor physical activity in the 10,000 Steps program: a matched case-control trial. J Med Internet Res 2012;14(2):e55 [FREE Full text] [doi: 10.2196/jmir.1950] [Medline: 22522112]
14. Boulos MN, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed Eng Online 2011;10:24 [FREE Full text] [doi: 10.1186/1475-925X-10-24] [Medline: 21466669]
15. Charpentier G, Benhamou PY, Dardari D, Clergeot A, Franc S, Schaepelynck-Belicar P, TeleDiab Study Group. The Diabeo software enabling individualized insulin dose adjustments combined with telemedicine support improves HbA1c in poorly controlled type 1 diabetic patients: a 6-month, randomized, open-label, parallel-group, multicenter trial (TeleDiab 1 Study). Diabetes Care 2011 Mar;34(3):533-539. [doi: 10.2337/dc10-1259] [Medline: 21266648]
16. Harvey P, Woodward B, Datta S, Mulvaney D. Data acquisition in a wireless diabetic and cardiac monitoring system. Conf Proc IEEE Eng Med Biol Soc 2011;2011:3154-3157. [doi: 10.1109/IEMBS.2011.6090860] [Medline: 22255009]
17. Kostikis N, Hristu-Varsakelis D, Arnaoutoglou M, Kotsavasiloglou C, Baloyiannis S. Towards remote evaluation of movement disorders via smartphones. Conf Proc IEEE Eng Med Biol Soc 2011;2011:5240-5243. [doi: 10.1109/IEMBS.2011.6091296] [Medline: 22255519]
18. Puiatti A, Mudda S, Giordano S, Mayora O. Smartphone-centred wearable sensors network for monitoring patients with bipolar disorder. Conf Proc IEEE Eng Med Biol Soc 2011;2011:3644-3647. [doi: 10.1109/IEMBS.2011.6090613] [Medline: 22255129]
19. Palmerini L, Mellone S, Rocchi L, Chiari L. Dimensionality reduction for the quantitative evaluation of a smartphone-based Timed Up and Go test. Conf Proc IEEE Eng Med Biol Soc 2011;2011:7179-7182. [doi: 10.1109/IEMBS.2011.6091814] [Medline: 22255994]
20. Rigoberto MM, Toshiyo T, Masaki S. Smart phone as a tool for measuring anticipatory postural adjustments in healthy subjects, a step toward more personalized healthcare. Conf Proc IEEE Eng Med Biol Soc 2010;2010:82-85. [doi: 10.1109/IEMBS.2010.5626162] [Medline: 21095641]
21. Meankaew P, Kaewkungwal J, Khamsiriwatchara A, Khunthong P, Singhasivanon P, Satimai W. Application of mobile-technology for disease and treatment monitoring of malaria in the "Better Border Healthcare Programme". Malar J 2010;9:237 [FREE Full text] [doi: 10.1186/1475-2875-9-237] [Medline: 20723223]
22. Rajput ZA, Mbugua S, Amadi D, Chepngeno V, Saleem JJ, Anokwa Y, et al. Evaluation of an Android-based mHealth system for population surveillance in developing countries. J Am Med Inform Assoc 2012 Jul 1;19(4):655-659. [doi: 10.1136/amiajnl-2011-000476] [Medline: 22366295]
23. Gregoski MJ, Mueller M, Vertegel A, Shaporev A, Jackson BB, Frenzel RM, et al. Development and validation of a smartphone heart rate acquisition application for health promotion and wellness telehealth applications. Int J Telemed Appl 2012;2012:696324 [FREE Full text] [doi: 10.1155/2012/696324] [Medline: 22272197]
24. Hii PC, Chung WY. A comprehensive ubiquitous healthcare solution on an Android™ mobile device. Sensors (Basel) 2011;11(7):6799-6815. [doi: 10.3390/s110706799] [Medline: 22163986]
25. Oresko JJ, Duschl H, Cheng AC. A wearable smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing. IEEE Trans Inf Technol Biomed 2010 May;14(3):734-740. [doi: 10.1109/TITB.2010.2047865] [Medline: 20388600]
26. Sicari R, Galderisi M, Voigt JU, Habib G, Zamorano JL, Lancellotti P, et al. The use of pocket-size imaging devices: a position statement of the European Association of Echocardiography. Eur J Echocardiogr 2011 Feb;12(2):85-87 [FREE Full text] [doi: 10.1093/ejechocard/jeq184] [Medline: 21216764]
27. Huang CC, Lee PY, Chen PY, Liu TY. Design and implementation of a smartphone-based portable ultrasound pulsed-wave Doppler device for blood flow measurement. IEEE Trans Ultrason Ferroelectr Freq Control 2012 Jan;59(1):182-188. [doi: 10.1109/TUFFC.2012.2171] [Medline: 22293750]
28. Anonymous. How to put your smartphone "on call." Applications that run on your cell phone put health and wellness aids just a touch away. Harv Womens Health Watch 2010 Dec;18(4):2-4. [Medline: 21268794]

29. Ly K. MHealth: better health through your smartphone. Community Pract 2011 Feb;84(2):16-17. [Medline: 21388038]
30. Gan KO, Allman-Farinelli M. A scientific audit of smartphone applications for the management of obesity. Aust N Z J Public Health 2011 Jun;35(3):293-294. [doi: 10.1111/j.1753-6405.2011.00707.x] [Medline: 21627732]
31. Cohn AM, Hunter-Reel D, Hagman BT, Mitchell J. Promoting behavior change from alcohol use through mobile technology: the future of ecological momentary assessment. Alcohol Clin Exp Res 2011 Dec;35(12):2209-2215. [doi: 10.1111/j.1530-0277.2011.01571.x] [Medline: 21689119]
32. Abroms LC, Padmanabhan N, Thaweethai L, Phillips T. iPhone apps for smoking cessation: a content analysis. Am J Prev Med 2011 Mar;40(3):279-285. [doi: 10.1016/j.amepre.2010.10.032] [Medline: 21335258]
33. Rabin C, Bock B. Desired features of smartphone applications promoting physical activity. Telemed J E Health 2011 Dec;17(10):801-803. [doi: 10.1089/tmj.2011.0055] [Medline: 22010977]
34. Anonymous. Smartphone app speeds registration. ED Manag 2011 Mar;23(3):28-29. [Medline: 21449510]
35. Takao H, Murayama Y, Ishibashi T, Karagiozov KL, Abe T. A new support system using a mobile device (smartphone) for diagnostic image display and treatment of stroke. Stroke 2012 Jan;43(1):236-239. [doi: 10.1161/STROKEAHA.111.627943] [Medline: 21998052]
36. León SA, Fontelo P, Green L, Ackerman M, Liu F. Evidence-based medicine among internal medicine residents in a community hospital program using smart phones. BMC Med Inform Decis Mak 2007;7:5 [FREE Full text] [doi: 10.1186/1472-6947-7-5] [Medline: 17313680]
37. Wong BM, Quan S, Cheung CM, Morra D, Rossos PG, Sivjee K, et al. Frequency and clinical importance of pages sent to the wrong physician. Arch Intern Med 2009 Jun 8;169(11):1072-1073. [doi: 10.1001/archinternmed.2009.117] [Medline: 19506178]
38. Wu RC, Morra D, Quan S, Lai S, Zanjani S, Abrams H, et al. The use of smartphones for clinical communication on internal medicine wards. J Hosp Med 2010 Dec;5(9):553-559. [doi: 10.1002/jhm.775] [Medline: 20690190]
39. Wu R, Rossos P, Quan S, Reeves S, Lo V, Wong B, et al. An evaluation of the use of smartphones to communicate between clinicians: a mixed-methods study. J Med Internet Res 2011;13(3):e59 [FREE Full text] [doi: 10.2196/jmir.1655] [Medline: 21875849]
40. Lo V, Wu RC, Morra D, Lee L, Reeves S. The use of smartphones in general and internal medicine units: a boon or a bane to the promotion of interprofessional collaboration? J Interprof Care 2012 Jul;26(4):276-282. [doi: 10.3109/13561820.2012.663013] [Medline: 22482742]
41. Menon A. Confessions of a Wilderness Fellow: I Can't Live Without My Smartphone, Can You? Perm J 2011;15(1):68-69. [Medline: 21505624]
42. Choi JS, Yi B, Park JH, Choi K, Jung J, Park SW, et al. The uses of the smartphone for doctors: an empirical study from samsung medical center. Healthc Inform Res 2011 Jun;17(2):131-138 [FREE Full text] [doi: 10.4258/hir.2011.17.2.131] [Medline: 21886874]
43. Mitchell JR, Sharma P, Modi J, Simpson M, Thomas M, Hill MD, et al. A smartphone client-server teleradiology system for primary diagnosis of acute stroke. J Med Internet Res 2011;13(2):e31 [FREE Full text] [doi: 10.2196/jmir.1732] [Medline: 21550961]
44. Rahme RJ, Fishman AJ, Hunt Batjer H, Bendok BR. The future is now: smartphones to join scalpels and stethoscopes? Neurosurgery 2012 Apr;70(4):N19-N20. [doi: 10.1227/01.neu.0000413224.88090.66] [Medline: 22426057]
45. Blaya JA, Fraser HS, Holt B. E-health technologies show promise in developing countries. Health Aff (Millwood) 2010 Feb;29(2):244-251. [doi: 10.1377/hlthaff.2009.0894] [Medline: 20348068]
46. Bellina L, Missoni E. Mobile cell-phones (M-phones) in telemicroscopy: increasing connectivity of isolated laboratories. Diagn Pathol 2009;4:19 [FREE Full text] [doi: 10.1186/1746-1596-4-19] [Medline: 19545373]
47. Tice AD. Gram stains and smartphones. Clin Infect Dis 2011 Jan 15;52(2):278-279 [FREE Full text] [doi: 10.1093/cid/ciq131] [Medline: 21288858]
48. Choi BG, Mukherjee M, Dala P, Young HA, Tracy CM, Katz RJ, et al. Interpretation of remotely downloaded pocket-size cardiac ultrasound images on a web-enabled smartphone: validation against workstation evaluation. J Am Soc Echocardiogr 2011 Dec;24(12):1325-1330. [doi: 10.1016/j.echo.2011.08.007] [Medline: 21925836]
49. Crawford I, McBeth PB, Mitchelson M, Tiruta C, Ferguson J, Kirkpatrick AW. Telementorable "just-in-time" lung ultrasound on an iPhone. J Emerg Trauma Shock 2011 Oct;4(4):526-527 [FREE Full text] [doi: 10.4103/0974-2700.86654] [Medline: 22090753]
50. Tseng D, Mudanyali O, Oztoprak C, Isikman SO, Sencan I, Yaglidere O, et al. Lensfree microscopy on a cellphone. Lab Chip 2010 Jul 21;10(14):1787-1792. [doi: 10.1039/c003477k] [Medline: 20445943]
51. Breslauer DN, Maamari RN, Switz NA, Lam WA, Fletcher DA. Mobile phone based clinical microscopy for global health applications. PLoS One 2009;4(7):e6320 [FREE Full text] [doi: 10.1371/journal.pone.0006320] [Medline: 19623251]
52. Zhu H, Yaglidere O, Su TW, Tseng D, Ozcan A. Cost-effective and compact wide-field fluorescent imaging on a cell-phone. Lab Chip 2011 Jan 21;11(2):315-322. [doi: 10.1039/c0lc00358a] [Medline: 21063582]
53. Sadasivam RS, Gathibandhe V, Tanik MM, Willig JH. Development of a point-of-care HIV/AIDS medication dosing support system using the Android mobile platform. J Med Syst 2012 Jun;36(3):1583-1591. [doi: 10.1007/s10916-010-9619-4] [Medline: 21057886]

54. Oehler RL, Smith K, Toney JF. Infectious diseases resources for the iPhone. Clin Infect Dis 2010 May 1;50(9):1268-1274 [FREE Full text] [doi: 10.1086/651602] [Medline: 20233061]
55. Low D, Clark N, Soar J, Padkin A, Stoneham A, Perkins GD, et al. A randomised control trial to determine if use of the iResus© application on a smart phone improves the performance of an advanced life support provider in a simulated medical emergency. Anaesthesia 2011 Apr;66(4):255-262. [doi: 10.1111/j.1365-2044.2011.06649.x] [Medline: 21401537]
56. Chang AY, Ghose S, Littman-Quinn R, Anolik RB, Kyer A, Mazhani L, et al. Use of mobile learning by resident physicians in Botswana. Telemed J E Health 2012 Feb;18(1):11-13. [doi: 10.1089/tmj.2011.0050] [Medline: 22171597]
57. Woods CA, Dumbleton K, Jones L, Fonn D. Patient use of smartphones to communicate subjective data in clinical trials. Optom Vis Sci 2011 Feb;88(2):290-294. [doi: 10.1097/OPX.0b013e3181ff9b80] [Medline: 21099440]
58. Yamada M, Aoyama T, Okamoto K, Nagai K, Tanaka B, Takemura T. Using a Smartphone while walking: a measure of dual-tasking ability as a falls risk assessment tool. Age Ageing 2011 Jul;40(4):516-519 [FREE Full text] [doi: 10.1093/ageing/afr039] [Medline: 21593058]
59. Rothschild JM, Lee TH, Bae T, Bates DW. Clinician use of a palmtop drug reference guide. J Am Med Inform Assoc 2002 Jun;9(3):223-229 [FREE Full text] [Medline: 11971883]
60. Terry M. Medical Apps for Smartphones. Telemed J E Health 2010 Feb;16(1):17-22. [doi: 10.1089/tmj.2010.9999] [Medline: 20070172]
61. Alzheimer's Association. 2010 Alzheimer's disease facts and figures. Alzheimers Dement 2010 Mar;6(2):158-194. [doi: 10.1016/j.jalz.2010.01.009] [Medline: 20298981]
62. Chen J, Park Y, Putzer GJ. An examination of the components that increase acceptance of smartphones among healthcare professionals. Electron J Health Inform 2010;5(2).

Abbreviations

AHRQ: Agency for Healthcare Research and Quality
ECG: electrocardiograph
FDA: Food and Drug Administration
GPS: global positioning system
ICD-9: International Classification of Diseases, 9th revision

---

Edited by G Eysenbach; submitted 11.11.11; peer-reviewed by R Wu, F Wodajo; comments to author 15.01.12; revised version received 06.02.12; accepted 01.08.12; published 27.09.12.

Please cite as:
Ozdalga E, Ozdalga A, Ahuja N
The Smartphone in Medicine: A Review of Current and Potential Use Among Physicians and Students
J Med Internet Res 2012;14(5):e128
URL: http://www.jmir.org/2012/5/e128/
doi:10.2196/jmir.1994
PMID:23017375

Original Paper

Smartphone Versus Pen-and-Paper Data Collection of Infant Feeding Practices in Rural China

Shuyi Zhang[1], MSc; Qiong Wu[1], B.Sci; Michelle HMMT van Velthoven[2], MSc; Li Chen[1], MSc; Josip Car[2], MD, PhD; Igor Rudan[3], MPH, MD, PhD, DSc; Yanfeng Zhang[1], MSc; Ye Li[1], B.Sci; Robert W Scherpbier[4], MPH, MD

[1]Department of Integrated Early Childhood Development, Capital Institute of Pediatrics, Beijing, China
[2]Global eHealth Unit, Department of Primary Care and Public Health, Imperial College London, London, United Kingdom
[3]Centre for Population Health Sciences and Global Health Academy, University of Edinburgh Medical School, Edinburgh, United Kingdom
[4]Section of Health and Nutrition, Water, Environment and Sanitation, UNICEF China, Beijing, China

Corresponding Author:
Yanfeng Zhang, MSc
Department of Integrated Early Childhood Development
Capital Institute of Pediatrics
No 2 Ya Bao Road
Chaoyang District
Beijing, 100020
China
Phone: 86 10 8569 5554
Fax: 86 10 8562 2025
Email: summyzh@yahoo.com.cn

Abstract

Background: Maternal, Newborn, and Child Health (MNCH) household survey data are collected mainly with pen-and-paper. Smartphone data collection may have advantages over pen-and-paper, but little evidence exists on how they compare.

Objective: To compare smartphone data collection versus the use of pen-and-paper for infant feeding practices of the MNCH household survey. We compared the two data collection methods for differences in data quality (data recording, data entry, open-ended answers, and interrater reliability), time consumption, costs, interviewers' perceptions, and problems encountered.

Methods: We recruited mothers of infants aged 0 to 23 months in four village clinics in Zhaozhou Township, Zhao County, Hebei Province, China. We randomly assigned mothers to a smartphone or a pen-and-paper questionnaire group. A pair of interviewers simultaneously questioned mothers on infant feeding practices, each using the same method (either smartphone or pen-and-paper).

Results: We enrolled 120 mothers, and all completed the study. Data recording errors were prevented in the smartphone questionnaire. In the 120 pen-and-paper questionnaires (60 mothers), we found 192 data recording errors in 55 questionnaires. There was no significant difference in recording variation between the groups for the questionnaire pairs (*P* = .32) or variables (*P* = .45). The smartphone questionnaires were automatically uploaded and no data entry errors occurred. We found that even after double data entry of the pen-and-paper questionnaires, 65.0% (78/120) of the questionnaires did not match and needed to be checked. The mean duration of an interview was 10.22 (SD 2.17) minutes for the smartphone method and 10.83 (SD 2.94) minutes for the pen-and-paper method, which was not significantly different between the methods (*P* = .19). The mean costs per questionnaire were higher for the smartphone questionnaire (¥143, equal to US $23 at the exchange rate on April 24, 2012) than for the pen-and-paper questionnaire (¥83, equal to US $13). The smartphone method was acceptable to interviewers, and after a pilot test we encountered only minor problems (eg, the system halted for a few seconds or it shut off), which did not result in data loss.

Conclusions: This is the first study showing that smartphones can be successfully used for household data collection on infant feeding in rural China. Using smartphones for data collection, compared with pen-and-paper, eliminated data recording and entry errors, had similar interrater reliability, and took an equal amount of time per interview. While the costs for the smartphone method were higher than the pen-and-paper method in our small-scale survey, the costs for both methods would be similar for a large-scale survey. Smartphone data collection should be further evaluated for other surveys and on a larger scale to deliver maximum benefits in China and elsewhere.

(J Med Internet Res 2012;14(5):e119) doi:10.2196/jmir.2183

KEYWORDS

Data collection; health survey; questionnaires; infant feeding; smartphone

Introduction

Undernutrition in infants and young children is highly prevalent in low- and middle-income countries and results in substantial mortality and morbidity [1]. Inadequate breastfeeding and complementary feeding are the key factors causing undernutrition in infants, which ultimately affects child survival [2]. Globally, suboptimal breastfeeding is estimated to be responsible for 1.4 million child deaths and 44 million disability-adjusted life-years [1]. Therefore, improving infant feeding practices in children from 0 to 23 months of age is critical to improve nutrition, health, and development of the children.

Breastfeeding and complementary feeding practices are key indicators for child health [3]. Accurate data on feeding practices is extremely important to start appropriate health interventions that can improve children's health. The Maternal, Newborn, and Child Health (MNCH) household survey (unpublished data, 2009) is an instrument for collecting data on the coverage of key child health interventions, delivery channels, reasons for coverage failure, and health expenditures. Originally, this instrument was developed by the World Health Organization (WHO) as a paper-based questionnaire for resource-limited settings and had been used in Cambodia, Papua New Guinea, and Vietnam [4]. One of the MNCH household survey modules is on breastfeeding and nutrition aiming to collect feeding information of children aged 0 to 23 months. In 2009, WHO and UNICEF jointly developed the guidelines on indicators for assessing infant and young child feeding practices [3], setting up a series of international standard infant and young child feeding coverage indicators.

For decades, pen-and-paper-based data collection has been the standard method for household surveys. However, this has several problems, including data collection and entry errors and the high costs for storage and double entry of data [5]. In the past 20 years, electronic methods of data collection have been developed on handheld devices such as personal digital assistants (PDAs) and more recently on mobile phones. Worldwide there are now about 6 billion mobile phones, of which 4.5 billion can be found in developing countries [6]. The growth in mobile phone subscriptions is led by China and India, which now have over 30% of the world's subscribers [7]. The use of mobile devices for the delivery of health care, also known as mHealth [8,9], has increasingly gained attention over the past years. However, a limited number of studies have evaluated the use of mobile phones as a data collection tool in developing settings. Our literature searches in electronic English-language health databases (The Cochrane Central Register of Controlled Trials, PubMed, EMBASE, WHO Global Health Library regional index, PsycINFO, Web of Science, MobileActive, KIT Information Portal, and mHealth in Low-Resource Settings) found 9 studies using mobile phones for health data collection [10-17], and we found no studies in the Chinese literature (Wanfang Data and the China National Knowledge Infrastructure).

Smartphones may be more suitable than low-end mobile phones for data collection, as smartphones have larger screens and can more easily accommodate complex functions (such as wireless uploading and downloading, screen touch typing, and photo or video capturing). Rapid developments in technology and falling prices of handsets make smartphones more accessible for data collection in developing settings. Smartphones share the advantages PDAs have for data collection, for example, the ability to combine the processes of data recording and data entry [18]. Smartphone software can be programmed to skip questions and give alerts when a question is answered incorrectly, which further improves data accuracy. Smartphone data collection may provide better data quality, less time consumption, and lower costs than with pen-and-paper data collection. However, the use of smartphones has some drawbacks, including that data can become corrupted when the device is damaged, and replacement costs are relatively high when the device is lost or damaged. Moreover, most surveys were originally designed for pen-and-paper use (the reference standard), and validation of questionnaires is required. There could be response bias between paper and electronic questionnaire versions [19], as patients may respond differently to questionnaires in different formats [20].

It is unknown whether smartphones can be effectively used for household survey data collection in rural China. This study aimed to compare the use of smartphones with the use of pen-and-paper for data collection of infant feeding practices with the MNCH household survey. We evaluated differences in data quality (data recording, data entry, open-ended answers, and interrater reliability), time consumption, costs, interviewer's perception, and problems encountered.

Methods

Study Design

This study compared two methods for MNCH household survey data collection: smartphone versus pen-and-paper. We randomly assigned mothers of infants aged 0 to 23 months to a smartphone or a pen-and-paper group. A pair of interviewers simultaneously interviewed mothers on infant feeding practices. Both interviewers used the same method, either the smartphone or the pen-and-paper method. The interviewers each recorded the mothers' responses separately at the same time. One interviewer asked the questions and the other interviewer assisted in providing details. We instructed the interviewers to change the leading and assisting roles in every interview. We compared the two data collection methods for differences in data quality (data recording, data entry, open-ended answers, and interrater reliability), time consumption, costs, perception of interviewers, and problems encountered. We undertook a small 2-day pilot study in July 2011 to test this setup.

Study Setting

We carried out the surveys in four village clinics in Zhaozhou Township, Zhao County, Hebei Province, China. More detailed information on the study setting can be found in Multimedia Appendix 1.

Participants

There were two types of participants: (1) interviewers, who interviewed the mothers in pairs, and (2) mothers, who were interviewed on their infant feeding practices. We recruited 10 students from Hebei Union University School of Public Health (1 second-year and 9 third-year students) as interviewers. We visited village doctors (in China, village doctors are familiar with all births in their catchment area and monthly report all pregnant women and newborns to their township hospital) and asked them to identify (by review of their records) all eligible mothers on their list and to invite them to come to the clinic. Mothers were eligible if they had children less than 24 months old; we excluded caregivers who were not the mother of the child. If there were more than 2 children of eligible age in a family, we collected information on the youngest child. Eligible mothers were requested to ask their neighbors to visit the clinic as well. The team visited village clinics one by one until we reached our sample size of 120.

Training of Interviewers

Our interviewers had experience with the MNCH household survey; 3 of them had participated in our pilot smartphone questionnaire study in July 2011 and all 10 participated in a baseline MNCH household survey (with a sample size of 1600) in August 2011. Although the interviewers were familiar with the survey, we provided them with additional training to reinforce their skills. The supervisors (study team members) thoroughly trained the interviewers on the use of the smartphone and pen-and-paper methods for 2 days. The training course included communication skills, explanation of questionnaires, demonstration, role plays, practice interviews with mothers, and group discussion throughout the course. Interviewers were encouraged to ask questions when they experienced any problem.

Recruitment

The study took place over 2 days in September 2011; on the first day we recruited mothers in three village clinics, and on the second day we reached our sample size in the fourth village. When a mother arrived in a village clinic, the village doctor informed her about the study and referred mothers who were interested in study participation to our supervisors. The supervisors obtained oral informed consent from the mother before the study. The interviewers obtained written informed consent and then questioned the mother in one of the five separate village clinic rooms. We gave each mother a towel (worth ¥5, equal to US $0.79 at an exchange rate of 6.3 on April 24, 2012) for her participation.

Randomization and Allocation

A supervisor gave mothers an identification number (the first mother was given number 1, the second mother number 2, and so on). Prior to the study we used Excel (version 2007; Microsoft Corporation, Redmond, WA, USA) to randomly assign each number to either the smartphone or the pen-and-paper group. Each of the 10 interviewers randomly drew a lot and was assigned to 1 of the 5 interview pairs in which they worked during the study. The 5 interview work pairs took turns when interviewing the mothers.

Data Collection and Entry Process

Pen-and-Paper Method

For mothers who were randomly assigned to the pen-and-paper method, an assigned pair of interviewers asked questions that were printed on a paper questionnaire. Each of the 2 interviewers separately recorded the interviewee’s response with a pen on a copy of the questionnaire. A supervisor collected the completed paper questionnaires and checked them immediately after the interview to ensure that any missing information or errors could be corrected before the mother left. Two students, who had experience with data entry, separately entered data with EpiData 3.1 (EpiData Software, Odense, Denmark). We compared the two files and resolved discrepancies by checking the original questionnaires.

Smartphone Method

For mothers who were randomly assigned to the smartphone method, an assigned pair of interviewers questioned the mothers by following the instructions on the smartphone. We used a Chinese smartphone (C8600; Huawei, Shenzhen, China) with an Android 2.2 system. Each interviewer recorded the mother’s response by touching the smartphone screen. The smartphone program checked the questions automatically in real time. The smartphone program showed a warning when an answer was out of range. Three questions could be automatically skipped based on the response of previous questions. Most of the questions were multiple-choice questions with one-response or multiple-response answers. Some questions required a number to be filled in, and for other variables one of the answer options of multiple-response answers was open ended (which made it possible to type in Chinese characters for the smartphone questionnaires or write down characters for the pen-and-paper questionnaires). Figure 1 shows screen shots of the main four types of questions. More detailed information on the smartphone questionnaire can be found in Multimedia Appendix 1.

Figure 1. Screen shots of single-choice, multichoice, number-filling, and text-filling questions for the smartphone questionnaire.

ID.4
与儿童的关系?
(单选)
1. 母亲
2. 父亲
3. （外）祖父母
9. 其他,[]

BN.7
下面我想了解你对初乳的一些看法（初乳是分娩后3天内分泌的乳汁，外观粘稠、发黄）。
(复选(2项))左="提及"; 右="未提及"
1. 初乳健康
2. 初乳干净
3. 初乳营养丰富
4. 初乳能够帮助孩子

BN.2
孩子应在多大时开始吃粥或捣碎的固体食物?
(单行输入)多少月，88=不知道

BN.15.k
其他液体
(单选)过去24小时吃了吗?
爽歪歪

Survey

We used the Breastfeeding and Nutrition module of MNCH household survey. This is a WHO standard household survey, which we translated into Chinese. We have used it over the past 2 years in several studies in Zhao County and made minor changes to make it appropriate for the Chinese context. More detailed information on the survey can be found in Multimedia Appendix 1.

Outcomes

Our primary outcome was data quality, which we defined as data recording errors and data entry errors in the pen-and-paper questionnaires; interrater reliability; and open-ended answer differences between the smartphone and pen-and-paper methods. Secondary outcomes were the differences between the two methods in time consumption for the data recording and entry, costs, interviewers' perceptions, and problems they encountered.

Data Quality

Data Recording Errors of the Paper Questionnaires

We report on data recording errors for the paper questionnaire only, as the smartphone program automatically alerted the interviewer when a mistake occurred (and therefore there were zero data recording errors in the smartphone questionnaires). The supervisors checked and corrected missing values or errors at the end of the interview. After the fieldwork, 1 of the supervisors filled in an error summary form with the frequency of all types of errors and numbers of questionnaires with errors. We counted the errors in Excel 2007 with the COUNT function.

Interrater Reliability

We defined interrater reliability as the percentage of overlap in recording between the 2 interviewers who interviewed the same mother using the same method [21].

Data Entry Errors of Paper Questionnaires

We report data entry errors for the pen-and-paper questionnaires only, as the smartphone questionnaire responses were automatically transferred to the database after uploading (according to the information technology experts, the smartphone data were not changed during transfer into the database). For the pen-and-paper questionnaire, though 2 people entered and checked the data separately, errors could still occur.

Open-Ended Answers

We counted the number of Chinese symbols for the 24 questions that had one open-ended answer.

Time Consumption for Data Recording and Entry

For the pen-and-paper questionnaire, the interviewers recorded the starting time and ending time of the interview. For the smartphone questionnaire, the software automatically recorded the duration of the interview. In addition, we recorded the time consumption of data entry and data cleaning for the pen-and-paper group.

Costs

To assess and compare costs, we used the same cost categories for the pen-and-paper and smartphone questionnaires categories (eg, training, travel, and accommodation). For the pen-and-paper method, we estimated the costs of printing and transporting the questionnaire, stationery, and data entry. In our smartphone costs assessment, we used an estimated local market price for renting the smartphone and the software. To compare the pen-and-paper and smartphone methods, we calculated cost by individual questionnaire. In addition, we used the cost data of two large-scale household surveys (sample sizes of 1200 and 1600) to show the costs for larger-scale surveys.

Perceptions of the Interviewers

Before the fieldwork, we gave each interviewer pen-and-paper questionnaires. We asked them to fill in the questionnaire and to give it back to the supervisor after the fieldwork. The questionnaire included general information (eg, age, sex, and education), experience with using a mobile phone or a smartphone, and perceptions of using the two methods for the interviews. We asked them if they liked the smartphone and the pen-and-paper methods, and the interviewer could respond on a scale from 1 to 5 (1, very bad; 2, bad; 3, ok; 4, good; and 5, very good). After the fieldwork, our supervisors and all the interviewers had a 30-minute group discussion about the differences between the pen-and-paper survey and the smartphone survey (such as time needed for an interview, which method was easier to communicate with the interviewees, and responses of the mothers during the interview).

Problems Encountered

We instructed the interviewers to record problems, such as loss of pens and program errors, as soon as possible on standard forms during the fieldwork. The supervisors collected all the forms after the fieldwork and wrote down the problems.

Analysis

Quantitative Analysis

We carried out statistical analysis with SPSS 16.0 for Windows (IBM Corporation, Somers, NY, USA). For data recording errors, we manually counted all the errors from the error summary form, which was filled in by our supervisors in the field. For interrater reliability, we manually compared all completed pen-and-paper questionnaires and smartphone questionnaires within each pair and report the percentage variation. We calculated the difference in interrater reliability between two groups with the chi-square test and we report percentages. We did not analyze interrater reliability with kappa statistics, as we were unable to use a test–retest design for this study. For data entry errors we compared the two files using EpiData 3.1, which gave a report on the nonmatching questionnaires. We report the differences in percentages. For time consumption, we compared the average time duration of data collection in a work pair for the two groups by the independent-samples *t* test. We tested whether the time consumption was normally distributed. We report the mean difference and standard deviation, and we used a significance level of .05. For costs, we compared the total costs between the two methods. For interviewers' perceptions of the smartphone and pen-and-paper questionnaires, we calculated the median of the scale from 1 to 5.

Qualitative Analysis

For interviewers' perceptions, we analyzed the transcription of the recorded discussion and noted all identified issues. We list all the problems that were experienced by the interviewers.

Sample Size

We based our sample size on the data entry errors. We assumed that the data entry error rate would be 30% for the pen-and-paper method and 0% for the smartphone method. We used an alpha error of .05 and beta error of .20 and calculated that the sample size would be 120 with 60 mothers per method. We assumed that 10% of the mothers would decline to participate or withdraw from the study. Therefore, we planned to recruit 67 mothers per method.

Ethical Approval

We obtained ethical approval from the Ethical Committee of the Capital Institute of Pediatrics in Beijing. This is a comparison study, which does not assess an intervention, and therefore is not registered with a randomized trial registry.

Data Security

We encrypted the data stored in the smartphones when the data were uploaded. The data could be decrypted only with special software. When the interviewer completed the questionnaire, the data were wirelessly uploaded into an Excel database via the Internet server and then saved in the memory card of the smartphone as a text file (for a backup). Only the supervisors could enter the database and make the necessary changes before the data were uploaded. After the data were uploaded, no changes could be made to the database. The supervisors collected the smartphones at the end of each fieldwork day and returned the smartphones (cleared of the data that were entered during the previous day) to the interviewers in the morning. The supervisors collected the completed paper questionnaires when the interview was finished and stored the questionnaires in a safely locked box.

Results

The village doctors informed 120 mothers who visited the village clinics about the study, and all mothers agreed to participate. Figure 2 shows a flow diagram of the recruitment of the 120 mothers: 60 received the paper-and-pen version and 60 received the smartphone version. Table 1 lists the number of interviewed mothers per village. The age and sex ratio of the mother's youngest infants were similar in the two groups.

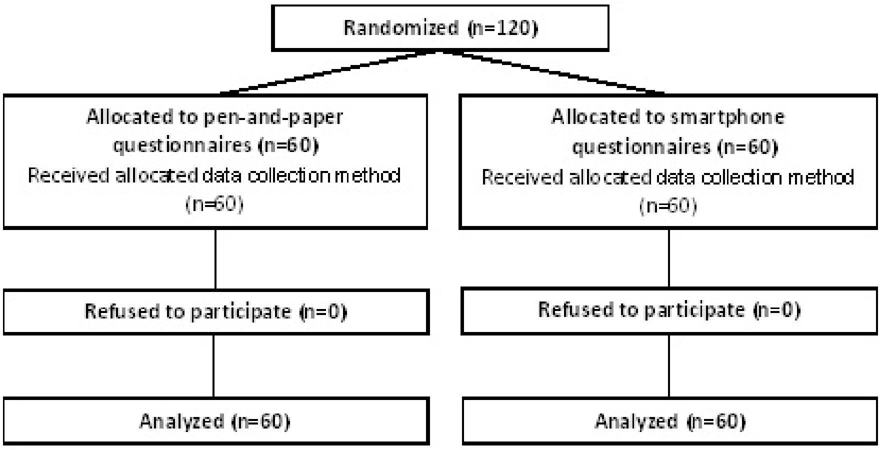
Table 1. Characteristics of interviewed mothers' child and number of mothers per village.

Characteristic	Total	Smartphone group	Pen-and-paper group	*P* value
Children				
Age (days), mean (SD)	353.88 (174.65)	323.95 (149.86)	383.82 (192.99)	.06[a]
Sex (male/female)	61/59	32/28	29/31	.59[b]
Number of mothers				
Total	120	60	60	
Village 1	25	16	9	
Village 2	50	23	27	
Village 3	24	11	13	
Village 4	21	10	11	

[a] 2-tailed *t* test.

[b] Pearson chi-square test.

Figure 2. Flow of study participants.



Data Quality

Recording Errors of Pen-and-Paper Questionnaires

Table 2 demonstrates that in 120 copies of the pen-and-paper questionnaires, 55 questionnaires contained errors. The most frequent error was a missing confirmation of the default option, which was found 156 times in 49 questionnaires.

Table 2. Recording errors in pen-and-paper questionnaires.

Type of error		Immediately after interview		Before data entry		Total	
		No. of questionnaires with error(s)	No. of errors	No. of questionnaires with error(s)	No. of errors	No. of questionnaires with error(s)	No. of errors
1	Missing confirmation of default option	49	156	0	0	49	156
2	Missing survey date	1	1	0	0	1	1
3	Wrong response for checking question	2	2	0	0	2	2
4	Two options circled	2	2	0	0	2	2
5	Wrong option chosen	3	4	0	0	3	4
6	Missed question	2	2	1	4	3	6
7	More than one option circled in single-choice question	1	1	0	0	1	1
8	Wrong ID number for interviewer	0	0	7	8	7	8
9	Wrong date	0	0	1	1	1	1
10	Wrong setting for database	0	0	11	11	11	11
Total		51[a]	168	20[a]	24	55[a]	192

[a] Total numbers of questionnaires with error(s). A questionnaire could have more than one error, but was counted as one copy. Therefore, the total number of questionnaires does not equal the total number of all types of errors.

Interrater Reliability Within Interviewer Pairs

We assessed the interrater reliability for the one-response and multiple-response answer variables and the number variables; this can be found in Table 3. The supervisors checked the two records for each mother and judged whether they were the same. For 35 of the 120 questionnaire pairs (20 pairs in the pen-and-paper group and 15 pairs in the smartphone group), there was no recording variation in the database, which was not significantly different between the groups (*P* = .32). There were 186 variables in the smartphone and 184 in the pen-and-paper questionnaire. In the smartphone questionnaire, 134 of 186 variables (72.0%) did not have any recording variation. In the pen-and-paper questionnaire, 126 of 184 variables (68.5%) did not have any recording variation. This was not significantly different between the groups (*P* = .45).

Table 3. Interrater reliability within interviewer pairs.

	Smartphone number/ total number	Pen-and-paper number/ total number	*P* value[a]
Questionnaire pairs with no recording variations	15/60	20/60	.32
Variables with no recording variations	134/186 (72.0%)	126/184 (68.5%)	.45

[a] Pearson chi-square test.

Open-Ended Answers

We manually counted the characters recorded in variables for the open-ended questions. There were 48 characters in the smartphone questionnaires and 76 in the pen-and-paper questionnaires.

Data Entry Errors of Pen-and-Paper Questionnaires

EpiData 3.1 showed that 65.0% (78 of 120 questionnaires) of the pen-and-paper records did not match and needed to be checked.

Time Consumption

Figure 3 and Multimedia Appendix 2 (xTable MA3) show that the mean duration of an interview was 10.60 (SD 2.49) minutes, with 10.22 (SD 2.17) minutes for the smartphone method and 10.83 (SD 2.94) minutes for the pen-and-paper method. The mean interview duration was not significantly different between the methods (*P* = .19).

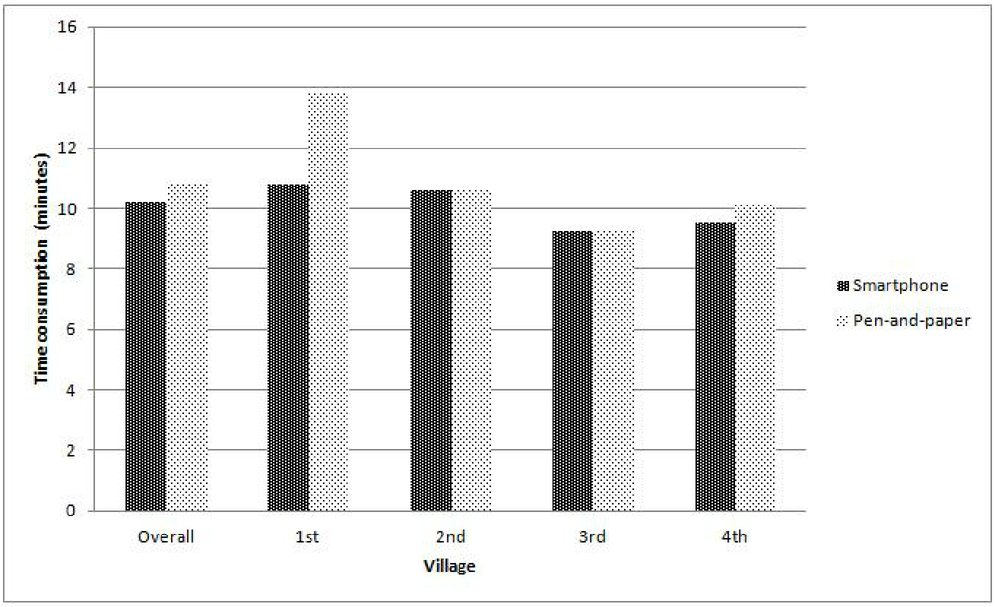
In the first village, the interviewers spent significantly more time for the pen-and-paper method (mean 13.78 minutes, SD 3.70) than for the smartphone method (mean 10.78 minutes, SD 2.37) for an interview (*P* = .02). We found no significant difference between the two methods in the following three villages.

For the pen-and-paper method, database completion took 16 hours (including data entry, checking, and data cleaning). For the smartphone method, database completion took half an hour. This can be found in Table 4.

Table 4. Time consumption for the survey process for the two methods.

Phase	Smartphone method	Pen-and-paper method
Preparation	7 days for printing	7 days for programming and installment for a team of 3 information technology engineers
Training	2 days	2 days
Fieldwork	2 days	2 days
Data pooling	3 hours for 2 persons to do double data entry, and 5 hours for 2 persons to check and clear the data	0.5 hour for 1 person

Figure 3. Time consumption for the conducting the interviews overall and per village for the smartphone and pen-and-paper methods.



Costs

Table 5 displays the total costs and costs per questionnaire, which are divided into two parts, logistics and questionnaire work. We estimated the costs based on local market prices. We spent ¥17,200 (US $2730) for the smartphone group (survey 1) and ¥9970 (US $1582) for the pen-and-paper group (survey 2). This included all items for preparation, training, fieldwork and data collection, and logistics. We used the following estimates: leasing the software cost ¥600 (US $95) per week (the software could not be leased per day); renting the sever cost ¥3000 (US $476) per week; labor cost ¥100 (US $16) per hour and ¥300 (US $48) per day; and postage was ¥336 (US $53) per 40 kg parcel (delivered by the National Post Office). Table 5 shows that the costs of both methods were similar for larger-scale household surveys. Multimedia Appendix 2 (xTable MA4) shows the general information on the basis of which we estimated the costs for our two methods (survey 1 and 2) and for two larger-scale surveys (one pen-and-paper and one smartphone).

Table 5. Costs (in US $1000) of four surveys in Zhao County, China.

Survey number	Type	No. of interviewees	Costs		Item costs					
					Logistics		Hardware and mailing		Data pooling	
			Total	Per questionnaire	Total	Percentage of all costs (%)	Total	Percentage of all costs (%)	Total	Percentage of all costs (%)
1	Smartphone	60	2.70	22.50	1.27	47.09	1.41	52.33	0.02	0.58
2	Pen-and-paper	60	1.57	13.05	1.27	81.24	0.04	2.71	0.25	16.05
3	Smartphone	1600	41.57	25.98	34.82	83.76	6.59	15.86	0.16	0.38
4	Pen-and-paper	1200	28.52	23.77	25.56	89.60	0.71	2.47	2.26	7.93

Perceptions of the Interviewers

We analyzed 9 of the 10 questionnaires given to the interviewers, as we received 1 blank questionnaire. Of these 9 interviewers, 8 had experience with doing pen-and-paper surveys before this study, and 4 of the 9 interviewers had experience with using a smartphone. On the question about whether the interviewers liked the survey method, their median score was 4 for the smartphone and 3 for the pen-and-paper method. A total of 4 of the interviewers marked the smartphone 1 point higher than the pen-and-paper, and 5 gave the same mark to the smartphone and the pen-and-paper method.

All the interviewers actively participated in the group discussions. For the pen-and-paper questionnaires, the main issues were being afraid of skipping questions, having to write down a lot on the paper that could waste time, perceiving a high risk of missing options that could not be easily found in the fields, having to carry the heavy questionnaires, and transportation difficulties. The only issue the interviewers identified using the smartphone was that if the program was unstable they could not go forward in the questionnaire. The interviewers experienced the following benefits using the smartphone: the automatic skipping function and error alerts took away the interviewers’ fear of making mistakes, one question per screen put more focus on the communication with the mother during the interview, the smartphone was portable and easy to handle, and data upload was quick. The interviewers mentioned that the mothers said that using the smartphone method was more modern and quicker than using the pen-and-paper method.

Problems Encountered

In the pen-and-paper group, only one abnormal event was recorded: an interviewer lost her pen and got a new pen from a supervisor immediately. In the smartphone group, five cases of abnormal conditions were recorded. All of them were about the system’s stability, such as that the system halted for a few seconds or that it shut off. We did not find that these recorded abnormalities caused data to be lost.

Discussion

Principal Results

Our study showed that using the smartphone to collect data on breastfeeding and nutrition, when compared with the pen-and-paper method, eliminated data recording and entry errors, had similar interrater reliability, and took an equal amount of time per interview. Fewer Chinese characters were entered in the smartphone questionnaire, which may indicate that the smartphone was less suitable for open-ended answers. While the costs for the smartphone method were higher than for the pen-and-paper method in our small-scale survey, the costs for both methods would be similar for a larger-scale survey. The smartphone method was acceptable to interviewers, and after pilot testing we encountered only minor problems.

Limitations

Our study had some limitations. First, it was hard to validate each interview when comparing the two methods. The test and retest methods are not feasible for the Breastfeeding and Nutrition module, which is based on 24-hour recall information. Therefore, we allocated 2 interviewers using the same method to interview the same mother, and then compared the interrater reliability of the two methods. However, different responses may have been given to the interviewers depending on the survey method (for example, the interviewers may have taken pen-and-paper questionnaires more seriously). We were unable to analyze this. Second, our sample size was relatively small, which may explain the relatively high standard deviations in the time consumption. Also, this may limit the generalizability of our results to other settings. However, at the time of development of the study, we did not have a better indicator than the data entry error assumptions. Third, we do not know how the mothers perceived the interview, as we did not collect data on the perceptions of mothers. However, the interviewers reported that the mothers found the smartphones modern and quicker. Fourth, all our interviewers were medical students and had experience with questionnaire interviewing. All were young and could easily learn how to use the smartphone; this may be different for data collectors in nonstudy situations. We expect that this will be a minor problem, as smartphone ownership is rapidly increasing. Strengths of our study include our pilot test of the smartphone software, which ensured that the system worked, thorough training of our interviewers, and our experience with undertaking the MNCH survey in Zhao County.

Comparison With Prior Work

While mobile devices have been used for health data collection over the past 20 years, we are unaware of any study using a mobile device or smartphone for data collection in China. In other countries, prior to the rapid development of smartphones, many studies used handheld computers such as PDAs for data collection [22-25]. A review showed that paper- and computer-collected, patient-reported outcomes are equivalent for both methods [26]. Previous studies found that data record and entry errors did not occur when an electronic device was used for data collection [27,28], which we confirmed in our study. A study demonstrated that the use of PDAs reduced data entry and transfer time by 23% [28]. Another study showed that it took less time to complete a questionnaire in the PDA group than in the pen-and-paper group [29]. Our study found that the smartphone can effectively use real-time upload and backup, and can prevent data loss problems, which was similar to the findings of a South African study using mobile phones as a data collection tool [5]. A study of an Android-based mHealth system found that users of the system felt it was easy to use and that it facilitated their work [30], which our interviewers also experienced.

In China, an estimated 70% of the population own a mobile phone, and they are widely used in both urban and rural areas [7]. In Zhao County, approximately 75% of the population and nearly all households have at least one mobile phone. Rapid developments in information and communication technology make Internet access increasingly affordable to people in China. The current third-generation network covers most of the counties (except for some remote areas) and facilitates data upload. The smartphone we used costs ¥1500 (US $230). As prices of handsets are falling rapidly, future studies may be able to use even cheaper smartphones with more functions. Additional functions such as a global positioning system, storing pictures and videos, taking photos, and recording sound could facilitate data collection.

Conclusion

This study is the first to show that smartphones can be successfully used for data collection in a rural setting in China. Accurate data are essential for the success of any public health survey and to inform appropriate interventions. Smartphone data collection can improve data quality by preventing data recording and entry errors. Also, increased efficiency at the data recording and entry stage is an important benefit of the smartphone method compared with the pen-and-paper method. This could lead to substantial time and cost savings. Smartphone data collection should be further evaluated for other surveys and on a larger scale to deliver maximum benefits to China and elsewhere.

Acknowledgments

We are indebted to all the interviewed mothers for participating in our study and to the medical students for their work. We thank the Ministry of Health of China, which provided funding support for the study. We are grateful to our colleagues at the Zhao Maternal and Child Health Hospital for coordination and support of the fieldwork. Also we would like to thank the software company for providing their expertise with developing the software. We are very grateful to Dr Laura Gunn from the Global eHealth Unit, Department of Primary Care and Public Health, Imperial College London, United Kingdom, for providing us with statistical support. Also we would like to thank the software company (BeijingWeiShenQingHong Technology CO., Ltd) for providing their expertise with developing the software.

Conflicts of Interest

None declared.

Multimedia Appendix 1

More detailed information on the Methods section of the paper.

[PDF File (Adobe PDF File), 2MB - jmir_v14i5e119_app1.pdf]

Multimedia Appendix 2

More detailed information on the Results section of the paper.

[PDF File (Adobe PDF File), 47KB - jmir_v14i5e119_app2.pdf]

References

1. Black RE, Allen LH, Bhutta ZA, Caulfield LE, de Onis M, Ezzati M, Maternal and Child Undernutrition Study Group. Maternal and child undernutrition: global and regional exposures and health consequences. Lancet 2008 Jan 19;371(9608):243-260. [doi: 10.1016/S0140-6736(07)61690-0] [Medline: 18207566]
2. World Health Organization, Regional Office for the Western Pacific. WHO/UNICEF Regional Child Survival Strategy: Accelerated and Sustained Action towards MDG 4. Manila: World Health Organization; 2006.
3. WHO, UNICEF, USAID, AED. UCDAVIS, IFPRI. Indicators for Assessing Infant and Young Child Feeding Practices. Part 1: Definitions. Geneva: World Health Organization; 2008.
4. Integrated Management of Childhood Illness (IMCI) Implementation in the Western Pacific Region. World Health Organization, Western Pacific Region. 2012. IMCI and Child Survival URL: http://www.wpro.who.int/child_adolescent_health/documents/info_package1.pdf [accessed 2012-08-07] [WebCite Cache ID 69jTVHuVH]
5. Tomlinson M, Solomon W, Singh Y, Doherty T, Chopra M, Ijumba P, et al. The use of mobile phones as a data collection tool: a report from a household survey in South Africa. BMC Med Inform Decis Mak 2009;9:51 [FREE Full text] [doi: 10.1186/1472-6947-9-51] [Medline: 20030813]
6. The World in 2011. International Telecommunication Union. 2011. ICT Facts and Figures: One Third of the World's Population is Online URL: http://www.slideshare.net/cokyfauzialfi/ict-facts-figures2011 [accessed 2012-05-22] [WebCite Cache ID 67rginuEw]
7. mobiThinking. dotMobi. 2012 Mar. Global Mobile Statistics 2012: All Quality Mobile Marketing Research, Mobile Web Stats, Subscribers, Ad Revenue, Usage, Trends URL: http://mobithinking.com/mobile-marketing-tools/latest-mobile-stats [accessed 2012-05-22] [WebCite Cache ID 67rguSfkW]
8. Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. J Biomed Inform 2012 Feb;45(1):184-198. [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]

9. Thirumurthy H, Lester RT. M-health for health behaviour change in resource-limited settings: applications to HIV care and beyond. Bull World Health Organ 2012 May 1;90(5):390-392 [FREE Full text] [doi: 10.2471/BLT.11.099317] [Medline: 22589574]
10. Skinner D, Donald S, Rivette U, Ulrike R, Bloomberg C, Charissa B. Evaluation of use of cellphones to aid compliance with drug therapy for HIV patients. AIDS Care 2007 May;19(5):605-607. [doi: 10.1080/09540120701203378] [Medline: 17505920]
11. Barrington J, Wereko-Brobby O, Ward P, Mwafongo W, Kungulwe S. SMS for Life: a pilot project to improve anti-malarial drug supply management in rural Tanzania using standard technology. Malar J 2010;9:298 [FREE Full text] [doi: 10.1186/1475-2875-9-298] [Medline: 20979633]
12. Asiimwe C, Gelvin D, Lee E, Ben Amor Y, Quinto E, Katureebe C, et al. Use of an innovative, affordable, and open-source short message service-based tool to monitor malaria in remote areas of Uganda. Am J Trop Med Hyg 2011 Jul;85(1):26-33 [FREE Full text] [doi: 10.4269/ajtmh.2011.10-0528] [Medline: 21734120]
13. Haberer JE, Kiwanuka J, Nansera D, Wilson IB, Bangsberg DR. Challenges in using mobile phones for collection of antiretroviral therapy adherence data in a resource-limited setting. AIDS Behav 2010 Dec;14(6):1294-1301. [doi: 10.1007/s10461-010-9720-1] [Medline: 20532605]
14. Kaewkungwal J, Singhasivanon P, Khamsiriwatchara A, Sawang S, Meankaew P, Wechsart A. Application of smart phone in "Better Border Healthcare Program": a module for mother and child care. BMC Med Inform Decis Mak 2010;10:69 [FREE Full text] [doi: 10.1186/1472-6947-10-69] [Medline: 21047412]
15. Svoronos T, Mjungu D, Dhadialla P, Luk R, Zue C, Jackson J, et al. D-tree International. 2010. CommCare: Automated Quality Improvement to Strengthen Community-Based Health URL: http://d-tree.org/wp-content/uploads/2010/05/Svoronos-Medinfo-CommCare-safe-pregnancy1.pdf [accessed 2012-08-07] [WebCite Cache ID 69jXAt7hd]
16. Grameen Foundation. Computer Science & Engineering, University of Washington. 2011 Mar. Mobile Technology for Community Health in Ghana: What it is and What Grameen Foundation has Learned so Far URL: http://www.cs.washington.edu/education/courses/cse490d/12sp/docs/MOTECH.pdf [accessed 2012-08-07] [WebCite Cache ID 69jVENqdm]
17. Alam M, Khanam T, Khan R, Raihan A, Chowdhury M. ClickMedix, LLC. 2010. Assessing the Scope for Use of Mobile Based Solution to Improve Maternal and Child Health in Bangladesh: A Case Study on Efficiency of Community Health Workers, Automated Risk Assessment of Patients and Web Based Data Collection. (Working Paper for ICTD London Conference) URL: http://www.mobileactive.org/files/file_uploads/ICTD2010%20Alam%20et%20al.pdf [accessed 2012-08-31] [WebCite Cache ID 6AKNdKls0]
18. Shirima K, Mukasa O, Schellenberg JA, Manzi F, John D, Mushi A, et al. The use of personal digital assistants for data entry at the point of collection in a large household survey in southern Tanzania. Emerg Themes Epidemiol 2007;4:5 [FREE Full text] [doi: 10.1186/1742-7622-4-5] [Medline: 17543099]
19. Bowling A. Mode of questionnaire administration can have serious effects on data quality. J Public Health (Oxf) 2005 Sep;27(3):281-291 [FREE Full text] [doi: 10.1093/pubmed/fdi031] [Medline: 15870099]
20. Juniper EF, Langlands JM, Juniper BA. Patients may respond differently to paper and electronic versions of the same questionnaires. Respir Med 2009 Jun;103(6):932-934. [doi: 10.1016/j.rmed.2008.10.019] [Medline: 19028085]
21. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. Int J Nurs Stud 2011 Jun;48(6):661-671. [doi: 10.1016/j.ijnurstu.2011.01.016] [Medline: 21514934]
22. Lane SJ, Heddle NM, Arnold E, Walker I. A review of randomized controlled trials comparing the effectiveness of hand held computers with paper methods for data collection. BMC Med Inform Decis Mak 2006;6:23 [FREE Full text] [doi: 10.1186/1472-6947-6-23] [Medline: 16737535]
23. Vanden Eng JL, Wolkon A, Frolov AS, Terlouw DJ, Eliades MJ, Morgah K, et al. Use of handheld computers with global positioning systems for probability sampling and data entry in household surveys. Am J Trop Med Hyg 2007 Aug;77(2):393-399 [FREE Full text] [Medline: 17690421]
24. Byass P, Hounton S, Ouédraogo M, Somé H, Diallo I, Fottrell E, et al. Direct data capture using hand-held computers in rural Burkina Faso: experiences, benefits and lessons learnt. Trop Med Int Health 2008 Jul;13 Suppl 1:25-30. [doi: 10.1111/j.1365-3156.2008.02084.x] [Medline: 18578809]
25. Galliher JM, Stewart TV, Pathak PK, Werner JJ, Dickinson LM, Hickner JM. Data collection outcomes comparing paper forms with PDA forms in an office-based patient survey. Ann Fam Med 2008;6(2):154-160 [FREE Full text] [doi: 10.1370/afm.762] [Medline: 18332408]
26. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. Value Health 2008;11(2):322-333. [doi: 10.1111/j.1524-4733.2007.00231.x] [Medline: 18380645]
27. Quinn P, Goka J, Richardson H. Assessment of an electronic daily diary in patients with overactive bladder. BJU Int 2003 May;91(7):647-652. [Medline: 12699477]
28. Lal SO, Smith FW, Davis JP, Castro HY, Smith DW, Chinkes DL, et al. Palm computer demonstrates a fast and accurate means of burn data collection. J Burn Care Rehabil 2000;21(6):559-61; discussion 558. [Medline: 11194811]

29. Yu P, de Courten M, Pan E, Galea G, Pryor J. The development and evaluation of a PDA-based method for public health surveillance data collection in developing countries. Int J Med Inform 2009 Aug;78(8):532-542. [doi: 10.1016/j.ijmedinf.2009.03.002] [Medline: 19369114]
30. Rajput ZA, Mbugua S, Amadi D, Chepngeno V, Saleem JJ, Anokwa Y, et al. Evaluation of an Android-based mHealth system for population surveillance in developing countries. J Am Med Inform Assoc 2012 Jul 1;19(4):655-659. [doi: 10.1136/amiajnl-2011-000476] [Medline: 22366295]

Abbreviations

MNCH: Maternal, Newborn, and Child Health
PDA: personal digital assistant
WHO: World Health Organization

Edited by G Eysenbach; submitted 23.05.12; peer-reviewed by I Svab; comments to author 24.07.12; revised version received 07.08.12; accepted 06.08.12; published 18.09.12.

Please cite as:
Zhang S, Wu Q, van Velthoven MHMMT, Chen L, Car J, Rudan I, Zhang Y, Li Y, Scherpbier RW
Smartphone Versus Pen-and-Paper Data Collection of Infant Feeding Practices in Rural China
J Med Internet Res 2012;14(5):e119
URL: http://www.jmir.org/2012/5/e119/
doi:10.2196/jmir.2183
PMID:22989894

Original Paper

Classification Accuracies of Physical Activities Using Smartphone Motion Sensors

Wanmin Wu[1], PhD; Sanjoy Dasgupta[2], PhD; Ernesto E Ramirez[1], MS; Carlyn Peterson[1], MS; Gregory J Norman[1], PhD

[1]Center For Wireless & Population Health Systems, Department of Family & Preventive Medicine, University of California, San Diego, La Jolla, CA, United States

[2]Department of Computer Science and Engineering, University of California, San Diego, La Jolla, CA, United States

Corresponding Author:
Gregory J Norman, PhD
Center For Wireless & Population Health Systems
Department of Family & Preventive Medicine
University of California, San Diego
9500 Gilman Drive
Dept. 0811
La Jolla, CA, 92093-0811
United States
Phone: 1 (858)534 9302
Fax: 1 (858) 534 9404
Email: gnorman@ucsd.edu

Abstract

Background: Over the past few years, the world has witnessed an unprecedented growth in smartphone use. With sensors such as accelerometers and gyroscopes on board, smartphones have the potential to enhance our understanding of health behavior, in particular physical activity or the lack thereof. However, reliable and valid activity measurement using only a smartphone in situ has not been realized.

Objective: To examine the validity of the iPod Touch (Apple, Inc.) and particularly to understand the value of using gyroscopes for classifying types of physical activity, with the goal of creating a measurement and feedback system that easily integrates into individuals' daily living.

Methods: We collected accelerometer and gyroscope data for 16 participants on 13 activities with an iPod Touch, a device that has essentially the same sensors and computing platform as an iPhone. The 13 activities were sitting, walking, jogging, and going upstairs and downstairs at different paces. We extracted time and frequency features, including mean and variance of acceleration and gyroscope on each axis, vector magnitude of acceleration, and fast Fourier transform magnitude for each axis of acceleration. Different classifiers were compared using the Waikato Environment for Knowledge Analysis (WEKA) toolkit, including C4.5 (J48) decision tree, multilayer perception, naive Bayes, logistic, k-nearest neighbor (kNN), and meta-algorithms such as boosting and bagging. The 10-fold cross-validation protocol was used.

Results: Overall, the kNN classifier achieved the best accuracies: 52.3%–79.4% for up and down stair walking, 91.7% for jogging, 90.1%–94.1% for walking on a level ground, and 100% for sitting. A 2-second sliding window size with a 1-second overlap worked the best. Adding gyroscope measurements proved to be more beneficial than relying solely on accelerometer readings for all activities (with improvement ranging from 3.1% to 13.4%).

Conclusions: Common categories of physical activity and sedentary behavior (walking, jogging, and sitting) can be recognized with high accuracies using both the accelerometer and gyroscope onboard the iPod touch or iPhone. This suggests the potential of developing just-in-time classification and feedback tools on smartphones.

(J Med Internet Res 2012;14(5):e130) doi:10.2196/jmir.2208

KEYWORDS

Activity classification; machine learning; accelerometer; gyroscope; smartphone

Introduction

It is widely recognized that lack of physical activity and excess of sedentary behavior are associated with increased health risks for obesity, type 2 diabetes, cardiovascular disease, depression, and all-cause mortality [1]. In the United States, physical inactivity is alarmingly prevalent. A study based on the 2003-2004 National Health and Nutrition Examination Survey data suggested that on average Americans spend about 55% of their waking time, or 7.7 hours per day [2], being sedentary. Another study based on the same data showed that less than 5% of adults meet the national 30 minutes/day guideline for physical activity [3].

Measurement of physical activity and sedentary behavior is a fundamental, yet nontrivial, task for developing effective intervention tools. Self-reported data are subject to bias and errors. Objective methods enabled by advancement in accelerometer technologies are gaining increasing attention. Researchers have explored different accelerometer-equipped monitoring devices, such as customized sensor boards [4,5], Actigraph accelerometer [6,7], DynaPort [8], and Pegasus activity monitors [9], for detecting activities. Although some of these devices are small, they are still an extra burden for users to wear.

More recently, smartphones equipped with accelerometers have become ubiquitous. Carried by people throughout the day, smartphones are an ideal platform for monitoring physical activity and sedentary behavior and for just-in-time intervention. Furthermore, they have powerful computational capabilities and allow development of customized applications that integrate monitoring and intervention. Researchers have used accelerometers on Nokia N95 phones [10,11] and Android phones [12] to detect common activities such walking, stair climbing, jogging, and sitting.

Similar to earlier studies [10-12], in this study the goal was to create a valid activity classification tool that uses sensors onboard today's smartphones. However, this study is distinguished by the following three characteristics.

Device

We conducted our study with the iPod Touch (Apple Inc., Cupertino, CA, USA). It has essentially the same sensors and computing platform as the iPhone, yet costs much less. We compared the obtained results with findings from the two previous studies that used the Nokia N95 and Android phones.

Sensor

While past work relied mainly on accelerometers on cell phones, we combined acceleration with orientation readings from the newly available gyroscope sensor. In June 2010, Apple became the first to introduce gyroscopes to a mobile phone with the launch of the iPhone 4. Since then, an increasing number of mobile phones have added gyroscopes on board. However, few researchers have explored the use of gyroscopes as a way to measure physical activity. This study demonstrated one of the first steps in assessing whether gyroscope readings are beneficial in classifying activities. As accelerometers measure acceleration, gyroscopes measure rotation. Our hypothesis was that combining these two complementary sensors could improve recognition accuracy of activities.

Activity Intensity

Previous research [10-12] classified common physical activities such as walking on stairs and walking on level ground without differentiating speed. However, it has been shown that the intensity of these activities matters: walking at a normal pace is classified as light physical activity with an intensity of <3 metabolic equivalents (work metabolic rate/resting metabolic rate), whereas brisk walking is considered moderate physical activity with an intensity of 3-6 metabolic equivalents [13]. The US Centers for Disease Control and Prevention guideline on physical activity for adults is 150 minutes of moderate-intensity aerobic activity (eg, brisk walking) every week. Thus, we differentiated the speed of common activities, such as walking and stair climbing, at normal and brisk paces.

Methods

Data Collection

Recruitment

Eligibility criteria were being 19–60 years of age, speaking English, having no existing medical conditions that prevent performing moderate-intensity physical activity, and being able to climb and descend stairs. We drew a convenience sample from the University of California, San Diego. Men were recruited through the university's campuswide listserv. About 43 men responded with interest in participating and 6 were recruited. We recruited 5 women within the Center for Wireless & Population Health Systems in the California Institute for Telecommunications and Information Technology and 5 from a pool of potential study participants at the Moores Cancer Center, San Diego, CA, USA. All participants signed study consent forms and all protocols were approved by the university's institutional review board. Table 1 shows the characteristics of participants.

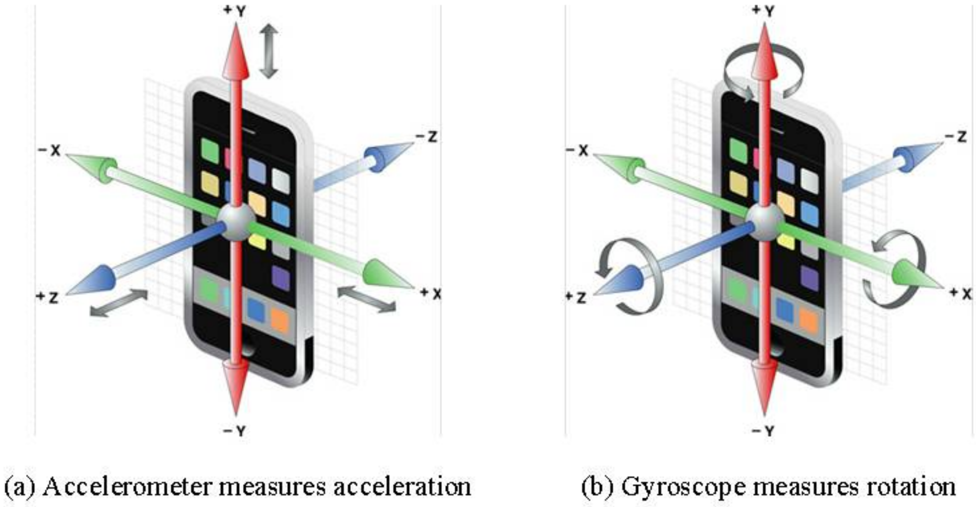
Hardware Platform

We used the iPod Touch (Apple Inc.) as the hardware platform for data collection. It has essentially the same accelerometer and gyroscope sensors as the iPhone (Apple Inc.) (Figure 1), one of the most widely used smartphones on the market. They run the same iOS operating system as well. The fourth-generation iPod Touch we employed used STMicroelectronics (Geneva, Switzerland) LIS331DLH accelerometer and L3G4200D gyroscope [14]. The iPod Touch is 11.18 cm high, 5.89 cm wide, and 0.71 cm deep, and weighs 100.9 g.

Table 1. Demographic characteristics of study participants (n = 16).

Characteristic	n	%
Age group (years)		
21–30	6	38%
31–40	3	19%
41–50	1	6%
51–60	6	38%
Gender		
Female	10	63%
Male	6	38%
Body mass index range (kg/m^2)		
18.5–24.9 (normal)	8	50%
25–29.9 (overweight)	1	6%
30–34.9 (moderately obese)	4	25%
35–40 (severely obese)	3	19%

Figure 1. Coordinate system of the iPod touch (figure courtesy of Apple Inc.).



(a) Accelerometer measures acceleration

(b) Gyroscope measures rotation

Software Setup

We developed an application on the iPod Touch for easy data collection. It allows users to specify an activity type and the device location (often referred to as the labeling or annotation step), start and stop data recording using a toggle button, and transmit collected data to research staff. We used the CMDeviceMotion class of iOS 4.2 application programming interface that encapsulates processed acceleration and gyroscope measurements. More specifically, we used the userAcceleration property of CMDeviceMotion to get the 3-axis acceleration (gravitational force) that the user imparts to the device (ie, total acceleration minus gravity), and the rotationRate property to get the device's rate of rotation (in radians per second) around three axes, with the gyroscope bias removed by Apple's proprietary Core Motion algorithms. Both accelerometer and gyroscope were configured to sample at a 30 Hz (33.33 milliseconds) rate.

Activities

We studied 13 activity types in total, 4 of which were paced by research staff in a laboratory setting on a treadmill, and the rest were self-paced by participants to simulate a free-living condition. The details of these activities are described in Table 2. As shown in the last column, these 13 activities were further grouped into 9 classes: slow walking, normal walking, brisk walking, jogging, sitting, normal upstairs, normal downstairs, brisk upstairs, and brisk downstairs. For example, the prescribed laboratory activity A2 (3.0 mph walking) and the self-paced activity A11 (400 m normal walking) belong to the same class, C2 (normal walking). That is, a classification system should be able to recognize both activities as normal walking.

Table 2. Physical activity type descriptions and classifications.

Activity name	Description	Class
Prescribed		
A1. 1.5 mph walking	Walking at 1.5 mph on a treadmill for 3 minutes	C1. Slow walking
A2. 3.0 mph walking	Walking at 3.0 mph on a treadmill for 3 minutes	C2. Normal walking
A3. 4.0 mph walking	Walking at 4.0 mph on a treadmill for 3 minutes	C3. Brisk walking
A4. 5.5 mph jogging	Jogging at 5.5 mph on a treadmill for 3 minutes	C4. Jogging
Self-paced		
A5. Sitting	Seated in a chair, remaining still	C5. Sitting
A6. Normal upstairs	Ascending one flight of stairs (19 steps) at a normal pace	C6. Normal upstairs
A7. Normal downstairs	Descending one flight of stairs (19 steps) at a normal pace	C7. Normal downstairs
A8. Brisk upstairs	Ascending one flight of stairs (19 steps) at a brisk pace	C8. Brisk upstairs
A9. Brisk downstairs	Descending one flight of stairs (19 steps) at a brisk pace	C9. Brisk downstairs
A10. 400 m slow walking	Walking for one lap around a 400 m track at a slow pace	C1. Slow walking
A11. 400 m normal walking	Walking for one lap around a 400 m track at a normal pace	C2. Normal walking
A12. 400 m brisk walking	Walking for one lap around a 400 m track at a brisk pace	C3. Brisk walking
A13. 400 m jogging	Jogging for one lap around a 400 m track.	C4. Jogging

Collection Protocol

Due to limited treadmill and track availability, not every participant completed the activities in the order specified in Table 2. Some participants followed that order, while others completed the 400 m track activities (A10–A13) first, followed by stair tasks (A6–A9), and finally the laboratory activities (A1–A5). The participant carried the iPod touch device in an armband for jogging and in a front shorts pocket for all other activities. When in the pocket, the device was oriented with the screen facing away from the body and the 30-pin connection port facing up. When the device was in the armband, the screen was faced away from the body with the 30-pin connection port facing down. Not all participants were able to complete all tasks. For example, approximately half of the participants were unable to perform all of the strenuous jogging activities (A4. 5.5 mph jog and A13. 400 m jog). The whole protocol took about 2 hours to complete. Each participant received US $50 as compensation.

Data Preprocessing

The time series of collected sensory data (30 Hz) were stored in a comma-separated values file per activity, per participant. The beginning and the end of all files were manually trimmed in a data preprocessing phase. This was because at the beginning of data recording the research staff had to start the data recording by pressing a toggle button and then place the device in the right position on the participant; at the end of the activity they collected the device and stopped data recording.

Feature Extraction

Window Size

The data vector containing 3-axis acceleration and 3-axis rotation rate recorded at a time instant is called a *sample*. To reduce noise and capture cyclic patterns of motion, features were not computed on each single sample, but on a sliding window of samples. Many studies have indicated the superiority of using a 1-second window size [4,11,15,16]; others have used larger window sizes such as 2 seconds [9] and 10 seconds [12] to capture more cyclic patterns. We experimentally compared window sizes of 1 second, 2 seconds, 5 seconds, and 10 seconds, and found that the 2-second window size (60 samples in our case) produced the best classification performance. The detailed comparison results are shown in the next section. Finally, the use of a 50% overlap between consecutive windows has been shown to be beneficial by past research [17]. Thus, we used a 1-second overlap in our sliding windows.

Features

As discussed before, we used an accelerometer and gyroscope as our signal sources. In the literature, different features have been employed for acceleration-based activity recognition, such as mean [4,5,9,12], variance [11,12], spectral entropy [4, 5], and fast Fourier transform coefficients [5,9,11,18]. To select features, we performed extensive comparative experiments on these various features. The following features (variants of 4 basic ones) produced the best classification results:

- *Mean* for each axis of acceleration, each axis of rotation rate (from gyroscope), and acceleration magnitude, computed as the square root of ($A^2_x + A^2_y + A^2_z$) in the sliding window
- *Standard deviation* for each axis of acceleration, each axis of rotation rate, and acceleration magnitude in the sliding window
- *Sum* of acceleration magnitude in the sliding window
- *Fast Fourier transform magnitude*, or magnitude of the first five coefficients of the fast Fourier transform power spectrum (for each axis of acceleration). As Preece et al [9] showed, this fast Fourier transform feature is overall the best-performing feature among all compared time, frequency, and wavelet features for all activities.

Results

Classifier Comparison

We used the Waikato Environment for Knowledge Analysis (WEKA) machine learning toolkit [19] to train and compare the performance of the classifiers C4.5 (J48) decision tree, multilayer perception, naive Bayes, logistic, and k-nearest neighbor (kNN). For all classifiers, the default WEKA settings (version 3-6-6) were used. We used 10-fold cross-validation for all experiments. Table 3 shows the comparison results using a 2-second window size. Among these basic classifiers, kNN generally produced the best accuracy results. Thus, we further applied meta-algorithms including boosting (AdaBoostM1) and bagging to the kNN classifier but observed no clear benefits as shown in Table 3.

Table 3. Comparison of classification accuracies by classifier.

Activity	kNN[a]	J48[b]	MLP[c]	Logistic	NB[d]	Boosting	Bagging
C1. Slow walking	94.1%	86.3%	90.8%	88.3%	61.3%	94.1%	94.1%
C2. Normal walking	92%	80.9%	84.6%	74.2%	55.7%	92%	92.2%
C3. Brisk Walking	90.1%	82.2%	85%	68.7%	64.9%	89.9%	90.1%
C4. Jogging	91.7%	91.7%	91.5%	92.2%	79%	92.2%	91.7%
C5. Sitting	100%	99.6%	100%	100%	98.5%	100%	100%
C6. Normal upstairs	69.8%	51%	42.7%	47.9%	30.2%	69.8%	69.8%
C7. Normal downstairs	79.4%	64.9%	54.6%	46.4%	32%	79.4%	77.3%
C8. Brisk upstairs	70.4%	69%	33.8%	19.7%	22.5%	70.4%	69%
C9. Brisk downstairs	52.3%	44.6%	24.6%	33.8%	35.4%	52.3%	43.1%
Weighted average	90.2%	83.0%	83.4%	77.2%	63.2%	90.2%	89.9%

[a] k-nearest neighbor.

[b] C4.5 decision tree.

[c] Multilayer perception.

[d] Naive Bayes.

In general, the kNN classifier achieved high accuracies for walking at different paces (90.1%–94.1%), jogging (91.7%), and sitting (100%). Stair walking proved to be the most challenging activity, with recognition accuracies ranging from 52.3% to 79.4%.

Table 4 presents the confusion matrix generated by kNN. Among all the misclassified sample segments (n = 274), a significant number were caused by the difficulty of differentiating walking at different speeds (n = 101) and differentiating walking on stairs from walking on a level ground (n = 103). Fortunately, compared with walking on level ground (which has a classification accuracy of 90.1%–94.1%), stair walking is only a small part of daily activities for most people. Confusion also existed between brisk walking and jogging (with n = 57 sample segments incorrectly classified). This may be due to the different speeds participants used in self-paced situations. It is interesting to focus on the spectrum from slow walking to jogging. These activities were almost never confused with staircase motion or with sitting. Moreover, the predicted activity was almost always either correct or one speed gradation off; for instance, slow walking was never mistaken for jogging.

Table 4. Confusion matrix (k-nearest neighbor classifier with accelerometer and gyroscope features).

Activity	Classified as...								
	C1	C2	C3	C4	C5	C6	C7	C8	C9
C1 = Slow walking	572	30	5	0	0	0	0	0	0
C2 = Normal walking	29	602	13	0	0	4	5	0	1
C3 = Brisk walking	7	17	475	25	0	0	1	0	2
C4 = Jogging	0	1	32	389	0	0	0	2	0
C5 = Sitting	0	0	0	0	266	0	0	0	0
C6 = Normal upstairs	8	15	2	0	0	67	4	0	0
C7 = Normal downstairs	6	7	3	0	0	4	77	0	0
C8 = Brisk upstairs	1	8	10	1	0	0	0	50	1
C9 = Brisk downstairs	0	16	14	0	0	1	0	0	34

Window Size Comparison

As discussed above, different window sizes have been used in the literature, including 1 second [3,9,15], 2 seconds [20], and 10 seconds [11]. We experimentally compared window sizes of 1 second, 2 seconds, 5 seconds, and 10 seconds using the kNN classifier, with overall accuracies of 87.7%, 90.2%, 88.5%, and 84.2%, respectively. The 2-second window size achieved the best overall classification performance in terms of weighted average accuracy.

Effect of Gyroscope

We made one of the first attempts to evaluate the effect of a gyroscope in measuring physical activities. Our hypothesis was that adding gyroscope data could improve the overall classification accuracy. This was confirmed by the results as shown in Table 5. Using both rotation rate (from the gyroscope) and acceleration features (from the accelerometer) with kNN resulted in higher accuracies for all activity classes than when using only acceleration features, with improvement ranging from 3.1% to 13.4%.

Gyroscope data are useful because almost all activities involve some sort of orientation change of the phone. This makes it a powerful complementary data source to the accelerometer, which only measures linear motion along specified directions.

Table 5. A comparison of classification accuracies using acceleration features only versus using both acceleration and rotation rate features (k-nearest neighbor classifier).

Activity	Acceleration	Acceleration+ rotation rate	Difference
C1. Slow walking	89.6%	94.1%	+4.5%
C2. Normal walking	85.8%	92%	+6.2%
C3. Brisk walking	78%	90.1%	+12.1%
C4. Jogging	85.4%	91.7%	+6.3%
C5. Sitting	100%	100%	0%
C6. Normal upstairs	65.6%	69.8%	+4.2%
C7. Normal downstairs	66%	79.4%	+13.4%
C8. Brisk upstairs	64.8%	70.4%	+5.6%
C9. Brisk downstairs	49.2%	52.3%	+3.1%
Weighted average	83.7%	90.2%	+6.5%

Comparison With Prior Work

Table 6 shows a comparison of classification accuracies obtained in our study against those reported in three previous studies. However, the differences between studies should be interpreted with caution because they can be attributed to many factors (as listed in Table 7). The most significant factor is that different datasets were used in each study.

The lack of a shared dataset in the research community makes cross-study comparison difficult, particularly on feature types, sliding window sizes, and classifiers. To accelerate future research on assessment of activity using smartphones, we are sharing our anonymized iPod touch dataset with the research community. The dataset is accessible through the iDash Data Repository [20].

Table 6. Accuracy comparison with prior work.

Activity	Lu et al [10]	Reddy et al [11]	Kwapisz et al [12]	Current study
Sitting (still)	97.7%	95.6%	95%	100%
Walking	96.6%	96.8%	91.7%	94.1% (slow), 92% (normal), 90.1% (brisk)
Running or jogging	98%	91.0%	98.3%	91.7%
Upstairs	ND[a]	ND	61.5%	69.8% (normal), 70.4% (brisk)
Downstairs	ND	ND	44.3%	79.4% (normal), 52.3% (brisk)

[a] Not done.

Table 7. Methodology comparison with prior work.

Feature	Lu et al [10]	Reddy et al [11]	Kwapisz et al [12]	Current study
Device	Nokia N95 phone (iPhone unevaluated)	Nokia N95 phone	Android phones	iPhone/iPod Touch
Signal sources	Accelerometer	Accelerometer	Accelerometer, global positioning system	Accelerometer, gyroscope
Features	Mean, variance, mean crossing rate, spectrum peak, sub-band energy (ratio), spectral entropy	Mean, SD, average, absolute difference, magnitude, time between peaks, and binned distribution	Mean, variance, energy, and the density functional theory energy coefficients	Mean, SD, sum, magnitude, fast Fourier transform magnitude
Window sizes	4 seconds	1 second	10 seconds	2 seconds
Classifier	Support vector machine	Decision tree + Discrete Hidden Markov Model	Multilayer perception	k-nearest neighbor
Activities	Cycling, running, being still, riding in a vehicle, walking	Cycling, running, being still, riding in a vehicle, walking	Upstairs, downstairs, running, being still, walking	Upstairs (at different speeds), downstairs (at different speeds), running, being still, walking (at different speeds)

Discussion

Principal Results

This study is the first step in our effort to develop integrated tools to measure and intervene in physical activity and sedentary behavior. Combining time and frequency features of both acceleration and gyroscope measurements from sensors onboard smartphones, we classified common categories of physical activity and sedentary behavior (sitting, walking, and jogging at different paces) with high accuracy (90.1%–94.1%); up and down stair walking were classified at 52.3%–79.4% accuracies. Including orientation readings from a gyroscope proved to be beneficial for recognizing all activities studied.

Limitation

We collected data using a convenience sample of participants. As motion pattern varies with individuals, future studies would benefit from using multiple demographic and physiological variables to inform participant designs. Furthermore, data were collected with the device placed in specific positions (armband for jogging, and shorts pocket for other activities). Jogging with only armband placement of the device likely influenced the signal pattern for this activity and may have contributed to the high classification accuracy we observed. Further investigation is needed to evaluate classification accuracies with more variable placement of the device (eg, hand, back pants pocket, or backpack). It will also be necessary to test the accuracy of activity classification in a free-living context, where individuals make natural transitions between activities such as sitting to standing and jogging to walking. Machine learning algorithms for classification often benefit from having diverse observations or subjects, because the machine can then learn more patterns of individual movement. Therefore, applying our classification methodology (features, window size, and classifiers) to a larger dataset would most likely result in higher accuracies.

We focused on classification of a somewhat narrow range of activities that pertained to ambulatory movements and sitting posture. Including other activities such as bicycling will be important to more fully capture the spectrum of physical activities in which people engage. However, classifying a wider range of activities might result in lower accuracies than were obtained in this study.

The nature of the false-positives shown in the confusion matrix was that when activities were misclassified it was usually by one speed gradation (eg, brisk walking misclassified as normal walking or jogging). This suggests that it might be possible to significantly improve accuracy by calibrating the prediction thresholds to individual users. This is an important area to explore in future work.

Conclusion

This study is among the first to validate smartphone sensors including an accelerometer and gyroscope for activity recognition. The results suggest clear benefits of using a gyroscope as an additional data source for classifying activities. Including other signal data sources from the phone such as its global positioning system may further improve the system, but only for specifically identifying outdoor activities, and with the potential cost of reducing the battery life of the smartphone. Other sensors such as heart rate monitors might also further improve identifying the intensity of activities (eg, brisk walking compared with jogging). However, the trade-off of the extra burden of wearing an additional sensor would limit the public health impact of our system.

This study provided important indications of the possibilities and limitations of using a smartphone as an activity data collector. This system has potential high ecological validity because it requires people to carry only one device that they commonly carry with them already. The next step in our research is to test an onboard classifier application on the phone that can prompt users when needed for annotations in order to learn and classify individual activity patterns with high accuracy. The final step will be testing the feedback component that can offer individually tailored prompts and suggestions to increase physical activity and decrease sedentary time.

Acknowledgments

We would like to thank all the study participants who helped with data collection. This project is supported by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54HL10846. The authors reported no financial disclosures.

Conflicts of Interest

None declared.

References

1. Bauman AE. Updating the evidence that physical activity is good for health: an epidemiological review 2000-2003. J Sci Med Sport 2004 Apr;7(1 Suppl):6-19. [Medline: 15214597]
2. Matthews CE, Chen KY, Freedson PS, Buchowski MS, Beech BM, Pate RR, et al. Amount of time spent in sedentary behaviors in the United States, 2003-2004. Am J Epidemiol 2008 Apr 1;167(7):875-881 [FREE Full text] [doi: 10.1093/aje/kwm390] [Medline: 18303006]
3. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
4. Bao L, Intille S. Activity recognition from user-annotated acceleration data. Lect Notes Comput Sci 2004;3001:1-17. [doi: 10.1007/978-3-540-24646-6_1]
5. Lester J, Choudhury T, Kern N, Borriello G, Hannaford B. A hybrid discriminative/generative approach for modeling human activities. In: Proceedings. 2005 Presented at: International Joint Conference on Artificial Intelligence (IJCAI); Jul 30-Aug 5, 2005; Edinburgh, UK URL: http://ijcai.org/Past%20Proceedings/IJCAI-05/PDF/0888.pdf
6. Albinali F, Intille S, Haskell W, Rosenberger M. Using wearable activity type detection to improve physical activity energy expenditure estimation. In: Proceedings. 2010 Presented at: 12th ACM international conference on Ubiquitous computing (UbiComp '10); Sep 26-29, 2010; Copenhagen, Denmark p. 311-320.
7. De Vries SI, Garre FG, Engbers LH, Hildebrandt VH, Van Buuren S. Evaluation of neural networks to identify types of activity using accelerometers. Med Sci Sports Exerc 2011 Jan;43(1):101-107. [doi: 10.1249/MSS.0b013e3181e5797d] [Medline: 20473226]
8. van Hees VT, van Lummel RC, Westerterp KR. Estimating activity-related energy expenditure under sedentary conditions using a tri-axial seismic accelerometer. Obesity (Silver Spring) 2009 Jun;17(6):1287-1292 [FREE Full text] [doi: 10.1038/oby.2009.55] [Medline: 19282829]
9. Preece SJ, Goulermas JY, Kenney LP, Howard D. A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. IEEE Trans Biomed Eng 2009 Mar;56(3):871-879. [doi: 10.1109/TBME.2008.2006190] [Medline: 19272902]
10. Lu H, Yang J, Liu Z, Lane ND, Choudhury T, Campbell TA. The Jigsaw continuous sensing engine for mobile phone applications. In: Proceedings. 2010 Presented at: 8th ACM Conference on Embedded Networked Sensor Systems (SenSys '10); Nov 3-5, 2010; Zurich, Switzerland p. 71-84.
11. Reddy S, Mun M, Burke J, Estrin D, Hansen M, Srivastava MB. Using mobile phones to determine transportation modes. ACM Trans Sens Netw 2010;6(2).
12. Kwapisz JR, Weiss GM, Moore SA. Activity recognition using cell phone accelerometers. SIGKDD Explor 2010;12(2):74-82 [FREE Full text]

13. Pate RR, Pratt M, Blair SN, Haskell WL, Macera CA, Bouchard C, et al. Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. JAMA 1995 Feb 1;273(5):402-407. [Medline: 7823386]
14. Dixon-Warren SJ. MEMS J. 2010. Motion sensing in the iPhone 4: MEMS accelerometer URL: http://www.memsjournal.com/2010/12/motion-sensing-in-the-iphone-4-mems-accelerometer.html [accessed 2012-09-12] [WebCite Cache ID 6AdEecOGW]
15. Huynh T, Schiele B. Analyzing features for activity recognition. In: Proceedings. 2005 Presented at: Conference on Smart Objects and Ambient Intelligence; Oct 12-14, 2005; Grenoble, France.
16. Kern N, Schiele B, Schmidt A. Multi-sensor activity context detection for wearable computing. Lect Notes Comput Sci 2003;2875:220-232. [doi: 10.1007/978-3-540-39863-9_17]
17. van Laerhoven K, Cakmakci O. What shall we teach our pants? 2000 Presented at: Fourth International Symposium on Wearable Computers (ISWC 2000); Oct 18-21, 2000; Atlanta, GA, USA.
18. Tapia EM, Intille SS, Haskell W, Larson K, Wright J, King A, et al. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. 2007 Presented at: 11th IEEE International Symposium on Wearable Computers; Sep 28-Oct 1, 2007; Boston, MA. USA.
19. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. SIGKDD Explor 2009;11(1):10-18 [FREE Full text]
20. University of California, San Diego. 2012 May 06. iDash: Integrating Data for Analysis, Anonymization, and Sharing URL: http://idash.ucsd.edu/ [accessed 2012-05-05] [WebCite Cache ID 67S5XRWOx]

Abbreviations

kNN: k-nearest neighbor
WEKA: Waikato Environment for Knowledge Analysis

Edited by G Eysenbach; submitted 07.06.12; peer-reviewed by G Dunton, E Hekler, V Gay; comments to author 26.06.12; revised version received 11.07.12; accepted 12.07.12; published 05.10.12.

Please cite as:
Wu W, Dasgupta S, Ramirez EE, Peterson C, Norman GJ
Classification Accuracies of Physical Activities Using Smartphone Motion Sensors
J Med Internet Res 2012;14(5):e130
URL: http://www.jmir.org/2012/5/e130/
doi:10.2196/jmir.2208
PMID:23041431

Original Paper

Clinicians' Expectations of Web 2.0 as a Mechanism for Knowledge Transfer of Stroke Best Practices

Isabelle David[1], BSC, MSc; Lise Poissant[1,2], PhD; Annie Rochette[2,3], PhD

[1]Institut de réadaptation Gingras-Lindsay-de-Montréal, Montreal, QC, Canada
[2]School of Rehabilitation, Faculty of Medicine, University of Montreal, Montreal, QC, Canada
[3]Centre de réadaptation Lucie-Bruneau, Montreal, QC, Canada

Corresponding Author:
Lise Poissant, PhD
Institut de réadaptation Gingras-Lindsay-de-Montréal
6300, Darlington Ave.
Montreal, QC, H3S 2J4
Canada
Phone: 1 514 340 2111 ext 2277
Fax: 1 514 340 2154
Email: lise.poissant@umontreal.ca
School of Rehabilitation
Faculty of Medicine
University of Montreal
7077, Parc Ave.
Montreal, QC,
Canada
Phone: 1 514 340 2111 ext 2277
Fax: 1 514 340 2154
Email: lise.poissant@umontreal.ca

Abstract

Background: Health professionals are increasingly encouraged to adopt an evidence-based practice to ensure greater efficiency of their services. To promote this practice, several strategies exist: distribution of educational materials, local consensus processes, educational outreach visits, local opinion leaders, and reminders. Despite these strategies, gaps continue to be observed between practice and scientific evidence. Therefore, it is important to implement innovative knowledge transfer strategies that will change health professionals' practices. Through its interactive capacities, Web 2.0 applications are worth exploring. As an example, virtual communities of practice have already begun to influence professional practice.

Objective: This study was initially developed to help design a Web 2.0 platform for health professionals working with stroke patients. The aim was to gain a better understanding of professionals' perceptions of Web 2.0 before the development of the platform.

Methods: A qualitative study following a phenomenological approach was chosen. We conducted individual semi-structured interviews with clinicians and managers. Interview transcripts were subjected to a content analysis.

Results: Twenty-four female clinicians and managers in Quebec, Canada, aged 28-66 participated. Most participants identified knowledge transfer as the most useful outcome of a Web 2.0 platform. Respondents also expressed their need for a user-friendly platform. Accessibility to a computer and the Internet, features of the Web 2.0 platform, user support, technology skills, and previous technological experience were found to influence perceived ease of use and usefulness. Our results show that the perceived lack of time of health professionals has an influence on perceived behavioral intention to use it despite favorable perception of the usefulness of the Web 2.0 platform.

Conclusions: In conclusion, female health professionals in Quebec believe that Web 2.0 may be a useful mechanism for knowledge transfer. However, lack of time and lack of technological skills may limit their use of a future Web 2.0 platform. Further studies are required with other populations and in other regions to confirm these findings.

(J Med Internet Res 2012;14(5):e121) doi:10.2196/jmir.2016

KEYWORDS

Qualitative research; health personnel; Internet; evidence-based practice; information dissemination

Introduction

Gaps continue to be observed between professional practice and scientific evidence [1-3]. To address this situation, health professionals are increasingly encouraged to adopt an evidence-based practice. In Canada, a significant proportion of scientific evidence is not integrated into health care services offered to stroke patients [4-7]. The application of an evidence-based practice requires the implementation of resources and tools facilitating knowledge transfer and exchange between researchers and professionals.

According to the Cochrane Effective Practice and Organisation of Care (EPOC) taxonomy [8], 10 interventions are known to improve the practice of health care professionals: (1) distribution of educational materials, (2) educational meetings, (3) local consensus processes, (4) educational outreach visits, (5) local opinion leaders, (6) patient-mediated interventions, (7) audit and feedback, (8) reminders, (9) marketing, and (10) mass media. In the stroke field, 61 practice guidelines or educational materials exist worldwide according to a subcommittee of the World Stroke Organization [9]. One of those guidelines is published through the combined efforts of the Canadian Stroke Network and the Heart and Stroke Foundation of Canada. This guideline, the *Canadian Best Practice Recommendations for Stroke Care* [10], is published every 2 years (since 2006). Following this national initiative, each province has identified priorities and initiated an approach specific to their province.

In addition to educational materials, various knowledge transfer strategies have emerged in stroke care to reduce gaps between research and practice. Those strategies hardly fit in the EPOC taxonomy. As an example, the Montreal Stroke Network was created in 2002 independently of the national and provincial initiatives mentioned previously. This network has linked three communities of practice (prevention, acute care, and rehabilitation) and is composed of professionals working with stroke patients in Greater Montreal, a large urban city in Quebec, Canada. As presented in Poissant [11], the objective of the Montreal Stroke Network members is to improve the quality of professional practice in order to better meet the needs of stroke survivors across the continuum of care and in the community. The Montreal Stroke Network brings together a large variety of stakeholders (eg, managers, occupational therapists, nurses, physicians, speech-language pathologists, and social workers) working in different organizations (eg, hospitals, intensive functional rehabilitation units, and rehabilitation centers focusing on social integration). Communities of practice within the Montreal Stroke Network successfully developed and implemented several clinical projects.

In addition to the Montreal Stroke Network, initiatives such as Stroke Engine (www.strokengine.ca) [12] and Evidence-Based Review of Stroke Rehabilitation (www.ebrsr.com) [13] are available online for clinicians, managers, patients, and families to improve stroke care. Those strategies demonstrate that the Internet is well integrated into health care delivery. A study conducted among 243 health professionals (general practitioners, practice nurses, and practice managers) showed that 81% of them use the Internet for their work [14]. However, using the Internet for the sole purpose of accessing knowledge is unlikely to translate into practice change [15]. Conversely, accessing the Web to get involved in a community of practice is likely to lead to behavioral changes.

By implementing virtual communities of practice, communities can take advantage of new Internet applications. Web 2.0 is a “new generation of the Internet where interaction is important” [16]. Available interactive capacities within Web 2.0 facilitate information exchange among organizations independent of their geographic location. According to several authors [17-20], blogs, podcasts, and wikis are the most popular Web 2.0 applications. There are several other Web 2.0 applications (eg, virtual libraries and discussion forums). Several examples of virtual communities of practice are published in the health care literature [21,22]; however, little is known about the factors that will play a role in transforming a “face-to-face” community of practice into a virtual community of practice using Web 2.0 applications.

The main objective of this study is to document health professionals’ perception of a future Web 2.0 platform supporting knowledge transfer and implementation of stroke clinical practice guidelines. The secondary objective is to identify differences in perceptions between members of the Montreal Stroke Network and professionals outside this network possibly interested in joining it via a Web 2.0 platform.

Methods

Design

To understand professionals’ experiences regarding Web 2.0 being introduced into their practices, a phenomenological approach was chosen. Phenomenology, as described by Husserl in 1910, focuses on the essential structure of individual experiences. It allows for study of what is behind a phenomenon by giving the participant the opportunity to express freely and openly about the phenomenon. The analysis of the expressed content demonstrates the essence of what is perceived as inevitable following a rigorous process of reduction [23].

Population

Clinicians and managers working with stroke patients in a health organization of Quebec were asked to participate in the study. To compare the needs of Montreal Stroke Network members with those outside this network, 3 groups were targeted. The first group consisted of members of the Montreal Stroke Network. The second group was composed of health professionals who were considered members of the network, but who remained inactive over the previous 2 years. Lastly, the third group gathered health professionals who did not know about the network and who were working in remote areas. The target sample size was 20 people to have a holistic representation of each profession in every stroke care continuum phase (ie, acute care, intensive functional rehabilitation, and community reintegration).

Recruitment

We used a purposive sampling strategy to ensure adequate representativeness of groups. Therefore, we used the membership list of the Montreal Stroke Network to identify and contact participants for the first 2 groups (active and inactive Montreal Stroke Network members). We contacted research coordinators of rehabilitation centers and acute care hospitals geographically distant from the city of Montreal and asked them to identify health professionals who could potentially participate in the study. Potential participants were contacted by email or phone.

Data Collection Method

We conducted individual semi-structured interviews with each participant from January to July 2010. Semi-structured interviews were preferred as the data collection method because they are consistent with a phenomenological approach. The interview guide was developed on the basis of available scientific literature on Web 2.0 capabilities, communities of practice, and technology acceptance. A first version of the interview guide was pretested with a health professional to verify interview length and clarity of the questions. Revisions were made by the research team. The interview guide was structured as follows. First, participants explained their job particulars, then they discussed their perception and their needs regarding a Web 2.0 platform more specifically. Issues related to strategies used to share knowledge, benefits of a community of practice, and technology use were also discussed. All interviews were conducted by the same member of the research team (ID). Because the interviews were conducted prior to the development of the platform, key Web 2.0 definitions were provided to participants to ensure standardized comprehension of the concepts. The length of interviews varied from 40 to 75 minutes. The interviews were conducted in French or in English at the participant's convenience. A reflective research diary was used for gathering information about contextualization and interpretation of data. Most interviews were face-to-face, but 5 interviews were conducted by telephone due to geographical distance. Written consent was obtained from all participants.

Data Analysis

The audio content of each interview was transcribed. A member of the research team (ID) verified the accuracy of the transcripts and conducted the content analysis. Two other research team members (LP and AR) reviewed and discussed the codes to make sure they had face validity. First, codes were assigned to units of meaning and they were grouped into themes. Next, the research team identified links between identified codes. Only codes that were considered essential to answer the research question were kept. Qualitative data analysis software (QDA Miner 3.2.6) was used for data management and to support a systematic analytic approach by the research team. The research diary was used to refine the results and keep track of decisions made during the analysis.

Ethics Approval

This study was approved by the Ethics Committee of the Centre for Interdisciplinary Research in Rehabilitation of Greater Montreal.

Results

Sample Description

Twenty-four health professionals agreed to be interviewed: 9 were active Montreal Stroke Network members ("Active members"), 9 were inactive Montreal Stroke Network members ("Inactive members"), and 6 were in the remote areas group ("Remote members") (see Table 1). Six people didn't answer the invitation and 1 person refused. The distribution of stroke patients seen by health professionals within each group was similar. The interviews were conducted in English with 2 participants (participants A3 and I9). All participants were women with an average age of 45 years (SD 9.64). Mean professional experience was 18 years (SD 8.75). Nearly two-thirds had access to an individual computer. Participants in the Active members group were significantly older (mean 51 years, range 39-64, *P* = .02) than participants in the Remote members group (mean 38 years, range 28-46 years).

Table 1. Characteristics of participants.

	Participant	Age (years)	Profession[a]	Professional experience (years)	Experience with stroke patient (years)	Clinical time (%)	Workplace[b]	Ratio (computer/individual)
Active members								
	A1	48	OT	25	19	50	IFRU	1:10
	A2	48	PT	26	22	100	IFRU	3:10
	A3	66	SLP	30	25	66	Acute care	1:1
	A4	62	Program manager	18	4	0	Community reintegration	1:1
	A5	40	Nurse	18	10	100	IFRU	1:1
	A6	39	Clinical-administrative manager	6	1	0	IFRU	1:1
	A7	47	Clinical nurse specialist	24	20	90	Acute care	1:1
	A8	49	OT	25	20	98	Acute care	3:5
	A9	64	Neurologist	32	20	50	Acute care	1:1
Inactive members								
	I1	39	OT	15	15	100	Acute care	3:5
	I2	44	OT	21	20	100	Health and social services center	1:2
	I3	46	Consultant nurse	23	11	0	Acute care	1:1
	I5	47	PT	24	5	100	Acute care	3:11
	I6	40	PT	17	10	100	Specialized acute care	1:6
	I7	55	SW	33	25	100	Acute care	1:1
	I8	33	Nurse	9	5	80	Acute care	1:1
	I9	48	Neurologist	14	14	85	Acute care	1:1
Remote members								
	R1	28	PT	3.5	3.5	100	Acute care	1:5
	R2	30	OT	7	2	100	Community reintegration	1:1
	R3	40	Clinical nurse specialist	10	2	100	Community reintegration	1:1
	R4	46	Neuro-psychologist	25	11	100	IFRU	2:5
	R5	46	Clinical coordinator	20	9	0	IFRU and community reintegration	1:1
	R6	37	Program manager	5	5	0	IFRU and community reintegration	1:1

[a] OT: occupational therapist; PT: physical therapist; SLP: speech-language pathologist; SW: social worker

[b] IFRU: intensive functional rehabilitation unit

Results of Objective 1: Health Professionals' Perception

Four themes documenting perception of Web 2.0 emerged from the interviews: (1) influence of external variables, (2) perceived usefulness, (3) perceived ease of use, and (4) time availability. External variables were associated to accessibility, system features, user support, technologic skills, tool experience, and profession. Those variables had an impact on perceived usefulness and perceived ease of use. Perceived usefulness was characterized with concepts referring to knowledge transfer, quality of care, and efficiency. Perceived ease of use was associated with 2 codes: user-friendliness and timely access to information. The results indicated that participant's profession, one of the external variables, influences the time availability for technology utilization. Finally, the combination of perceived usefulness and perceived ease of use was creating the behavioral intention to use the technology. The results suggested that time availability influences system utilization in addition to behavioral intention. Examples of quotes supporting these results are presented in Table 2.

Table 2. Themes, codes, definitions, and sample statements.

Theme	Code	Definition	Statement
External variables	Accessibility	Computers performance and availability; quality of the Internet connection at work	"We have old computers" (participant I1)
	System features	Characteristics of Web 2.0 applications and exchanges	See Table 3
	User support	Informatics support available at work	"We have an informatics department here. Whenever something doesn't work, we call them and they can fix it rapidly normally" (participant A2)
	Technological skills	Level of skills to use computer and the Internet	"I am skilled to do what I have to do" (participant A7)
	Tool experience	Previous emotional experience with a Web 2.0 platform	"I could spend my whole day here. I had to stop, because it could take my whole day, I'm very interested, it could take up my whole day" (participant A3)
Perceived usefulness	Facilitates knowledge transfer	Opportunity to learn and stay up-to-date through exchanges with other members of a Web 2.0 platform	"I think it might be interesting, the opening, to have contact with other people working in the same field. Especially if it's interactive, I think that's good too, to have access to information with an easier way than right now" (participant A1); "To have access to what is done in other organizations" (participant I3); and "To seek the others' expertise." (participant I6)
	Increased quality of care	Opportunity to make changes to improve care through a Web 2.0 platform	"It can get answers to people...which may be useful in their practice" (participant A4)
	Allows tasks to be accomplished more quickly	Opportunity to decrease time spent to search information or to do other tasks through a Web 2.0 platform	"Why reinvent something when it already exists?" (participant R5)
	Perceived uselessness	Personal and organizational resistance to change	"From a management point of view, I have a concern with how clinicians will use this tool and how much time they will spend on it" (participant A6)
Perceived ease of use	User-friendly	Intuitive learning of a Web 2.0 platform	"If it's complicated, it might unmotivate me" (participant I7)
	Timely access to information	Optimizing the time fit between an informational need and its answer through a Web 2.0 platform	"If I have a problem, I need a quick response" (participant A7)
Time availability	-	Available time to learn about best practices and to search on the Internet	"Nobody has the time to do that" (participant I8)
Behavioral intention to use	Positive	Expected platform use	"By using it, if everything is going well, I will use it more and more often" (participant R4)
	Negative	Unexpected platform use	"I don't think I will go on it (participant the platform) every day" (participant I8)

External Variables

Within external variables influencing perceived usefulness and perceived ease of use, two were related to the technology (accessibility and features of the Web 2.0 platform), three were related to users (technologic skills, tool experience, and profession), and one was related to technological support provided to users. Accessibility to computers appeared to be influenced by the type of profession. Indeed, the ratio of computers to professionals could be as low as 1:10 for occupational therapists and physical therapists (see Table 1). In comparison, neurologists, speech-language pathologists, clinical nurse specialists, and workers with a management role (program manager or clinical coordinator) each had a personal computer (1:1 ratio). Many participants also reported having access to poorly performing computers and slow Internet connections at work.

Because of its complexity and in light of our interview process, 5 subcodes were attributed to the features of the Web 2.0 platform. These subcodes were identification, Web 2.0 applications, animation, look and feel, and membership fees (see Table 3).

Table 3. System features of subcodes.

Subcode	Definition	Statement
Identification	Password to access the platform. Once connected, personal information is revealed.	"Personally, I wouldn't have trouble identifying myself: where I'm from, what is my profession, where I work, my name...But some people might be less comfortable with this" (participant A5)
Web 2.0 applications	Perceived relevance of Web 2.0 applications (eg, blogs, podcast, and forums). Concerns about the quality, the relevance, and the variety of exchanges.	"I want quality answers" (participant R3)
Animation	Designated person to stimulate and organize exchanges.	"Someone will monitor that? Someone will manage that?" (participant A1)
Look and feel	Platform visual.	"If it is attractive, it is for sure an advantage" (participant I5)
Membership fees	Money to pay for the membership.	"When it's free, it's evident that I will try it for a time period" (participant I3)

Identification referred to the log-in process to start using the Web 2.0 platform and to the identification of participants once on the platform. Given the general abundance of passwords individuals have to deal with, participants perceived a secured log-in process as a barrier. Once on the platform, the respondents' perception of the importance of identifying themselves by their name, profession, and workplace was divided. Half of the participants felt that names should appear, one-quarter believed that people should have the choice to identify themselves or not, and one-quarter did not want to identify themselves at all. In terms of accessing information about a members' profession, more than half agreed with it being displayed, and no one completely disagreed. The remaining respondents were ambivalent or would give members the choice to display it or not. Most participants agreed that anonymity could influence the assessment of the quality of the information.

Professionals didn't express issues regarding patient confidentiality within virtual exchanges. It seemed like they were fully aware that written communication within a Web platform must ensure confidentiality and respect ethics rules. However, some of them were enthusiastic and some were reluctant about patients' prospective access to a section of the platform.

High variability was seen among participants about the most relevant Web 2.0 applications for professional practice. Nevertheless, some preferences were observed. Access to videos and a list of pertinent websites were preferred to podcasts and wikis. Two other potentially useful applications were virtual libraries and calendars because they facilitate access to documents and events that shared by colleagues. Email notifications were appealing to professionals. Although some were apprehensive about the potentially large volume of notifications, others saw notifications as a means to save time. Asynchronous exchanges on a Web 2.0 platform emerged as more applicable to professional practice and time management than synchronous exchanges. A preference was also noted in favor of discussion forums rather than blogs. Professionals gave added value to forums that were perceived as more time efficient because of the way the threads are organized. Participants mentioned that the rating of forum threads would be more useful than knowing the number of times the thread was read. Respondents believed that threads must be relevant and contain reliable and quality information.

> *"It is easy to write things in the forums, but...there are those who get carried away and put too much."* [participant I6]
>
> *In fact, some professionals only wanted to access expert opinions: "The problem that I see with forums is that you have to read the opinion of everybody...All I want is the expert."* [participant I8]

In addition to this concern for quality, discussion or information sharing should give professionals the feeling that they are learning while staying focused on clinical practice.

> *"If the level of discussions weren't interesting, I would eventually stop. Or if the level was, you know...the level was interesting and the questions were serious, then I'd go."* [participant A3]
>
> *"Theoretical and scientific information is easy to access. I think that it is more the things in day-to-day life, information in the field that are...that are less accessible."* [participant R5]

An expressed challenge was the importance of having varied topics to interest all members regardless of their profession and workplace. Finally, the update frequency was an important feature to consider. The need to access up-to-date information integrating evidence-based knowledge and innovative and yet-to-be-proven knowledge was frequently mentioned during interviews.

> *"We do not have the time to go use [a Web 2.0 platform], because that changes all the time."* [participant I1]
>
> *"There are sites that you consult and you return to see them, and realize that it's been a year and they have not been updated."* [participant I6]

All respondents thought facilitators should oversee the knowledge management process to increase the credibility of a Web 2.0 platform. Almost one-third of participants said that the "look and feel" would have an impact on their use of the platform. Lastly, according to some respondents, membership fees would constitute a barrier.

For most respondents, an efficient information technology (IT) department was already in place in their work environment. The relevance of having additional support for a Web 2.0 platform did not come out in the interviews.

According to respondents' self-assessment of their technology skills, most professionals working in remote areas had good computer skills. Active and Inactive members said they had limited skills.

Some participants mentioned they had a pleasant previous experience with a blog or a forum related to their job. Participant A3 claimed becoming over-addicted to this type of tool in the past, and was forced to stop using it. Participant I3 claimed to have a growing interest in forums or blogs, but was still underusing Web 2.0 platforms. Participant I4 occasionally used a discussion forum. Some other respondents (participants R2, R4, and R5) used these types of platforms outside their professional life. Others did not report any previous experience.

Perceived Usefulness

Participants identified three main uses of Web 2.0. The primary use mentioned by the majority of respondents was knowledge transfer. According to them, through interactions and discussions, Web 2.0 platforms offer opportunities to learn and remain up-to-date. It is also a means to facilitate information gathering.

Respondents perceived that young health professionals and professionals working in remote areas would be most likely to benefit from knowledge exchanges that would take place on the Web 2.0 platform. Some participants (participants I3, I8, I9, and R6) mentioned that Web 2.0 applications would be more useful for patients than professionals.

For several health professionals, a Web 2.0 platform would offer an opportunity to link research evidence to clinical practice.

> *"It could be grouped by things that are...with evidence-based...and those looking for the clinical aspect...we make sure to be on the right track than to use trial or error."* [participant I4]

A second use of a Web 2.0 platform was linked to its capacity to increase quality of care. Participants felt that knowledge transfer and exchanges via the Web 2.0 platform would eventually increase the quality of work through behavior changes related to service delivery and care standardization.

However, concerns were expressed about avoiding duplication of resources.

> *"It would be good if the website gathers the information...instead of spreading it."* [participant R1]

The third use of the platform was to allow professionals to accomplish tasks more efficiently. A Web 2.0 platform could reduce time spent searching for information by expanding their sources of information through a larger network of colleagues. It might also be useful for health care providers because it pools efforts made in other workplaces. Respondents also felt the Web 2.0 platform could reduce travel time for meetings.

> *"I can have the same question as someone else, and it's already there. It saves a search that is regularly done. I find that is practical."* [participant A1]
>
> *"We may lose less time to build things, but just adapt them to our setting."* [participant I3]

Elements describing the participants' resistance to the introduction of Web 2.0 into their practices cannot be ignored. This resistance was noted more frequently among managers and late-career professionals (eg, participants A6, I7, and R6). For example, a manager was afraid that employees would use that kind of tool for entertainment instead of work. A lack of organizational recognition was also identified as a barrier to the use of the platform.

> *"It is not just the time to treat patients, we also need time to...for other things. It is less recognized in my organization. It's really, really focused on how many patients you have seen."* [participant I6]

Some participants considered that other means, such as face-to-face discussions or email, are sufficient for their knowledge transfer needs.

> *"If I have a question, I will use my email and will send it to someone."* [participant I8]

Finally, another description of perceived uselessness referred to the inequities between academic and non-academic organizations. Here is how a professional from a non-academic organization expresses her view: "Budget is not the same, things are not the same, so it's not necessarily easy and you often feel incompetent when you compare your practices" (participant I1). This participant negatively perceived Web 2.0 exchanges with professionals working in academic organizations.

Perceived Ease of Use

To encourage people to use a Web 2.0 platform, it has to be as user-friendly as possible: "If it is easy to use, we will use it more" (participant R2). Providing clear and well-organized information was seen as a key factor to limit learning required to use the platform and to save time.

> *"That things are well organized and we do not get lost in finding information, it is also good to encourage us to consult more often."* [participant A2]

Being able to access the information at any time, in any place, was seen as a major advantage of Web 2.0 platforms over face-to-face or phone exchanges. To meet respondents' expectations, pages on a Web 2.0 platform must download quickly, and answers to questions should be made available in a timely fashion.

> *"If I have a problem, I have to quickly have access to get a fast answer. Because I will not go back three times."* [participant I7]

Time Availability

Almost all respondents mentioned they lacked time to read about scientific evidence and to use tools such as Web 2.0 platforms. Professionals explained that these tasks were not necessarily valued in workplaces. Individuals who did not mention this barrier were managers or clinical coordinators (participants A4 and R5).

Behavioral Intention to Use

The vast majority of participants expressed their intention to use the platform. Utilization would be gradual and would vary depending on members' needs (participants I1, I2, I5, R3, and R4).

> *"The better it goes, the more we will return, and it becomes somewhat automatic."* [participant R4]

A realistic frequency of use was defined as once per week. Respondents expressed an "intention to use" ranging from 2 to 15 minutes per visit. Professionals expressed a low tendency to interact actively on the platform. They were more likely to access or view information than to contribute or add new information. This phenomenon was also expressed by a respondent who already uses Web 2.0 in her practice. Some participants (participants I8, R1, and R4) expressed little or no intention to use the Web 2.0 platform. According to participant I8 (33 years), nobody has the time to use a Web 2.0 platform as part of his or her job. Participants R1 (28 years) and R4 (46 years) do not think they will use this type of platform.

Results of Objective 2: Group Differences

Some differences were seen between the 3 groups studied. First, Remote members seemed to have better access to computers. Moreover, this group seemed to have access to a larger network of colleagues outside their organization. Indeed, Active and Inactive members interacted less often with colleagues outside their organization compared with participants working in remote areas. Lastly, the possibility of increasing quality of work was not mentioned by the group working in remote areas, although half of the respondents in the other 2 groups expressed it. No difference concerning available time was observed.

Discussion

The main objective of this study was to document professionals' perception regarding the use of Web 2.0 in their clinical practices. Four out of 5 themes that emerged from our content analysis correspond to the Technology Acceptance Model (TAM) concepts (see Figure 1) [24,25]. Only, the "time availability" theme did not correspond to this model. This conceptual model was developed by Davis in 1980 to better understand why individuals accept or reject technology [24,25]. Although it is not specific to the health care system, it has been widely used to understand IT adoption by health professionals [26]. The TAM has been used with various professionals (eg, nurses [27,28], physical therapists [29], and occupational therapists [30]) using different types of information and communication technologies.

Like Van Schaik [29], our study demonstrates that technology is perceived as a support to evidence-based practice. More specifically, knowledge transfer is identified as the main use of the Web 2.0 platform. In the health care field, this is a major issue [31,32]. According to the Canadian Institutes of Health Research, knowledge transfer is "a dynamic and iterative process that includes the synthesis, dissemination, exchange, and ethically sound application of knowledge to improve the health of Canadians, provide more effective health services and products, and strengthen the health care system" [33]. Knowledge transfer should promote exchanges among patients, health professionals, managers, and researchers. In our study, professionals mainly expressed their need to share clinical experiences among themselves. It is interesting to observe that information searching based on collaborative networks is part of Eysenbach's Medicine 2.0 definition [34]. Eysenbach named this concept "apomediation." But it is important to remember when new knowledge is emerging, clinical experience, or tacit knowledge, is as important as explicit knowledge [35].

For respondents, ease of use of a Web 2.0 platform translates into a platform that does not require prior training (ie, is intuitive). This is important in the current health care system in Quebec where resources dedicated to training are scarce and time availability for integrating new knowledge is limited. As observed by Spallek [36], up-to-date and timely information are foundational elements of an emergent community of practice. The use of a Web 2.0 platform to obtain information in a timely fashion seems especially important to professionals working in remote areas despite their access to a large network of colleagues. Active members, working in urban regions, are probably part of larger teams and their organizations may have more human and financial resources (eg, documentation center, affiliated researchers, and clinical coordinator). This may reflect that professionals still tend to consult their immediate peers when facing complex situations, something that professionals in remote areas cannot do because of limited access to on-site peers.

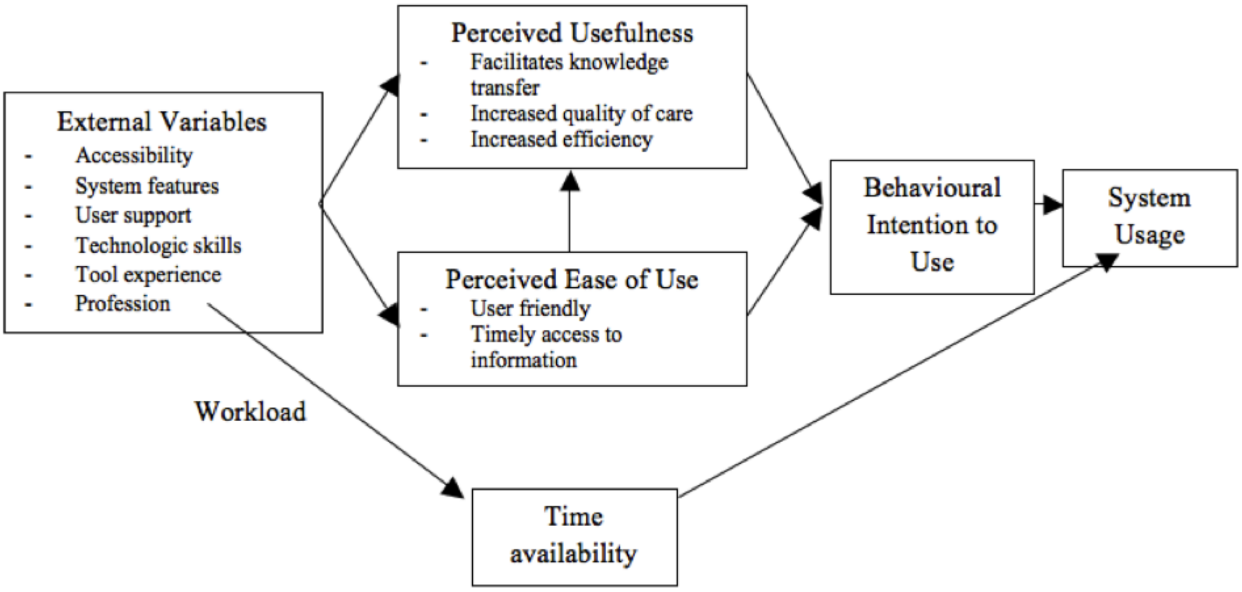
Despite high motivation from respondents to use a Web 2.0 platform that would be easy to use and would offer added value to their practice, several barriers remain. Time was the most often reported barrier. This is consistent with the results reported in other studies that looked at Internet or technology use among various professionals (eg, physicians, nurses, managers, and physical therapists) [14,29,37,38]. This lack of time is also discussed in studies about professionals' involvement in virtual communities of practice (eg, medical imaging administrators and emergency clinicians) [21,22]. Having available time is absolutely essential to participate in a community. This time availability is noted by Wenger et al [39] as being an important element to consider in the early stages of community development. Indeed, time is required before community members can see the added value it gives to their work. This reality is a challenge because respondents expressed their need for rapid answers. This situation reveals a contradiction where professionals will have to make coherent choices. Although they expect to save time by using a Web 2.0 platform to answer their knowledge needs, they will need to invest time to avoid disruption in the flow of information that will be conveyed by the platform. Organizations will need to address this issue to optimize professionals' time and further research will need to be conducted to assess the utilization of a Web 2.0 platform given the short amount of time professionals are willing to spend on such knowledge transfer tools.

During content analysis, classification of external variables influencing perceived usefulness and ease of use was particularly challenging. These variables cover a large spectrum of fields and characteristics with no clear pattern in the selection of external variables across studies [26]. Despite this, several studies have mentioned the low competency of health professionals in using tools available on the Web (eg, databases, virtual communities of practice, and other online information) [14,22,37,40]. Our study supports this observation in that a significant proportion (37.5%) of respondents estimated they had low technological skill. Information technology training should be part of university programs and continuing education sessions for health professionals to enhance their professional skills and encourage behavior changes. According to cyberpsychology, technological skills are correlated with age [41,42]. This relationship has not been identified in our results because the participants who had no intention of using a Web 2.0 platform were between 33 and 46 years. Moreover, no relationship was identified between expressed intention to use and technological skills.

Although it is not shown in our adaptation of the TAM, some respondents perceived a Web 2.0 platform as having little or no use for knowledge transfer. They did not perceive the need to introduce a new knowledge transfer strategy illustrating the existence of other means of knowledge transfer, whereas others reported lack of support from the organization, putting more emphasis on the number of patients treated, two concepts defined by Paré [43] as "vision clarity" and "organizational flexibility."

When comparing the 3 subgroups, very little difference is observed. Professionals working in remote areas may represent a group who would easily accept the introduction of a Web 2.0 platform into their practices because they seem to have better access to computers. However, according to these results, they were less likely to perceive this new knowledge transfer strategy as a useful one because of already well-established networks outside their workplace. Moreover, the fact that members of the Montreal Stroke Network and people within the periphery of the Montreal Stroke Network interact less with colleagues outside their organization could possibly explain why only 2 participants from the Montreal Stroke Network mentioned that it is important that the Web 2.0 platform meet their information needs at the right time, whereas this need was reported by almost all respondents from the other 2 groups.

Figure 1. Modified Technology Acceptance Model (TAM) reflecting perceived factors influencing Web 2.0 adoption by health professionals.



Limitations

Our study has some limitations. First, all respondents were women. However, according to the 2006 Canadian Census [44], 4 out of 5 health workers are women (80%). The fact that only 4 participants worked strictly with stroke patients could be seen as a limitation since the future Web 2.0 platform is intended to be focused on stroke care. Again, our study population is representative of the current organization of care, where health professionals must often deal with multiple clienteles [45]. Another limitation refers to empirical saturation. Despite the number of interviews (n = 24), empirical saturation in each group was probably not reached. We used a purposive sampling approach to recruit our participants when participants should usually be chosen based on the intensity of their experience of the phenomenon under the phenomenological approach. Respecting this condition would have limited our capacity to recruit since Web 2.0 applications are still emerging in the health care field. To compensate for this situation, we asked respondents about their level of skills with technology and we had a variety at each level. The interview as a data collection method possibly created a social desirability bias related to the expressed intention to use the Web 2.0 platform. Further studies are needed to take into consideration the organization and system levels in addition to the individual level as recommended by Karsh [46]. Lau [47] already demonstrated that health policy makers have to be involved in promotion of Web 2.0 utilization. Lastly, it is important to be aware that in qualitative research results may be transferable, but are not necessarily generalizable.

Conclusion

In this study, we aimed to understand professionals' perceptions and needs regarding the introduction of a future Web 2.0 platform into their practices. Previous studies have shown that a positive attitude is often associated with a high level of technology acceptance and adoption [48,49]. Our results reveal that professionals consider Web 2.0 to be very useful for knowledge transfer. However, lack of time and lack of technological skills are limitations to their future use of this technology. The introduction of Web 2.0 platforms undoubtedly requires a change in work habits. Professionals still seem to be inclined to use general search engines (eg, Google) to meet their information needs, whereas sites more specific to their profession could allow them to access more relevant information. Eventually, it might be interesting to investigate patients' perceptions of Web 2.0 platforms technology because this technology may be of interest to them (as stated by some of the respondents). It also might be interesting to explore health professional openness to introduce Web-based exchanges with stroke patients such as Nordqvist [50] addressed it for health professionals and diabetes patients.

Acknowledgments

This research was supported by a grant from the Canadian Institutes of Health Research. The funding agency had no involvement with this paper. Lise Poissant and Annie Rochette have a salary award by FRSQ.

Conflicts of Interest

None declared.

References

1. Glasgow RE, Emmons KM. How can we increase translation of research into practice? Types of evidence needed. Annu Rev Public Health 2007;28:413-433. [doi: 10.1146/annurev.publhealth.28.021406.144145] [Medline: 17150029]
2. Lemieux-Charles L, McGuire W, Blidner I. Building interorganizational knowledge for evidence-based health system change. Health Care Manage Rev 2002;27(3):48-59. [Medline: 12146783]
3. Salbach NM, Jaglal SB, Korner-Bitensky N, Rappolt S, Davis D. Practitioner and organizational barriers to evidence-based practice of physical therapists for people with stroke. Phys Ther 2007 Oct;87(10):1284-1303 [FREE Full text] [doi: 10.2522/ptj.20070040] [Medline: 17684088]
4. Salbach NM, Guilcher SJ, Jaglal SB, Davis DA. Determinants of research use in clinical decision making among physical therapists providing services post-stroke: a cross-sectional study. Implement Sci 2010;5:77 [FREE Full text] [doi: 10.1186/1748-5908-5-77] [Medline: 20946678]
5. Dumoulin C, Korner-Bitensky N, Tannenbaum C. Urinary incontinence after stroke: identification, assessment, and intervention by rehabilitation professionals in Canada. Stroke 2007 Oct;38(10):2745-2751 [FREE Full text] [doi: 10.1161/STROKEAHA.107.486035] [Medline: 17823380]
6. Korner-Bitensky N, Desrosiers J, Rochette A. A national survey of occupational therapists' practices related to participation post-stroke. J Rehabil Med 2008 Apr;40(4):291-297 [FREE Full text] [doi: 10.2340/16501977-0167] [Medline: 18382825]
7. Rochette A, Korner-Bitensky N, Desrosiers J. Actual vs best practice for families post-stroke according to three rehabilitation disciplines. J Rehabil Med 2007 Sep;39(7):513-519 [FREE Full text] [doi: 10.2340/16501977-0082] [Medline: 17724549]
8. Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. Health Technol Assess 2004 Feb;8(6):iii-iv, 1 [FREE Full text] [Medline: 14960256]

9. World Stroke Organization Stroke Guideline Sub-Committee. Stroke Society of Australasia. 2009 Jun. Clinical Practice Guideline Development Handbook for Stroke Care URL: http://www.strokesociety.com.au/index.php?option=com_docman&task=doc_download&gid=185&Itemid=196 [accessed 2011-11-07]
10. Canadian Stroke Strategy. Canadian Best Practice Recommendations for Stroke Care (update 2010). 2010. URL: http://www.hsf.sk.ca/siss/documents/2010_BP_ENG.pdf [accessed 2011-11-07] [WebCite Cache ID 6ASpUqQxI]
11. Poissant L, Ahmed S, Riopelle RJ, Rochette A, Lefebvre H, Radcliffe-Branch D. Synergizing expectation and execution for stroke communities of practice innovations. Implement Sci 2010;5:44 [FREE Full text] [doi: 10.1186/1748-5908-5-44] [Medline: 20529305]
12. Canadian Stroke Network. 2010. Stroke Engine URL: http://strokengine.ca/ [accessed 2011-11-20] [WebCite Cache ID 63LFYyt8Q]
13. Teasell R. EBRSR: Evidence-Based Review of Stroke Rehabilitation. 2011. URL: http://www.ebrsr.com/ [accessed 2011-11-07] [WebCite Cache ID 6ASpbO7DW]
14. Doney L, Barlow H, West J. Use of libraries and electronic information resources by primary care staff: outcomes from a survey. Health Info Libr J 2005 Sep;22(3):182-188. [doi: 10.1111/j.1471-1842.2005.00561.x] [Medline: 16109146]
15. Glasgow RE, Goldstein MG, Ockene JK, Pronk NP. Translating what we have learned into practice. Principles and hypotheses for interventions addressing multiple behaviors in primary care. Am J Prev Med 2004 Aug;27(2 Suppl):88-101. [doi: 10.1016/j.amepre.2004.04.019] [Medline: 15275677]
16. Van De Belt TH, Engelen LJ, Berben SA, Schoonhoven L. Definition of Health 2.0 and Medicine 2.0: a systematic review. J Med Internet Res 2010;12(2):e18 [FREE Full text] [doi: 10.2196/jmir.1350] [Medline: 20542857]
17. Mathieu J. Blogs, podcasts, and wikis: the new names in information dissemination. J Am Diet Assoc 2007 Apr;107(4):553-555. [doi: 10.1016/j.jada.2007.02.027] [Medline: 17383254]
18. Boulos MN, Maramba I, Wheeler S. Wikis, blogs and podcasts: a new generation of Web-based tools for virtual collaborative clinical practice and education. BMC Med Educ 2006;6:41 [FREE Full text] [doi: 10.1186/1472-6920-6-41] [Medline: 16911779]
19. McLean R, Richards BH, Wardman JI. The effect of Web 2.0 on the future of medical practice and education: Darwikinian evolution or folksonomic revolution? Med J Aust 2007 Aug 6;187(3):174-177. [Medline: 17680746]
20. Rethlefsen ML, Rothman DL, Mojon DS. Internet Cool Tools for Physicians. Berlin: Springer; 2009.
21. Nagy P, Kahn CE, Boonn W, Siddiqui K, Meenan C, Knight N, et al. Building virtual communities of practice. J Am Coll Radiol 2006 Sep;3(9):716-720. [doi: 10.1016/j.jacr.2006.06.005] [Medline: 17412156]
22. Curran JA, Murphy AL, Abidi SS, Sinclair D, McGrath PJ. Bridging the gap: knowledge seeking and sharing in a virtual community of emergency practice. Eval Health Prof 2009 Sep;32(3):312-325. [doi: 10.1177/0163278709338570] [Medline: 19696084]
23. Giorgi A. De la méthode phénoménologique utilisée comme mode de recherche qualitative en sciences humaines: théorie, pratique et évaluation. In: Poupart J, Deslauriers JP, Groulx LH, Laperrière A, Mayer R, Pires AP, editors. La recherche qualitative. Montréal: Gaëtan Morin; 1997:113-169.
24. Davis FD. A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results. Cambridge, MA: Massachusetts Institute of Technology; 1986.
25. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Management Science 1989;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
26. 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. [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
27. Rawstorne P, Jayasuriya R, Caputi P. Issues in predicting and explaining usage behaviors with the technology acceptance model and the theory of planned behavior when usage is mandatory. 2000 Presented at: Proceedings of the twenty first international conference on Information systems; 2000; Brisbane p. 35-44.
28. Tung FC, Chang SC, Chou CM. An extension of trust and TAM model with IDT in the adoption of the electronic logistics information system in HIS in the medical industry. Int J Med Inform 2008 May;77(5):324-335. [doi: 10.1016/j.ijmedinf.2007.06.006] [Medline: 17644029]
29. Van Schaik P, Bettany-Saltikov JA, Warren JG. Clinical acceptance of a low-cost portable system for postural assessment. In: Behaviour & Information Technology. Abingdon, Oxford: Taylor and Francis; 2002:47-57.
30. Schaper LK, Pervan GP. ICT and OTs: a model of information and communication technology acceptance and utilisation by occupational therapists. Int J Med Inform 2007 Jun;76 Suppl 1:S212-S221. [doi: 10.1016/j.ijmedinf.2006.05.028] [Medline: 16828335]
31. Oborn E, Barrett M, Racko G. Cambridge Judge Business School. 2010 May. Knowledge Translation in Healthcare: A Review of the Literature (Working Paper Series) URL: http://www.jbs.cam.ac.uk/research/working_papers/2010/wp1005.pdf [accessed 2011-11-08] [WebCite Cache ID 6ASsfF2Mi]
32. Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? J Contin Educ Health Prof 2006;26(1):13-24. [doi: 10.1002/chp.47] [Medline: 16557505]
33. Tetroe J. Knowledge Translation at the Canadian Institutes of Health Research: A Primer. Focus: Technical brief 2007;18:1-8 [FREE Full text]

34. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
35. Nonaka I, Reinmoeller P, Senoo D. Integrated IT systems to capitalize on market knowledge. In: von Krogh G, editor. Knowledge Creation: A Source of Value. New York: Palgrave; 2000:36-60.
36. Spallek H, Butler BS, Schleyer TK, Weiss PM, Wang X, Thyvalikakath TP, et al. Supporting emerging disciplines with e-communities: needs and benefits. J Med Internet Res 2008;10(2):e19 [FREE Full text] [doi: 10.2196/jmir.971] [Medline: 18653443]
37. Younger P. Internet-based information-seeking behaviour amongst doctors and nurses: a short review of the literature. Health Info Libr J 2010 Mar;27(1):2-10. [doi: 10.1111/j.1471-1842.2010.00883.x] [Medline: 20402799]
38. Yarbrough AK, Smith TB. Technology acceptance among physicians: a new take on TAM. Med Care Res Rev 2007 Dec;64(6):650-672. [doi: 10.1177/1077558707305942] [Medline: 17717378]
39. Wenger E, McDermott R, Snyder WM. Cultivating Communities of Practice: A Guide to Managing Knowledge. Boston: Harvard Business School Press; 2002.
40. Kloda LA, Bartlett JC. Clinical information behavior of rehabilitation therapists: a review of the research on occupational therapists, physical therapists, and speech-language pathologists. J Med Libr Assoc 2009 Jul;97(3):194-202. [doi: 10.3163/1536-5050.97.3.008] [Medline: 19626145]
41. Prensky M. Digital Natives, Digital Immigrants. On the Horizon 2001;9(5):1-6.
42. Chiang I, Huang C. Characterizing Web users' degree of Web 2.0-ness. Journal of the American Society for Information Science and Technology 2009;60(7):1349-1357. [doi: 10.1002/asi.21058]
43. Paré G, Sicotte C, Poba-Nzaou P, Balouzakis G. Clinicians' perceptions of organizational readiness for change in the context of clinical information system projects: insights from two cross-sectional surveys. Implement Sci 2011;6:15 [FREE Full text] [doi: 10.1186/1748-5908-6-15] [Medline: 21356080]
44. Statistics Canada. Census of Population. 2006 URL: http://www12.statcan.gc.ca/census-recensement/2006/dp-pd/index-eng.cfm [accessed 2011-11-20] [WebCite Cache ID 1321813382531795]
45. Ministère de la Santé et des Services sociaux. L'architecture des services de santé et des services sociaux: Les programmes-services et les programmes-soutien. 2004 Jan. URL: http://publications.msss.gouv.qc.ca/acrobat/f/documentation/2004/04-710-01.pdf [accessed 2012-09-05] [WebCite Cache ID 6AStilCjf]
46. Karsh BT, Escoto KH, Beasley JW, Holden RJ. Toward a theoretical approach to medical error reporting system research and design. Appl Ergon 2006 May;37(3):283-295. [doi: 10.1016/j.apergo.2005.07.003] [Medline: 16182233]
47. Lau AS. Hospital-based nurses' perceptions of the adoption of Web 2.0 tools for knowledge sharing, learning, social interaction and the production of collective intelligence. J Med Internet Res 2011;13(4):e92 [FREE Full text] [doi: 10.2196/jmir.1398] [Medline: 22079851]
48. Paré G, Sicotte C, Jacques H. The effects of creating psychological ownership on physicians' acceptance of clinical information systems. J Am Med Inform Assoc 2006;13(2):197-205 [FREE Full text] [doi: 10.1197/jamia.M1930] [Medline: 16357351]
49. Alquraini H, Alhashem AM, Shah MA, Chowdhury RI. Factors influencing nurses' attitudes towards the use of computerized health information systems in Kuwaiti hospitals. J Adv Nurs 2007 Feb;57(4):375-381. [doi: 10.1111/j.1365-2648.2007.04113.x] [Medline: 17291201]
50. Nordqvist C, Hanberger L, Timpka T, Nordfeldt S. Health professionals' attitudes towards using a Web 2.0 portal for child and adolescent diabetes care: qualitative study. J Med Internet Res 2009;11(2):e12 [FREE Full text] [doi: 10.2196/jmir.1152] [Medline: 19403464]

Abbreviations

EPOC: Cochrane Effective Practice and Organisation of Care
QDA: qualitative data analysis
TAM: Technology Acceptance Model

Edited by G Eysenbach; submitted 24.11.11; peer-reviewed by S Davis; comments to author 17.01.12; revised version received 20.02.12; accepted 30.05.12; published 13.09.12.

Please cite as:
David I, Poissant L, Rochette A
Clinicians’ Expectations of Web 2.0 as a Mechanism for Knowledge Transfer of Stroke Best Practices
J Med Internet Res 2012;14(5):e121
URL: http://www.jmir.org/2012/5/e121/
doi:10.2196/jmir.2016
PMID:23195753

Original Paper

Understanding the Factors That Influence the Adoption and Meaningful Use of Social Media by Physicians to Share Medical Information

Brian S McGowan[1], PhD; Molly Wasko[2], PhD; Bryan Steven Vartabedian[3], MD; Robert S Miller[4], MD, FACP; Desirae D Freiherr[5], MS; Maziar Abdolrasulnia[5], PhD

[1]Education Technology Consultant, Blue Bell, PA, United States

[2]Department of Management, Information Systems and Quantitative Methods, University of Alabama at Birmingham, Birmingham, AL, United States

[3]Baylor College of Medicine, Baylor College, Houston, TX, United States

[4]Breast Cancer Program, The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins Green Spring Station, Lutherville, MD, United States

[5]CE Outcomes, LLC, Birmingham, AL, United States

Corresponding Author:
Brian S McGowan, PhD
Education Technology Consultant
225 Canterbury Ct
Blue Bell, PA, 19047
United States
Phone: 1 267 603 2510
Fax: 1 205 259 1501
Email: drbrianmcgowan@gmail.com

Abstract

Background: Within the medical community there is persistent debate as to whether the information available through social media is trustworthy and valid, and whether physicians are ready to adopt these technologies and ultimately embrace them as a format for professional development and lifelong learning.

Objective: To identify how physicians are using social media to share and exchange medical information with other physicians, and to identify the factors that influence physicians' use of social media as a component of their lifelong learning and continuing professional development.

Methods: We developed a survey instrument based on the Technology Acceptance Model, hypothesizing that technology usage is best predicted by a physician's attitudes toward the technology, perceptions about the technology's usefulness and ease of use, and individual factors such as personal innovativeness. The survey was distributed via email to a random sample of 1695 practicing oncologists and primary care physicians in the United States in March 2011. Responses from 485 physicians were analyzed (response rate 28.61%).

Results: Overall, 117 of 485 (24.1%) of respondents used social media daily or many times daily to scan or explore medical information, whereas 69 of 485 (14.2%) contributed new information via social media on a daily basis. On a weekly basis or more, 296 of 485 (61.0%) scanned and 223 of 485 (46.0%) contributed. In terms of attitudes toward the use of social media, 279 of 485 respondents (57.5%) perceived social media to be beneficial, engaging, and a good way to get current, high-quality information. In terms of usefulness, 281 of 485 (57.9%) of respondents stated that social media enabled them to care for patients more effectively, and 291 of 485 (60.0%) stated it improved the quality of patient care they delivered. The main factors influencing a physician's usage of social media to share medical knowledge with other physicians were perceived ease of use and usefulness. Respondents who had positive attitudes toward the use of social media were more likely to use social media and to share medical information with other physicians through social media. Neither age nor gender had a significant impact on adoption or usage of social media.

Conclusions: Based on the results of this study, the use of social media applications may be seen as an efficient and effective method for physicians to keep up-to-date and to share newly acquired medical knowledge with other physicians within the medical community and to improve the quality of patient care. Future studies are needed to examine the impact of the meaningful use of social media on physicians' knowledge, attitudes, skills, and behaviors in practice.

(J Med Internet Res 2012;14(5):e117) doi:10.2196/jmir.2138

KEYWORDS

Social media; continuing medical education; physicians and social media; physician-physician relationship; oncologists; primary care physicians; education technology; physicians' practice patterns

Introduction

The amount of information required for medical practice is growing at an exponential rate, and the ability for one physician to stay completely abreast of the entirety of this knowledge base has long since been surpassed [1]. Physicians in primary care fields and data-intensive specialties such as oncology bear a particularly heavy burden in consuming and managing the amount of information available to them [1-3]. Over the next decade, the cognitive limitation of the traditional model, wherein physicians are expected to learn, retain, and call upon an ever-expanding body of medical knowledge, will become more challenging to navigate. New models for learning and sharing will be needed.

Social learning theory has long been applied to medical education [2,4,5]. In the past, these explorations focused on simple connections derived from training pedigree, geography, and shared memberships in medical societies or associations, and connectedness was largely episodic (eg, annual meetings, committee work, and listservs). However, with the emergence of social media, the concept of social learning can encompass a myriad of nontraditional connections and uses.

Social media websites and applications are online environments where users contribute, retrieve, and explore content primarily generated by fellow users. As opposed to more traditional forms of information and communication technologies used in health care organizations, the content generated through social media is typically created by users for users, thus allowing knowledge and support to flow more effectively through a professional social network, and allowing answers and support to be more effectively leveraged across a professional social network [6-9].

Despite a growing body of literature highlighting both the promises and controversies associated with social media usage in health care, there are surprisingly few empirical studies on those most affected by it—physicians themselves [10-12]. There is also variation in how social media were selected, perceptions were collected, and usage was examined, thus limiting the value of conclusions [13,14]. For instance, one recent survey of the literature identified 46 unique definitions of social media across 44 articles of Health 2.0 and Medicine 2.0 publications [15]. Additionally, while patients are embracing social media technologies to share information with other patients and health care experts, practicing physicians seem to be more reluctant to move into a new age of collaborative health care [16].

We used a theoretical framework for assessing and predicting the adoption of social media by physicians for the specific use of sharing medical knowledge and lifelong learning, and explored whether adoption differs between two specialties—primary care and oncology—in rapidly changing medical knowledge environments. We developed the research model based on the Technology Acceptance Model (TAM) (Multimedia Appendix 1), hypothesizing that technology usage is best predicted by a physician's attitudes toward the technology, perceptions about the technology's usefulness and ease of use, and individual factors such as personal innovativeness and beliefs.

Methods

Study Design

We conducted a cross-sectional study of physicians in the fields of primary care and oncology who practice in the United States to test the following primary hypothesis: physicians who perceive social media as easy to use and useful and who have positive attitudes toward its use are more likely to share medical knowledge with other physicians through social media. This protocol was approved by the Western Institutional Review Board (Olympia, WA, USA).

Model, Measures, and Data Collection

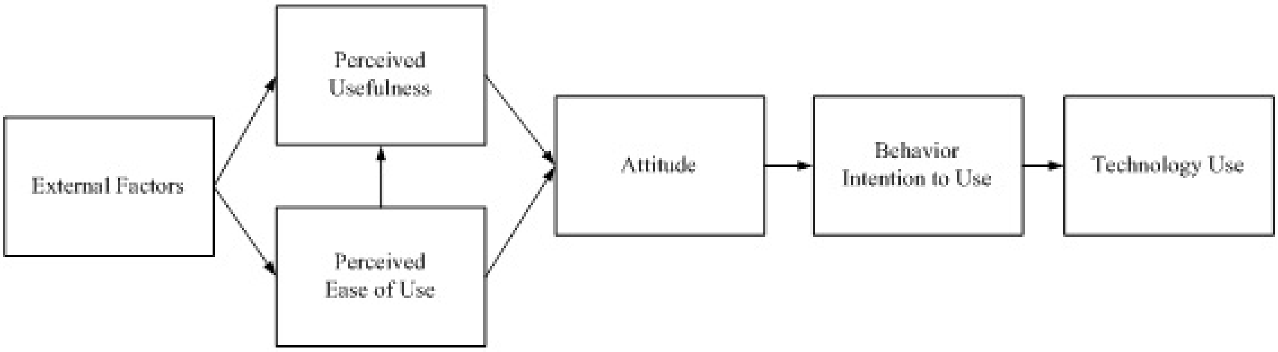
We designed the survey to test the theoretical framework posited in TAM [17]. TAM proposes that an individual's acceptance of a technology is determined by its perceived usefulness and perceived ease of use. The model predicts that ease of use and usefulness will influence an individual's attitudes toward, intention to use, and acceptance of the technology (Figure 1) [17].

Survey questions were adapted from previously published scales. We field tested the survey instrument for clarity and comprehensiveness with 2 physicians in the intended target audience prior to implementation. Using multi-item scales for each construct, the survey assessed (1) the perceived barriers to social media adoption, (2) motivations to adopt social media, including desire to advance the profession, personal innovativeness, and access to peers, (3) attitudes toward social media, (4) perceived ease of use of social media, (5) perceived usefulness of social media, and (6) usage of social media to share medical knowledge with other physicians. The final instrument included 27 items assessing the constructs of interest. Response categories for barriers, motivations, perceived ease of use, and perceived usefulness (independent variables) consisted of a 7-point scale ranging from strongly disagree to strongly agree. Attitudes toward social media usage were assessed using 10-point semantic differential scales. The outcome was the frequency of use of social media to share medical knowledge with other physicians (dependent variable). Response categories for current frequency of use were never, rarely, monthly, once a week, 3 times a week, daily, and many times a day (Multimedia Appendix 1).

For this study, we defined social media as Internet-based applications that allow for the creation and exchange of user-generated content, including services such as social networking, professional online communities, wikis, blogs, and microblogging. We defined use as the exchange of information, advice, ideas, reports, and scientific discoveries with other physicians in the medical community. Additional questions were used to understand adoption on a social media application-specific basis. We did not use these data in the TAM analyses, but to provide a more granular perspective on current levels of use and future intention to use each application.

A national sample of 1695 physicians was randomly selected from the American Medical Association's Physician Masterfile: 699 were practicing in oncology and 996 were practicing in primary care. We sent an email invitation to all physicians in the sample in March 2011 to participate in the survey. An honorarium of US $50 was offered for completing the survey.

Figure 1. The Technology Acceptance Model predicts that ease of use and usefulness will influence an individual's attitudes towards, intention to use, and acceptance of the technology.



Analysis

Prior to performing hypothesis tests, we assessed the reliability and validity of the survey items. To assess reliability, which indicates the degree of agreement between the multiple items making up a construct, we determined the Cronbach alphas for our variables. The average Cronbach alpha was .92, and all constructs were higher than the recommended threshold of .70. Discriminant validity is useful to demonstrate the extent to which a construct of interest differs from others. To assess discriminant validity, we used a principal component factor analysis to test that the various items loaded highest on their theoretical constructs. We ran a 1-factor analysis containing all multi-item constructs using varimax rotation and extracting 8 factors; all items loaded on their expected factors at a level of .69 or higher, indicating adequate discriminant validity.

We used hierarchical regression analysis to test the theoretical model. Hierarchical regression analysis makes it possible to test whether a set of variables, entered as a block and in a theoretically justified order, adds significantly to variance already explained by a prior set of variables. In the first step, we entered demographic variables. In the second step, we entered the individual factors (barriers, motivations to advance the professional community, personal innovativeness, and peer access). In the third step, we entered attitudes toward social media, perceived ease of use, and perceived usefulness. The specified order of entry allows the rigorous testing of the effects of individual factors over and above the demographic variables, as well as the impact of attitudes and perceptions above all the previously entered factors (Multimedia Appendix 2).

To assess the severity of multicolinearity, which occurs when 2 or more predictor variables in a multiple regression are highly correlated, we calculated the variance inflation factor for each variable. All variance inflation factors were below the generally accepted 5.0 cut-off. We also performed the analysis dropping the usefulness variable from the regressions. The amount of variance explained in the final model dropped from .57 to .52, and the pattern of results was the same. We therefore included the usefulness variable in our final model. In reporting the data, we used *P* = .05, 2-sided, as the criterion for statistical significance of the estimated effects of the independent variables on the frequency of social media usage by physicians. All analyses were performed with PASW statistics software, version 18 (IBM Corporation, Somers, NY, USA).

Results

We received responses from 491 of the 1695 physicians contacted, resulting in a response rate of 28.97%. However, 6 respondents self-classified as nonpracticing physicians were removed and a sample of 485 was analyzed (186 oncologists and 299 primary care physicians; Table 1). To assess response bias, we compared the results on key attitudinal questions and demographic variables between early and late responders. There were no significant differences between early and late responders, minimizing the concern of response bias in our sampling frame. To assess nonresponse bias, we compared the demographics of our sampling frame with the overall demographics of primary care physicians and oncologists in the United States and found no discernible differences, minimizing the threat of nonresponse bias.

Table 1. Sample characteristics.

	Oncology (n = 186)	Primary care (n = 299)
Degree (MD/DO), n (%)	186 (100.0%)	299 (100.0%)
Male gender, n (%)	140 (75.3%)	216 (72.2%)
Years since medical school graduation, mean (SD)	24 (10)	24 (9)
Practice location, n (%)		
Urban	88 (47.3%)	70 (23.4%)
Suburban	82 (44.1%)	179 (59.9%)
Rural	17 (9.1%)	50 (16.7%)
Practice setting, n (%)		
Solo	23 (12.4%)	102 (34.1%)
Group	128 (68.8%)	178 (59.5%)
Medical school	17 (9.1%)	3 (1.0%)
Nongovernment hospital	9 (4.8%)	8 (2,7%)
Major professional activity, n (%)		
Direct patient care	181 (97.3%)	292 (97.7%)
Other	5 (2.7%)	7 (2.3%)

The study was designed to understand at a granular level the current adoption and the intent to adopt social media for the exchange of information, advice, ideas, reports, and scientific discoveries with other physicians in the medical community. Figure 2 shows what applications and platforms were being used by respondents. By providing a list of specific applications within the survey, our intent was to underscore the broader definition of social media used within the TAM analyses. Across all applications, awareness was high with 78%–98% of respondents claiming to be aware of the application. Current use varied on an application-specific basis from 33 of 485 (6.8%) for Twitter to 252 of 485 (52.0%) for online physician-only communities (such as Sermo, Ozmosis, or medical society membership sites). For each application there was a subset of respondents (between 5% and 33%) who claimed that they “will never use” the application for the exchange of information, advice, ideas, reports, and scientific discoveries with other physicians in the medical community. But for most applications (except restricted online communities), the largest portion of the respondents identified themselves as currently unlikely or unsure about their intent to use.

Figure 3 shows the frequency distribution of social media usage. Respondents indicated how frequently they (1) were using social media to *contribute* medical knowledge to other physicians, (2) were using social media to *seek* specific information about a medical problem or situation, and (3) were using social media to *scan or explore* medical knowledge for new insights. Overall, 117 of 485 (24.1%) of respondents used social media daily to scan or explore medical information, whereas 69 of 485 (14.2%) contributed new information via social media on a daily basis. These numbers rose to 296 of 485 (61.0%) scanning and 223 of 485 (46%) contributing once a week or more.

Among the variables that TAM explores is the general attitudes that respondents have toward the usefulness of social media for the exchange of information, advice, ideas, reports, and scientific discoveries with other physicians in the medical community. Figure 4 (part A) shows how respondents felt about the use of social media along 3 dimensions: perceived risk, perceived usefulness, and perceived quality of information. Approximately one-third of respondents found social media to be an essential use of time, to be beneficial, and to return high-quality information. Figure 4 (part B) shows how respondents perceived their engagement and use of social media to affect their competency and clinical performance. Approximately 60% of respondents (281 of 485) stated that social media enabled them to care for patients more effectively and improved the quality of patient care they delivered (291 of 485).

Table 2 shows the correlations between constructs and the variance inflation factors. Although the majority of correlations were modest, 2 variables were very strongly correlated with the perceptions of usefulness of the technology: attitudes toward usage (.80) and frequency of usage (.72). These variables indicate that respondents who had strong positive attitudes about using social media and had found using social media to be useful in enhancing their performance and patient care were significantly more likely to be frequent users of social media.

Table 2. Correlation between variables and variance inflation factors (VIFs).

Variable	Specialty	Year	Gender	Patients/week	Barriers	Advance community	Innovativeness	Peer access	Attitudes	Ease of use	Usefulness	VIF
Specialty											1.064	
P value												
Graduation year	.017											1.11
P value	.71											
Gender	.027	–.16										1.093
P value	.55	.001										
Patients per week	.177	.021	–.134									1.066
P value	.000	.65	.003									
Barriers	–.047	.114	.014	.047								1.417
P value	.30	.01	.75	.42								
Advance the professional community	.010	–.056	.018	.004	–.210							1.223
P value	.83	.22	.69	.92	.000							
Personal innovativeness	–.090	–.112	–.124	.056	–.201	.305						1.524
P value	.047	.01	.006	.22	.000	.000						
Peer access	.035	–.141	.052	.023	–.408	.316	.484					2.509
P value	.44	.002	.26	.61	.000	.000	.000					
Attitudes	.078	–.165	.084	.048	–.344	.350	.426	.689				3.109
P value	.09	.000	.07	.29	.000	.000	.000	.000				
Ease of use	–.009	–.274	.038	.048	–.491	.342	.418	.502	.474			1.811
P value	.85	.000	.41	.29	.000	.000	.000	.000	.000			
Usefulness	.043	–.160	.047	.030	–.384	.343	.463	.726	.802	.493		3.426
P value	.35	.000	.30	.51	.000	.000	.000	.000	.000	.000		
Frequency of use	.026	–.121	.075	.064	–.274	.280	.439	.638	.661	.478	.718	
P value	.57	.007	.10	.16	.000	.000	.000	.000	.000	.000	.000	

Table 3 shows results from the hierarchical regression analysis. In the first step that included the demographic variables, the only significant predictor of usage was years since medical school, indicating the physicians who were younger were likely to use social media more frequently; however, the amount of variance explained was less than 2%. The control variable for specialty was not significant, indicating that there were no significant differences in the frequency of usage patterns between oncologists and primary care physicians. In step 2, we added the barriers and individual factors to the model. The amount of variance explained in the frequency of usage increased to 43%, with the variables personal innovativeness and gaining access to influential peers being the key predictors. Age was no longer significant, however, but gender became significant in step 2.

In step 3, the final model, we explained 57% of the variance in frequency of usage. The demographic variables were no longer significant. Barriers, which was not significant in the first 2 steps, became significant but in a positive direction. This is surprising given that barriers to use has a significant negative bivariate correlation with usage frequency, and intuitively we expected that the higher the barriers to usage, the less frequently physicians would use social media. This indicates that once attitudes toward social media and perceptions of its usefulness and ease of use are taken into account, respondents frequently use social media even though the perceived barriers are high. The ability to gain access to influential peers remained significant, indicating that respondents would use social media more frequently because they are motivated by accessing learning and decision-making resources based on the collective knowledge of their peers. Positive attitudes toward social media usage and perceptions about its ease of use and usefulness were also significant predictors of usage frequency.

Although specialty was not significant in the hierarchical regression model, one of the goals of this research was to determine whether there were differences not only in the frequency of social media usage, but also in the predictors explaining usage frequency. These data suggest that oncologists are more likely to be influenced by motivations of personal innovativeness, while primary care physicians are more likely to be influenced by having access to peers (Table 3). Both groups were influenced by positive attitudes toward social media, ease of use, and usefulness.

Table 3. Hierarchical regression results (standardized beta).

Variable	Step 1: Demographics		Step 2: Barriers and motivations		Step 3: Full model		Oncologists		Primary care	
	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value
Specialty	.019	.69	.014	.70	–.003	.92	NA[a]	NA	NA	NA
Graduation year	–.108	.02	–.012	.74	.039	.21	0.076	.16	0.008	.83
Gender	.067	.15	.074	.04	.046	.15	0.058	.28	0.026	.50
Patients per week	.073	.12	.050	.16	.033	.29	0.013	.80	0.040	.29
Barriers			–.013	.72	.083	.02	0.110	.08	0.060	.16
Advance the professional community			.058	.12	–.022	.50	–0.103	.06	0.027	.53
Personal innovativeness			.171	.000	.070	.06	.133	.02	0.006	.90
Peer access			.523	.000	.169	.000	0.061	.43	.254	.000
Attitudes					.153	.004	.178	.04	.143	.04
Ease of use					.154	.000	.247	.000	.105	.04
Usefulness					.407	.000	.412	.000	.384	.000
R2 adjusted	.015	.03	.428	.000	.567	.000	.529	.000	.594	.000
Change in *R2*			0.415	.000	.139	.000				

Figure 2. Respondents' current use and intention to use social media.

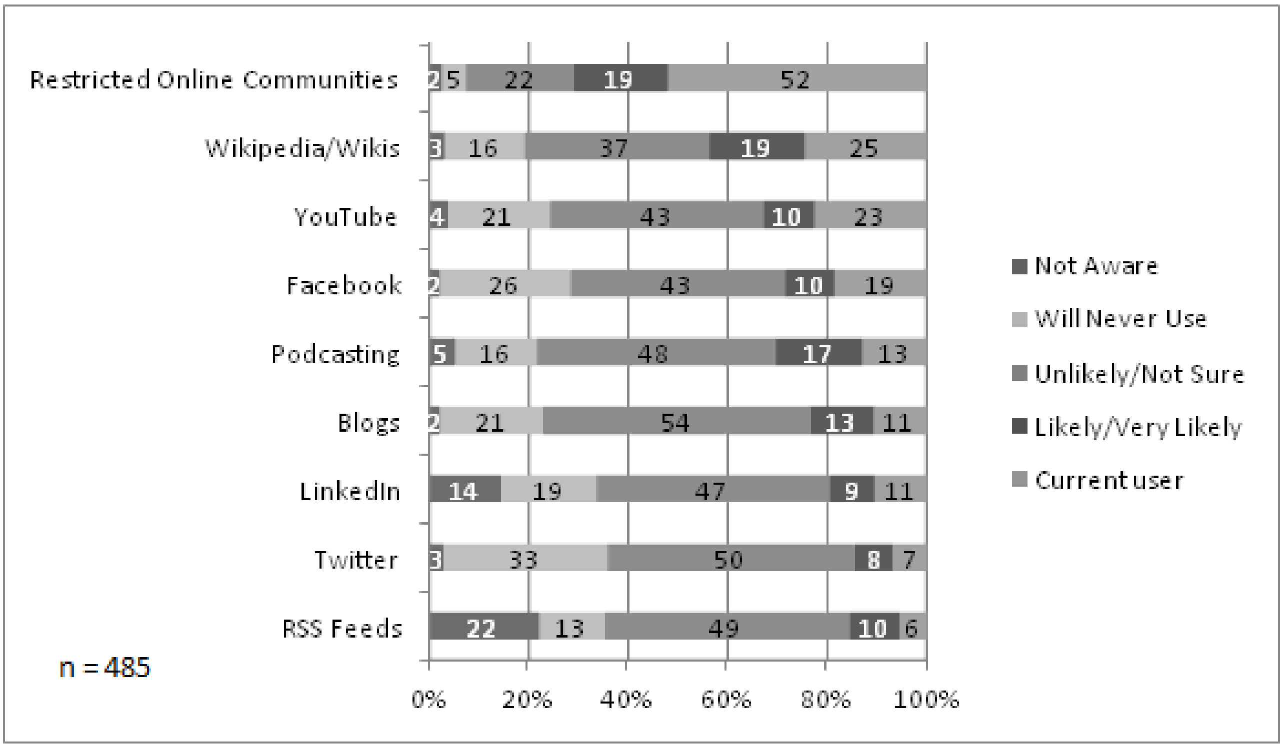


Figure 3. Physicians' frequency of using social media to contribute medical knowledge to other physicians, to seek specific information about a medical problem or situation, and to scan or explore medical knowledge for new insights.

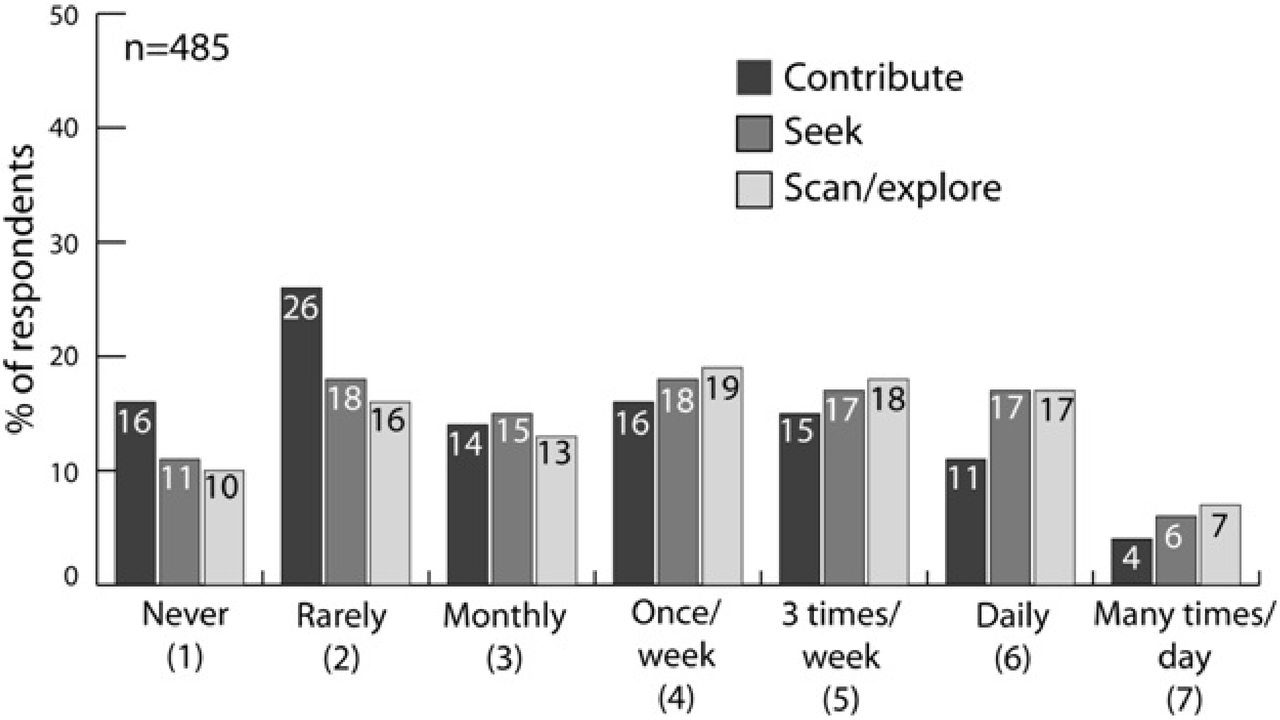
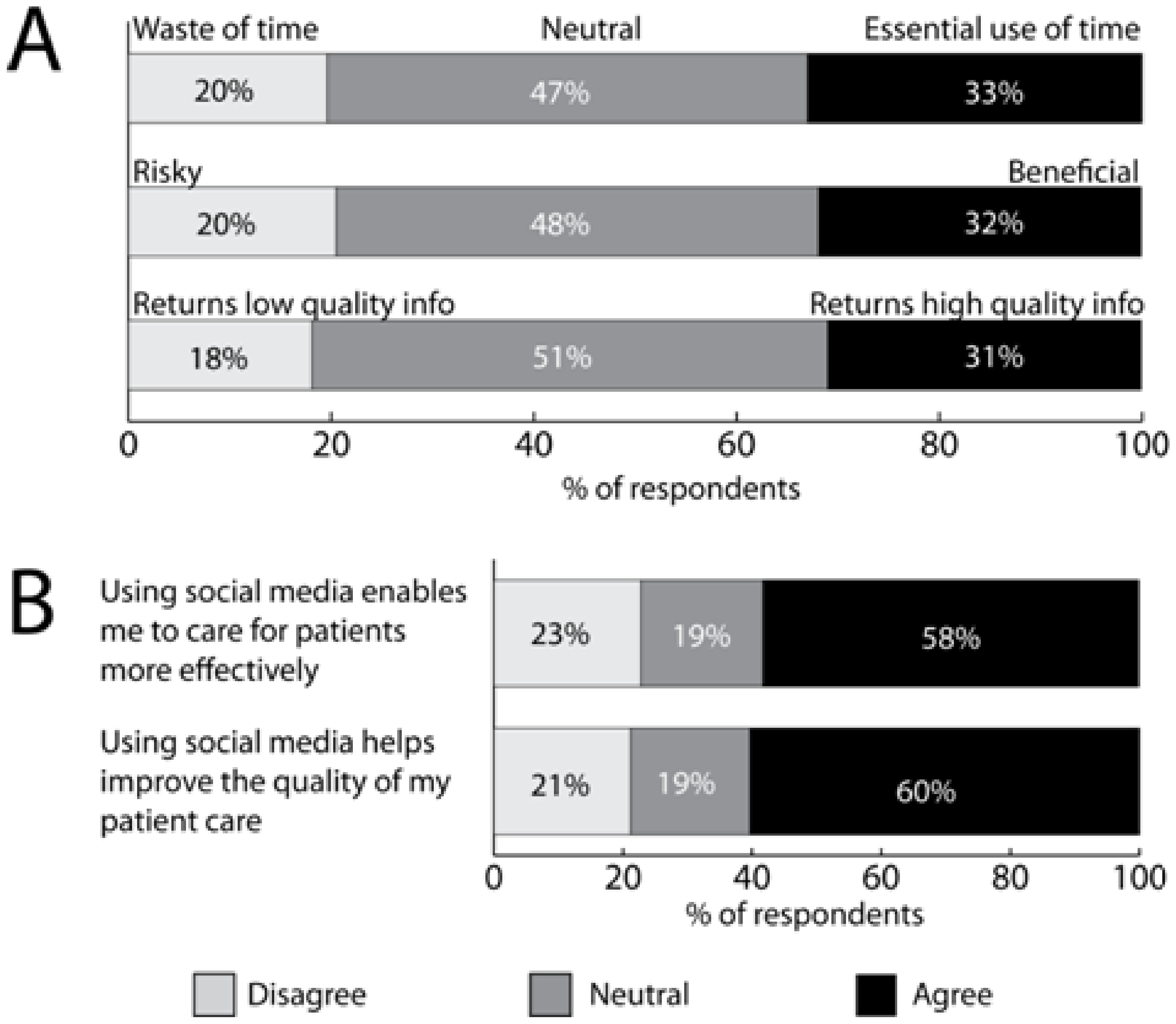


Figure 4. Respondents expressed how they felt about the use of social media along 3 dimensions: perceived risk, perceived usefulness, and perceived quality of information (part A). Part B shows how respondents perceived their engagement and use of social media to affect their competency and clinical performance. n = 485.



Discussion

As the amount of medical knowledge required for patient care continues to expand, social media technologies may provide an efficient and effective tool for educating and informing practicing physicians. Our findings suggest that although a small percentage of respondents were using social media on a daily basis to seek, scan, or contribute medical knowledge with other physicians, up to half of the respondents were taking these actions on a weekly basis. The portion rose to more than 70% of respondents using social media on at least a monthly basis.

From this study sample, it appears that the frequency of social media usage is influenced primarily by positive attitudes toward the technology, perceiving that the technology is easy to use, and perceiving the technology to be useful to achieve better performance outcomes. Conversely, factors found to be nonsignificant included demographic variables typically perceived as important, such as years since graduation (a proxy for age), gender, patients seen per week (a proxy for how busy a physician is), and type of specialty. This finding is consistent with other studies, which have shown practice-related characteristics to be unassociated with use of Internet-based communication technologies [14].

Social media usage was clearly application specific and, once we moved beyond the general definition of social media that was used in the TAM analyses, we found a 6- to 7-fold variation in the extent of adoption across the list of applications we explored. Interestingly, more than half of respondents had adopted online physician-only communities for the purpose of exchanging medical information with other physicians. While future studies should explore the predictors and barriers to adoption at an application-specific level, it stands to reason from these data that the elements of TAM themselves are application specific. For example, respondents might see online communities as a less-risky and higher-quality source of medical knowledge than more broadly open social media applications such as Twitter, LinkedIn, or Facebook.

In this study sample, one key difference between oncologists and primary care physicians was the underlying factors influencing usage. Oncologists were more likely to be motivated to use social media out of a sense of personal innovativeness. This could, in part, be due to a characteristic of the professional culture of oncologists regarding a perceived need to be on the cutting edge of science and clinical practice [18,19]. In contrast, primary physicians were more likely to be motivated to use social media out of a need to have access to and be influenced by peer physicians [20]. In addition to individual factors and attitudes, respondents were far more likely to use social media to share medical knowledge with other physicians when they perceived that learning the technology was easy for them and when it resulted in useful performance outcomes, such as increased practice efficiency and enhanced patient outcomes. Furthermore, though perceived barriers were high, respondents were still willing to use social media more frequently if they have found it to be useful. These findings indicate that, for now, the key factors influencing frequent usage are experiential factors, and they are achieved only after initial adoption of the technology and a period of usage. But it is important to note that there are no definitive studies demonstrating that the use of social media for the exchange of medical information with other physicians as a component of their lifelong learning and continuing professional development leads to more learned physicians or better patient outcomes. For now, personal experience and anecdotes are likely to be the primary drivers of these positive attitudes.

Perhaps, as physicians increasingly experiment with social media technologies, these tools may provide an efficient and effective means for staying abreast of the vast amount of medical knowledge required to deliver patient care. This might be transformative in medicine, as traditional lecture-based continuing medical education has been shown to be largely ineffective in changing physician behavior at the same time that medical knowledge is changing at the fastest pace in history [21-23]. Social media technologies could complement (or even replace) continuing medical education for physicians as either an informal or formal learning channel [24-26]. But for now, how social media channels are the vehicles through which physicians are exposed to emerging information that has the potential to inform or change practice remains an open question [27,28].

As one of the first studies to examine the factors that influence the frequency of social media usage by physicians to share medical knowledge with other physicians, this study has several strengths and limitations. The first strength of this research was establishing clear definitions for our key constructs of interest, specifically definitions for what constitutes social media and clarifying the type of usage of interest (to share medical knowledge with other physicians). The second important strength was the grounding of this research in theory based on TAM, and the use of previously validated, multi-item survey scales to ensure the reliability and validity of the findings. Third, this survey studied two medical specialties characterized by a rapidly changing and dynamic medical knowledge base.

However, limitations of the study include a narrow focus on two medical specialties and one specific definition of usage; a need for a better understanding of the barriers to using social media for lifelong learning; and a need for a better understanding of why respondents indicated they would never use certain channels to exchange medical advice with other physicians. It should be noted that the first limitation is closely related to one of the strengths of this study: the definition of usage was narrowly focused on sharing medical knowledge with other physicians. Therefore, our findings are not generalizable to other types of usage: the use of social media by physicians to treat or to educate patients; nor are they generalizable to physicians' personal use. In addition, there is little prior research directing our exploration of the barriers to use as they uniquely relate to the exchange of medical information with other physicians as a component of their lifelong learning and continuing professional development. Therefore, this study relied on more general barriers, including risks related to privacy, access to social media applications in practice, and time available to use and explore these technologies. Our predictive analyses probably would have yielded even stronger results had we had a more robust understanding of use-specific barriers.

Future studies should examine potential differences between other populations of physicians and other types of health care professionals (specifically, rural and urban; and emergency professions and public health professionals) in terms of their use of social media to share and exchange medical knowledge. Studies should also examine different types of social media usage beyond the sharing of knowledge with other physicians.

Conclusions

The amount of information that a practicing clinician must learn, understand, and apply in practice is growing at unprecedented levels and has long surpassed our cognitive capacities. Social media and social learning models in general provide an important opportunity to manage this information overload, but only if the media are being used effectively. This study demonstrates that the adoption of social media to exchange information and medical knowledge with other physicians is strongly dependent on the perceived usefulness of the technology and the general attitudes physicians have toward the value these technologies offer. Efforts should be made to further explore these predictors of use. These follow-up studies must be conducted with rigor and must move the science of professional learning and development forward in discernible steps to allow physicians to fully embrace a collaborative approach to care.

Acknowledgments

Funding was provided by the Pfizer, Inc. New York, NY, USA. Research was conducted independently from the funding organization.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Construct definitions.

[PDF File (Adobe PDF File), 43KB - jmir_v14i5e117_app1.pdf]

Multimedia Appendix 2

Survey Instrument.

[PDF File (Adobe PDF File), 110KB - jmir_v14i5e117_app2.pdf]

References

1. Abernethy AP, Etheredge LM, Ganz PA, Wallace P, German RR, Neti C, et al. Rapid-learning system for cancer care. J Clin Oncol 2010 Sep 20;28(27):4268-4274. [doi: 10.1200/JCO.2010.28.5478] [Medline: 20585094]
2. Olsen L, Aisner D. The Learning Healthcare System: Workshop Summary. Washington, DC: National Academies Press; 2007.
3. Orzano AJ, Ohman-Strickland PA, Patel M. What can family medicine practices do to facilitate knowledge management? Health Care Manage Rev 2008;33(3):216-224. [doi: 10.1097/01.HMR.0000324909.49766.de] [Medline: 18580301]
4. Kaufman DM, Mann KV. Teaching and learning in medical education: how theory can inform practice. In: Swanwick T, editor. Swanwick T, editor. Understanding Medical Education: Evidence, Theory and Practice. Hoboken, NJ: Wiley-Blackwell; 2010:16-36.
5. Braddock CH, Eckstrom E, Haidet P. The “new revolution” in medical education: fostering professionalism and patient-centered communication in the contemporary environment. J Gen Intern Med 2004 May;19(5 Pt 2):610-611. [doi: 10.1111/j.1525-1497.2004.45003.x] [Medline: 15109334]
6. Within3 News & Events. Within3. 2011 Apr 12. American College of Surgeons Launches Collaborative Online Community to Connect Rural Surgeons around the Country URL: https://within3.com/press-release/american-college-of-surgeons-launches-collaborative-online-community-to-connect-rural-surgeons-around-the-country.html [accessed 2012-04-16] [WebCite Cache ID 66yUVanQr]
7. Chretien KC, Azar J, Kind T. Physicians on Twitter. JAMA 2011 Feb 9;305(6):566-568. [doi: 10.1001/jama.2011.68] [Medline: 21304081]
8. Ohio State Medical Association. 2010. Social Networking and the Medical Practice: Guidelines for Physicians, Office Staff and Patients URL: http://www.osma.org/files/documents/tools-and-resources/running-a-practice/social-media-policy.pdf [accessed 2012-04-16] [WebCite Cache ID 66yUftRJY]
9. Peregrin T. Time to tweet: social networking for surgeons. Bull Am Coll Surg 2011 Feb;96(2):46-48. [Medline: 21452633]
10. Lau AS. Hospital-based nurses' perceptions of the adoption of Web 2.0 tools for knowledge sharing, learning, social interaction and the production of collective intelligence. J Med Internet Res 2011;13(4):e92 [FREE Full text] [doi: 10.2196/jmir.1398] [Medline: 22079851]
11. Archambault PM, Bilodeau A, Gagnon MP, Aubin K, Lavoie A, Lapointe J, et al. Health care professionals' beliefs about using wiki-based reminders to promote best practices in trauma care. J Med Internet Res 2012;14(2):e49 [FREE Full text] [doi: 10.2196/jmir.1983] [Medline: 22515985]
12. Wiecha J, Heyden R, Sternthal E, Merialdi M. Learning in a virtual world: experience with using second life for medical education. J Med Internet Res 2010;12(1):e1 [FREE Full text] [doi: 10.2196/jmir.1337] [Medline: 20097652]
13. Bosslet GT, Torke AM, Hickman SE, Terry CL, Helft PR. The patient-doctor relationship and online social networks: results of a national survey. J Gen Intern Med 2011 Oct;26(10):1168-1174. [doi: 10.1007/s11606-011-1761-2] [Medline: 21706268]
14. Cooper CP, Gelb CA, Rim SH, Hawkins NA, Rodriguez JL, Polonec L. Physicians who use social media and other internet-based communication technologies. J Am Med Inform Assoc 2012 May 25. [doi: 10.1136/amiajnl-2011-000628] [Medline: 22634078]
15. Van De Belt TH, Engelen LJ, Berben SA, Schoonhoven L. Definition of Health 2.0 and Medicine 2.0: a systematic review. J Med Internet Res 2010;12(2):e18 [FREE Full text] [doi: 10.2196/jmir.1350] [Medline: 20542857]
16. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
17. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. Manag Inf Syst Q 1989;13:319-340.

18. Goss TF, Picard EH, Tarab A. Boston Healthcare Associates. 2012 Jun. Recognizing Value in Oncology Innovation URL: http://www.bostonhealthcare.com/objects/pdfs/BHAWhitepaperFINAL060412.pdf [accessed 2012-07-02] [WebCite Cache ID 68rjEO2wS]
19. Association of Community Cancer Centers. 2012. 2012 ACCC Innovator Awards URL: http://accc-cancer.org/meetings/NOC2012-innovatorawards.asp [accessed 2012-07-02] [WebCite Cache ID 68rjbPECB]
20. Nutting PA, Miller WL, Crabtree BF, Jaen CR, Stewart EE, Stange KC. Initial lessons from the first national demonstration project on practice transformation to a patient-centered medical home. Ann Fam Med 2009;7(3):254-260 [FREE Full text] [doi: 10.1370/afm.1002] [Medline: 19433844]
21. Davis D, O'Brien MA, Freemantle N, Wolf FM, Mazmanian P, Taylor-Vaisey A. Impact of formal continuing medical education: do conferences, workshops, rounds, and other traditional continuing education activities change physician behavior or health care outcomes? JAMA 1999 Sep 1;282(9):867-874. [Medline: 10478694]
22. Davis DA, Thomson MA, Oxman AD, Haynes RB. Changing physician performance. A systematic review of the effect of continuing medical education strategies. JAMA 1995 Sep 6;274(9):700-705. [Medline: 7650822]
23. Davis DA, Thomson MA, Oxman AD, Haynes RB. Evidence for the effectiveness of CME. A review of 50 randomized controlled trials. JAMA 1992 Sep 2;268(9):1111-1117. [Medline: 1501333]
24. Cain J, Policastri A. Using Facebook as an informal learning environment. Am J Pharm Educ 2011 Dec 15;75(10):207. [doi: 10.5688/ajpe7510207] [Medline: 22345726]
25. Yamout SZ, Glick ZA, Lind DS, Monson RA, Glick PL. Using social media to enhance surgeon and patient education and communication. Bull Am Coll Surg 2011 Jul;96(7):7-15. [Medline: 22315896]
26. Paton C, Bamidis PD, Eysenbach G, Hansen M, Cabrer M. Experience in the Use of Social Media in Medical and Health Education. Contribution of the IMIA Social Media Working Group. Yearb Med Inform 2011;6(1):21-29. [Medline: 21938320]
27. Wright A, Bates DW, Middleton B, Hongsermeier T, Kashyap V, Thomas SM, et al. Creating and sharing clinical decision support content with Web 2.0: Issues and examples. J Biomed Inform 2009 Apr;42(2):334-346. [doi: 10.1016/j.jbi.2008.09.003] [Medline: 18935982]
28. Hughes B, Joshi I, Wareham J. Health 2.0 and Medicine 2.0: tensions and controversies in the field. J Med Internet Res 2008;10(3):e23 [FREE Full text] [doi: 10.2196/jmir.1056] [Medline: 18682374]

Abbreviations

TAM: Technology Acceptance Model

Edited by G Eysenbach; submitted 16.04.12; peer-reviewed by T Klose, A Fuad; comments to author 18.06.12; revised version received 17.07.12; accepted 14.07.12; published 24.09.12.

Please cite as:
McGowan BS, Wasko M, Vartabedian BS, Miller RS, Freiherr DD, Abdolrasulnia M
Understanding the Factors That Influence the Adoption and Meaningful Use of Social Media by Physicians to Share Medical Information
J Med Internet Res 2012;14(5):e117
URL: http://www.jmir.org/2012/5/e117/
doi:10.2196/jmir.2138
PMID:23006336

Original Paper

Disordered Eating in a Digital Age: Eating Behaviors, Health, and Quality of Life in Users of Websites With Pro-Eating Disorder Content

Rebecka Peebles[1], MD; Jenny L Wilson[2], MD; Iris F Litt[3], MD; Kristina K Hardy[4], PhD; James D Lock[5], MD, PhD; Julia R Mann[6], BA; Dina LG Borzekowski[7], EdD, EdM

[1]The Craig Dalsimer Division of Adolescent Medicine, Department of Pediatrics, The Children's Hospital of Philadelphia, Perelman School of Medicine at The University of Pennsylvania, Philadelphia, PA, United States

[2]Division of Pediatric Neurology, Department of Pediatrics, The Children's Hospital of Philadelphia, Perelman School of Medicine at The University of Pennsylvania, Philadelphia, PA, United States

[3]Department of Pediatrics, Stanford University, Stanford, CA, United States

[4]Department of Neurology, Children's National Medical Center, Washington, DC, United States

[5]Division of Child and Adolescent Psychiatry, Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Palo Alto, CA, United States

[6]Rice University, Houston, TX, United States

[7]Department of Health, Behavior and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

Corresponding Author:
Rebecka Peebles, MD
The Craig Dalsimer Division of Adolescent Medicine
Department of Pediatrics
The Children's Hospital of Philadelphia, Perelman School of Medicine at The University of Pennsylvania
3401 Civic Center Blvd
11NW Room 19
Philadelphia, PA, 19104
United States
Phone: 1 215 590 6864
Fax: 1 215 590 4708
Email: peeblesr@email.chop.edu

Abstract

Background: Much concern has been raised over pro-eating disorder (pro-ED) website communities, but little quantitative research has been conducted on these websites and their users.

Objective: To examine associations between levels of pro-ED website usage, disordered eating behaviors, and quality of life.

Methods: We conducted a cross-sectional, Internet-based survey of adult pro-ED website users. Main outcomes were Eating Disorder Examination Questionnaire (EDE-Q) and Eating Disorder Quality of Life (EDQOL) scores.

Results: We included responses from 1291 participants; 1254 (97.13%) participants were female. Participants had an average age of 22.0 years and a mean body mass index of 22.1 kg/m^2; 24.83% (296/1192) were underweight; 20.89% (249/1192) were overweight or obese. Over 70% of participants had purged, binged, or used laxatives to control their weight; only 12.91% (163/1263) were in treatment. Mean EDE-Q scores were above the 90th percentile and mean EDQOL scores were in the severely impaired range. When compared with moderate and light usage, heavy pro-ED website usage was associated with higher EDE-Q global (4.89 vs 4.56 for medium and 4.0 for light usage, $P < .001$) and EDQOL total scores (1.64 vs 1.45 for medium and 1.25 for light usage, $P < .001$), and more extreme weight loss behaviors and harmful post-website usage activities. In a multivariate model, the level of pro-ED website usage remained a significant predictor of EDE-Q scores.

Conclusions: Pro-ED website visitors reported many disordered eating behaviors, although few had been treated. Heavy users reported poorer quality of life and more disordered eating behaviors.

(J Med Internet Res 2012;14(5):e148) doi:10.2196/jmir.2023

KEYWORDS

Eating disorder; social network; anorexia nervosa; bulimia nervosa; pro-eating disorder website; pro-anorexia website; pro-bulimia website; pro-ED; pro-ana; pro-mia

Introduction

The Internet offers numerous websites that can affect the health of vulnerable users. Of particular concern are pro-eating disorder (pro-ED) website communities (also called pro-anorexia or pro-ana, and pro-bulimia or pro-mia), where individuals may learn about, discuss, and reinforce disordered eating behaviors [1]. Pro-recovery websites promote discussion more related to fighting an eating disorder, although online content can overlap between pro-recovery and pro-ED communities [2].

Content on pro-ED websites includes "thinspiration" (images or text for the purpose of inspiring weight loss), techniques to assist in weight loss, and interactive forums [2-6]. Some sites promote eating disorders as a lifestyle choice, offering encouragement for extreme dieting and exercise behaviors and assistance in avoiding detection by family and medical providers. Other websites aim to support visitors at various stages of illness, recognizing the dangers of disordered eating and offering content dedicated to treatment and recovery [4]. See Figure 1 for an example of a mock website.

These websites can have deleterious effects on the user [7-10]. Participants exposed to a pro-ED site for 25 minutes were more likely to show negative affect, perceive themselves as heavier, and exercise or think about weight [11]. Pro-ED visitors have also displayed higher body dissatisfaction, restriction, and bulimic activity than controls [12]. A pilot study of 76 adolescents who had been in treatment for an eating disorder found that over a third had visited pro-ED sites. Practically all (96%) of these pro-ED website users reported learning new weight loss or purging techniques from the sites [1].

Great concern has been raised over pro-ED websites; however, relatively little is known about their users. No study has examined a large group of pro-ED website users from a clinical perspective, nor have associations between disordered eating patterns and escalating levels of site visitation been described. The purpose of this study was to examine the demographics, media use patterns, and eating behaviors of pro-ED website visitors, and the degree to which website usage correlates with disease severity and quality of life.

Figure 1. Mock pro-eating disorder website with typical content.

I hope that those who currently have eating disorders can find love and support here. This is meant to be a place of understanding and acceptance.

This is a pro-Ana, pro-ED site made to support people who already have an eating disorder. Please do not come here to criticize or judge. If you have come here in hopes of developing an eating disorder, you should leave right now. Eating disorders are not cool or glamorous. They are not a quick fix. They are not a diet. They are a living, breathing hell. So please, I beg you, don't let this monster into your life.

Methods

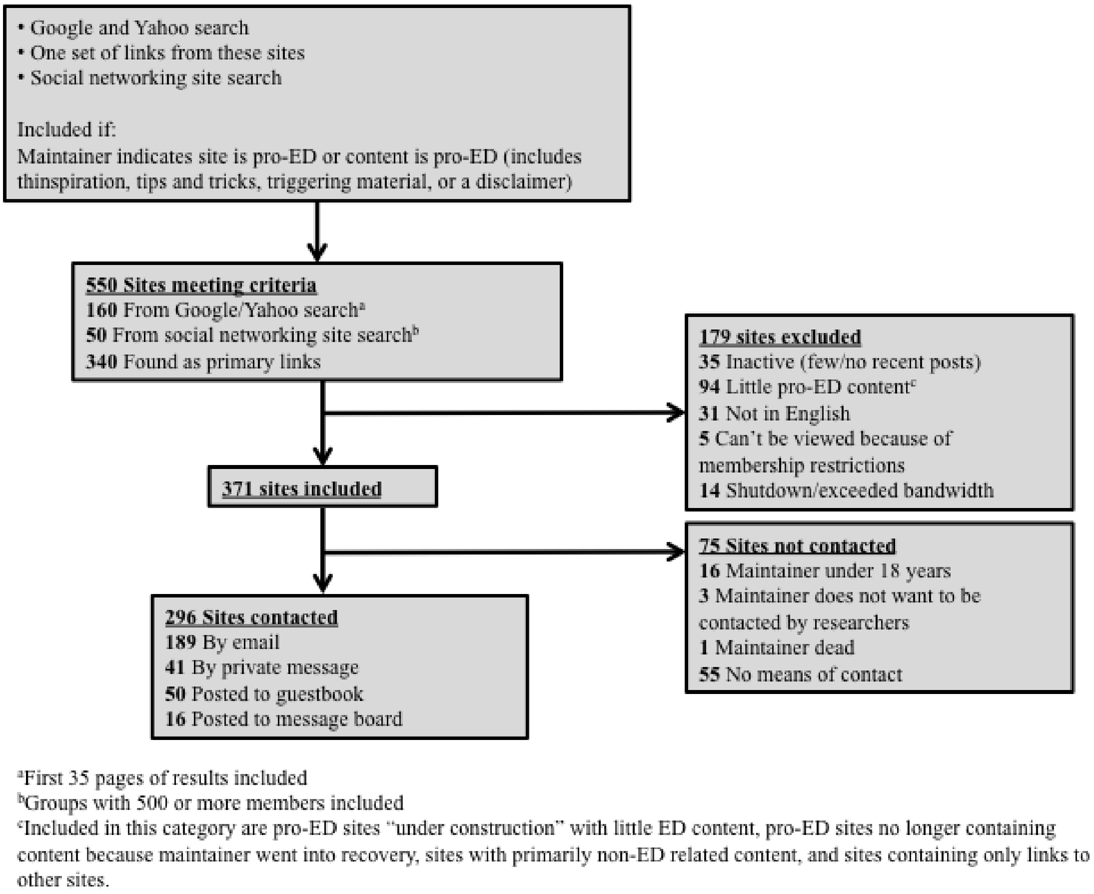
We recruited adult users of websites with pro-ED content via a dedicated link established on identified websites. All protocols were approved by the Stanford University Panel on Medical Research in Human Subjects, Stanford, CA, USA, and were compliant with the Health Insurance Portability and Accountability Act of 1996. Informed consent was obtained from all participants.

Website Search and Inclusion

We developed a comprehensive new methodology to access users from a broad spectrum of websites with pro-ED content. We searched Google and Yahoo!, entering the keywords “pro-ana” and “pro-mia”. All 700 URL addresses on the first 35 pages of search results for either engine were examined. Websites were included if they displayed any of the following: (1) a declaration that the website was pro-ED, (2) a disclaimer or warning to stay away from the site if the visitor was in recovery or did not have an eating disorder, (3) the term thinspiration, or (4) information on disordered eating behaviors in a framework intended to inform the disorder (tips and techniques). We further searched these websites for links to similar sites. These second-generation websites were also examined and included if they met the above criteria. Additionally, we searched the three most popular open social networking sites as determined by comscore Media Metrix [13] for the keywords pro-ana, ana, pro-mia, and mia. Webrings (collections of related websites) or interest groups found in these searches were included if they met the inclusion criteria and had at least 500 members.

Our methods involved no deception. Only English-language sites with active participation or updates within the last 12 months were included. Websites or webrings were excluded if their maintainers indicated that they were less than 18 years of age or did not wish to be involved in research. Figure 2 details the search results.

Figure 2. Website search results. [a] First 35 pages of results included. [b] Groups with 500 or more members included. [c] Included in this category are pro-ED (eating disorder) sites under construction with little ED content, pro-ED sites no longer containing content because maintainer went into recovery, sites with primarily non-ED-related content, and sites containing only links to other sites.



Recruitment

We contacted the maintainers of 296 websites, describing the study and its purpose, and asking maintainers to post a link to our study website, which described both the study and the online survey. Participation time was described as 30–40 minutes. On entering the study site, potential participants were asked to confirm that they were over 18 years of age and to give informed consent before being directed to the survey. No incentives were used to encourage participation. No identifiers were collected and no cookies were used. IP addresses were not tracked or limited in their responses. Participants were able to contact the study researchers via email. Survey responses were entered online and automatically captured into a data file.

Survey

Our online survey contained 193 items, covering demographics, self-reported heights and weights, disordered eating, quality of life, treatment history, overall health, self-harm, and website usage of both pro-ED and pro-recovery websites (see Multimedia Appendix 1). Participants were asked to self-designate their race or ethnicity according to categories defined by the US National Institutes of Health. The Eating Disorder Examination Questionnaire (EDE-Q) assessed the degree of disordered eating and thoughts [14], the core set of the US Centers for Disease Control and Prevention's (CDC) Health-Related Quality of Life Healthy Days Measure [15] was used as a generic quality-of-life measure, and the Eating Disorder Quality of Life (EDQOL) measure [16] assessed eating disorder-specific health-related quality of life. Finally, we adapted a survey that our team had previously used to evaluate pro-ED and pro-recovery website usage [1]. Participants were also asked if they lied on the survey, and if so, these responses were excluded from analyses. The survey was piloted within all members of the research team to be certain it was functional online prior to fielding the questionnaire.

Outcome Measures

Main outcome measures were scores on the EDE-Q and EDQOL. Secondary outcomes were the number of unhealthy days on the Healthy Days Measure and common indicators of health in patients with an eating disorder, such as history of admission to hospitalization, osteopenia, or missed menses. The predictor variable for all primary analyses was the level of pro-ED website usage.

Weight Calculations

We calculated body mass index (BMI) using participants' self-reported heights and weights, using the equation BMI = weight in kilograms/(height in meters)^2. Reported BMI values were divided by a median BMI of 21.7 for women and 23.0 for men, based on growth charts developed by the CDC [17], to obtain percentages of median body weight for each participant. Participants were asked for their highest and lowest weights and ages at those times, and definitions of overweight and obesity at those times followed CDC guidelines for children and adults.

Statistics

We used standard descriptive and frequency testing to describe the data. Associations were assessed using chi-square testing, Student *t* test, Pearson correlations, and analysis of variance, followed by Tukey test for post hoc comparisons. Multivariate regression analyses were used to stratify factors predictive of disordered eating behaviors and impaired quality of life. Responses were not weighted in any manner. Significance level was set at .05. Statistical analysis was conducted using SPSS 19.0 (IBM Corporation, Somers, NY, USA).

Results

Response Rate

While nonresponse bias cannot be assessed in anonymous online surveys, we compared the number of visits (hits) to the study website with the number of completed questionnaires to determine a proxy response rate. Over a 10-week period between May and July 2006, there were 3341 hits to our study website and 1456 completed surveys, resulting in a proxy response rate of 43.58%.

Our final analysis included data from 1291 participants, representing 38.64% of initial survey site hits. A total of 227 participants admitted to lying on the survey (15.59% of respondents). Of these, we excluded 165 from analyses because they lied about their age (n = 127) or reported being less than 18 years of age (n = 38). Of those remaining, 70 reported lying about their weight, so we did not analyze their heights and weights. A total of 18 lied about something other than age and weight, with 13 stating that they had done their best to be truthful but had estimated some answers. The remaining 5 lied on very specific portions of the survey, so we excluded answers to those corresponding questions. We discovered no duplicate responses on analyses. One respondent gave nonsensical responses and was excluded.

With regard to the completeness of responses in crucial data on primary predictor and outcome variables, only 1.63% (n = 21) responses were "break-offs" (surveys with less than 50% of questions answered), and another 3.72% (n = 48) were partially complete (50%-80% of questions answered), according to standard definitions of these values in survey reporting as defined by the American Association for Public Opinion Research [18].

Characteristics of Website Users

Table 1 describes clinical and demographic characteristics reported by participants at the time of the survey, disordered eating behaviors, and views on their own health. Many reported being overweight (282/1188, 23.73%) or obese (406/1188, 34.17%) in the past, with a mean %median BMI at their highest weight of 128.59% (range 70–298, SD 32.71). The mean %median BMI at the lowest reported weight was 88.6% (range 44–233, SD 19.0). While many (800/924, 86.6%) self-diagnosed an eating disorder, 67.62% (873/1291) had never been in treatment, 87.09% (1100/1263) were not currently being treated, and 39.20% (499/1273) had ever had a formal diagnosis of an eating disorder. Underweight participants were more likely to have been treated than normal-weight or overweight respondents (41.6% vs 30.4% or 24.9%, $\chi^2_2 = 18.9$, $P < .001$).

Table 1. Clinical and demographic characteristics of study participants (N = 1291).

Characteristic	n	%	Mean	SD	Range
Age (years)	1260		22	5.1	18–55
Gender (n=1291)					
Female	1254	97.13			
Male	37	2.87			
Ethnicity (n=1291)					
White	1100	85.21			
Hispanic/Latino	63	4.88			
Asian	35	2.71			
African American or black	30	2.32			
American Indian/Alaskan Native	17	1.32			
Native Hawaiian/Pacific Islander	4	0.31			
Other	30	2.32			
Marital status (n=1285)					
Married	134	10.38			
Unmarried	1151	89.16			
Employment/student status (n=1284)					
Student	758	58.71			
Employed	360	27.89			
Combined employed/student	37	2.87			
Unemployed and not a student	129	10.00			
BMI[a] (kg/m^2)	1192		22.1	5.7	12.1–59.0
%Median body weight	1192		101.9	26.4	56–272
Weight class by CDC [b] criteria (n=1192)					
Underweight (BMI <18.5)	296	24.83			
Normal weight (BMI 18.5–24.9)	647	54.28			
Overweight (BMI 25–29.9)	148	12.42			
Obese (BMI ≥30)	101	8.47			
ED [c] diagnosis if formally diagnosed (n=498)					
Anorexia nervosa	225	45.2			
Bulimia nervosa	104	20.9			
EDNOS[d]	151	30.3			
Binge eating disorder	8	1.6			
Did not specify	1	0.2			
Age at dieting onset (years)	1236		13	3.7	
Age at ED onset (years)	1031		14.3	3.7	3–40
Disease duration (years)	1006		7.6	5.6	0–43
Activities in the last month					
Counting of calories, fat, or carbohydrates (n = 1265)	1169	92.41			
Compulsive exercise (n = 1260)	965	76.59			
Secretive eating (n = 1270)	952	74.96			
Purging					

Characteristic	n	%	Mean	SD	Range
Ever (n = 1269)	969	76.36			
Last month (n = 1251)	717	57.31			
Age at onset	964		15.9	3.8	6-40
Binge eating					
Ever (n = 1269)	1068	83.96			
Last month (n = 1271)	843	66.33			
Age at onset	1019		14.5	3.8	3-35
Laxative use					
Ever (n = 1273)	700	54.99			
Last month (n = 1264)	443	35.05			
Age at onset	672		17.6	3.9	10-40
Diet pill use					
Ever (n = 1271)	917	72.15			
Last month (n = 1265)	551	43.66			
Age at onset	909		17.1	3.5	10-48
Used >7 times per week	602	65.9			
Age at first ED treatment (years)	402		17.4	3.9	7.0–38.0
Admission to hospitalization					
In past 30 days (n = 1243)	20	1.61			
Ever (n = 1253)	170	13.57			
Missed menses in last year (n = 1238)					
None	472	38.13			
<3	248	20.03			
≥3	358	28.92			
None, but I am on the pill	148	11.95			
I have never had a period	12	0.97			
Diagnosis of low bone density (n = 1183)	110	9.30			
Mental health diagnoses					
Depression (n = 1271)	748	58.85			
Anxiety (n = 1271)	530	41.70			
ADHD[e] (n = 1264)	136	10.76			
Other (n = 1121)	367	32.74			
Ever treated with psychiatric medication (n = 1259)	648	51.47			
Self-harm					
In last 30 days (n = 1270)	454	35.75			
Ever (n = 1271)	988	77.73			
Cutting (n = 1291)	841	65.14			
Burning (n = 1291)	285	22.08			
Scratching (n = 1291)	558	43.22			
Other (n = 1291)	351	27.18			
Views on own health (n = 1257)					
Eating disordered by choice	389	30.95			

Characteristic	n	%	Mean	SD	Range
Sick	320	25.46			
Recovering or trying to recover	159	12.65			
Healthy	140	11.14			
Other or combination of above	251	19.97			
Support pro-ED movement (n = 1227)					
Completely or very much	539	43.93			
A little bit or somewhat	487	39.69			
Not at all	201	16.38			

[a] Body mass index (kg/m^2).

[b] US Centers for Disease Control and Prevention.

[c] Eating disorder.

[d] Eating disorder not otherwise specified.

[e] Attention deficit/hyperactivity disorder

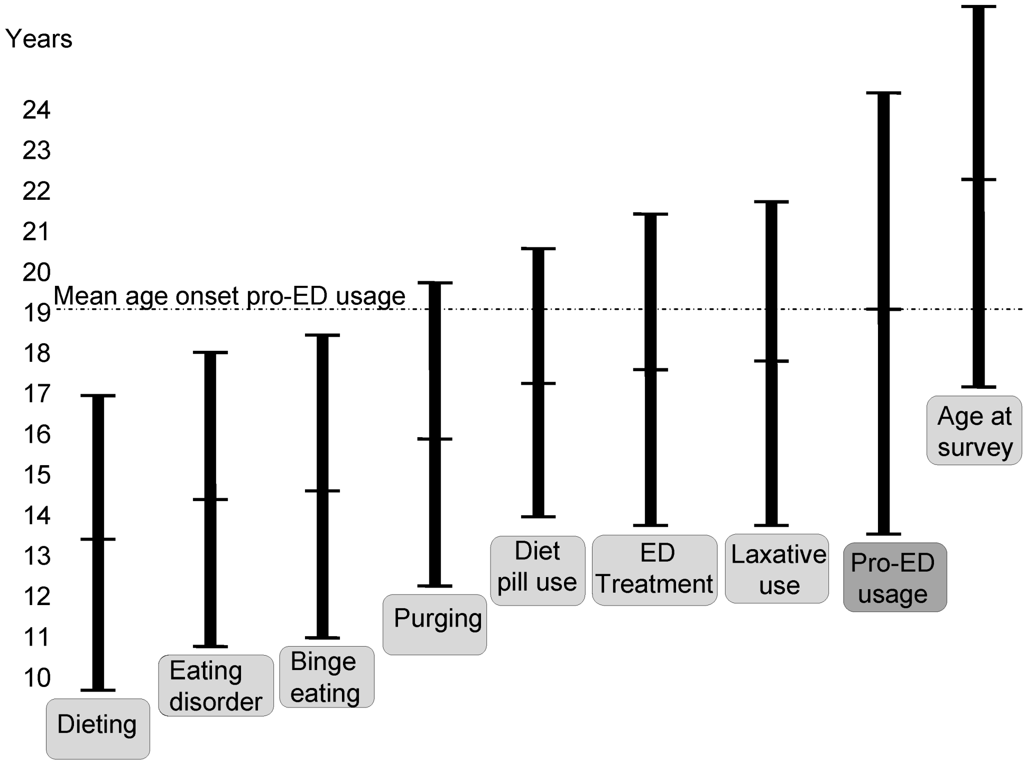
Mean scores on EDE-Q subscale and global scores exceeded the 90th percentile for young adult female norms. Mean scores on subscales were 4.56 (SD 1.25, 95–99 percentile) for Restraint, 3.83 (SD 1.29, 95–99 percentile) for Eating Concern, 5.18 (SD 0.94, 90–95 percentile) for Shape Concern, and 4.86 (SD 1.07, 95–99 percentile) for Weight Concern. The EDE-Q global score showed significant pathology with a mean of 4.61 (SD 0.96, 95–99 percentile).

On the CDC Healthy Days Measure, 5.23% (67/1280) indicated their health was poor, 24.14% (309/1280) fair, 39.61% (507/1280) good, 22.42% (287/1280) very good, and 8.59% (110/1280) excellent. Despite these encouraging answers, participants reported a mean of 6.5 (SD 7.8) physically unhealthy days in the past month, 18.4 mentally unhealthy days (SD 9.9), and 21.1 (SD 9.9) unhealthy days overall. On average, participants reported 7.5 (SD 8.3) days in which their activity was limited in the last month. Moreover, 24.00% (248/1033) reported they had to stop school, and 17.49% (192/1098) needed to stop working in the past because of their eating disorder. As well, 38.06% (427/1122) spent less time on their school or work, 36.10% (422/1169) spent less time in recreational activities, and 58.70% (725/1235) spent less time with friends over the past month, owing to their eating disorder.

EDQOL subscales were all in the severely impaired range, with participants demonstrating a mean psychological score of 2.78 (SD 0.74), physical/cognitive score of 1.97 (SD 0.87), financial score of 0.55 (SD 0.82), and work/school score of 0.63 (SD 0.81). Overall, the global quality-of-life score on this measure was 1.49 (SD 0.60), also in the severely impaired range.

The reported average age at onset of visiting pro-ED websites was higher than the reported age at onset of other disordered eating behaviors and years after participants felt their eating disorder had begun (Figure 3).

Figure 3. Self-reported age of onset of dieting behaviors and pro-eating disorder (ED) website usage. Mean age of onset of ED-related activities indicated by center horizontal bar, with vertical standard deviation error bars. Horizontal dashed line represents mean onset of pro-ED usage.



Website Usage

When asked from which sources they obtained the most eating disorder information in the last 30 days, 86.65% (1077/1243) of participants reported a website, either pro-ana (611/1243, 49.16%), pro-mia (4.42%, 55/1243), pro-recovery (44/1243, 3.54%), or a general health website (27/1243, 2.17%), and the Internet in general (340/1243, 27.35%). Only a few (39/1243, 3.14%) reported consulting books, and only 1.21% (15/1243) reported consulting physicians. The dominance of the Internet for this population was not surprising, given that 35.73% (453/1268) spent more than 4 hours on the Internet daily, 34.94% (443/1268) spent 2–4 hours each day, and 23.11% (293/1268) spent 1–2 hours each day. Only 6.23% (79/1268) spent less than 1 hour on the Internet daily.

Most participants learned about pro-ED sites by chance (403/1228, 32.82%) or by reading about them (308/1228, 25.08%). A third (466/1269, 36.72%) indicated that they had visited a pro-recovery site in the past 30 days, and 62.20% (790/1270) had ever visited such a site. Table 2 presents additional details regarding the online activities and feelings of this population about pro-ED and pro-recovery websites.

Table 2. Online activities and feelings of the study overall population on pro-eating disorder (pro-ED) websites and of the subgroup using pro-recovery websites.

	Pro-ED websites			Pro-recovery websites		
	n	%	Mean (SD)	n	%	Mean (SD)
Reported activities or feelings						
Age first visited (years)	1218		18.9 (5.5)	735		19.3 (4.8)
Hours/week in last 30 days	1147		10.5 (18.2)	681		2.3 (6.8)
Have their own site	212	17.24		42	5.7	
Felt supported by website in last 30 days	n = 1210			n = 692		
Extremely	351	29.01		46	6.6	
Very	385	31.82		66	9.5	
Somewhat	254	20.99		151	21.8	
A little	115	9.50		183	26.4	
Not at all	105	8.68		246	35.5	
Found a community like self on website	n = 1220			n = 705		
Completely	412	33.77		59	8.4	
Very much	335	27.46		78	11.1	
Somewhat	201	16.48		116	16.5	
A little	152	12.46		203	28.8	
Not at all	120	9.84		249	35.3	
Most important reason for accessing site in last 30 days	n = 1211			n = 658		
Motivation for weight loss	447	36.91		89	13.5	
Weight loss tips	192	15.85		54	8.2	
Meeting people	63	5.20		13	2.0	
Tips on hiding eating disorder	48	3.96		32	4.9	
Curiosity	95	7.84		220	33.4	
Support	318	26.26		190	28.9	
Help with recovery	2	0.17		22	3.3	
Activity at sites	n = 1289			n = 790		
Read posts	1120	86.89		534	67.6	
Visited chat room	307	23.82		89	11.3	
Posted messages	696	54.00		167	21.1	
Read diaries or blogs	913	70.83		358	45.3	
Created my own diary or blog	433	33.59		86	10.9	
Learned about weight loss methods	789	61.21		218	27.6	
Learned about diet pills, laxatives, or weight loss supplements	621	48.18		197	24.9	
Learned about places to purchase new diet pills, laxatives, or weight loss supplements	294	22.81		80	10.1	
Post-website use activity				n = 790		
Changed eating habits	671	54.55		137	18.6	
Used new diet pills, laxatives, or weight loss supplements in past 30 days	210	17.21		37	5.2	
Used new diet pills, laxatives, or weight loss supplements ever	454	37.09		128	17.9	

	Pro-ED websites			Pro-recovery websites		
	n	%	Mean (SD)	n	%	Mean (SD)
New weight loss or purging methods in past 30 days	391	32.26		99	13.6	
New weight loss or purging methods ever	748	60.86		224	30.7	
Self-harm in past 30 days	68	5.60		40	5.6	
Self-harm ever	180	14.65		95	12.9	
Site components that were motivational for continuing eating disorder or recovery						
Photographs and artwork	1097	89.70		426	58.8	
Forum postings and chat rooms	1040	85.60		552	77.0	
Diet and exercise information	1086	89.16		499	69.3	
Diaries, journal entries, and blogs	1017	83.50		551	76.3	

Website Usage Level and Outcomes

We grouped participants based on the frequency and duration of pro-ED website usage. Light users (n = 199, 16.1%) were those who reported less than 1 month of website usage regardless of frequency, or 1–12 months of usage with a frequency of less than once a month. Heavy users (n = 513, 41.5%) were those who used the websites at least daily and had visited them for 12 or more months. All other participants were considered medium users (n = 525, 42.4%). Mean hours/week of website usage were 3.02 (SD 4.8), 7.8 (SD 12.8), and 16.1 (SD 23.7, $P < .001$) for light, medium, and heavy users, respectively. Differences between groups in outcomes are presented in Table 3.

Table 3. Pro-eating disorder (pro-ED) website usage and associated eating disorder activities and outcomes.

Characteristic	Total	Light (n = 199)	Medium (n = 525)	Heavy (n = 513)	χ^2/*F*	*df*	*P* value
Age at survey (years), mean	22.0	23.1	22	21.6	6.1[a]	2	.008
Current %median body weight, mean	101.9	106.3	101.4	100.5	3.5[b]	2	.006
Eating Disorder Examination Questionnaire score, mean							
Restraint	4.56	3.86	4.55	4.86	50.4[c]	2	≤.001
Eating concerns	3.83	3.24	3.71	4.17	43.8[c]	2	≤.001
Shape concerns	5.18	4.65	5.15	5.4	49.1[c]	2	≤.001
Weight concerns	4.86	4.27	5.82	5.13	51.6[c]	2	≤.001
Global	4.61	4	4.56	4.89	69.2[c]	2	≤.001
Healthy days (in last 30 days)							
Overall health poor (n = 1280)	67 (5.23%)	4.5	3.8	7.2	29.2[c]	8	≤.001
Overall health excellent (n = 1280)	110 (8.59%)	14.1	9	6.4	29.2[c]	8	≤.001
Unhealthy physical days, mean	6.5	5.3	5.8	7.7	11.2[c]	2	≤.001
Unhealthy mental days, mean	18.4	16.6	17.6	20	11.2[c]	2	≤.001
Sum unhealthy days, mean	21.1	19.1	20.4	22.6	11.2[c]	2	≤.001
Days of limited activities, mean	7.5	7	7.1	8.3	3.0[d]	2	.05
Eating Disorder Quality of Life score, mean							
Psychological	2.78	2.43	2.74	2.95	39.0[c]	2	≤.001
Physical	1.97	1.60	1.92	2.17	33.7[c]	2	≤.001
Financial	0.55	0.47	0.49	0.64	5.3[a]	2	.005
Work/school	0.63	0.43	0.59	0.76	13.1[c]	2	≤.001
Total	1.50	1.25	1.45	1.64	34.7[c]	2	≤.001
Amenorrhea (n = 1082)	367 (33.92%)	25.1	28.7	42.4	29.2[c]	4	≤.001
Low bone mineral density (n = 1183)	110 (9.30%)	8.9	8.7	10.1	.574	2	.750
Age started dieting (years), mean	13.3	14.3	13.5	12.6	16.7[c]	2	≤.001
Age eating disorder began (years), mean	14.3	14.7	14.6	13.9	4.2[b]	2	.015
Disease duration (years), mean	7.6	8.1	7.5	7.6	.564	2	.569
History of eating disorder treatment (n = 1291)	419 (32.46%)	23.6	28.0	39.2	22.5[c]	2	≤.001
History of hospitalization (n = 1253)	170 (13.57%)	6.6	12.3	17.3	15.1[c]	2	.001
Age first binged (years), mean	14.5	14.7	14.5	14.5	.238	2	.788
Binge eating							
Last 30 days (n = 1271)	843 (66.33%)	66	66.5	65.8	.051	2	.975
Ever (n = 1272)	1068 (83.96%)	80.2	81.1	88.3	12.4[a]	2	.002
Age first purged (years), mean	15.9	16.4	16.2	15.6	4.5[b]	2	.012
Purging							
Last 30 days (n = 1251)	717 (57.31%)	48.5	53.2	65.1	22.3[c]	2	≤.001
Ever (n = 1269)	969 (76.36%)	66.3	72.8	84.4	32.9[c]	2	≤.001

Characteristic	Total	Light (n = 199)	Medium (n = 525)	Heavy (n = 513)	χ^2/F	*df*	*P* value
Age first used laxatives (years), mean	17.6	18.4	17.9	17.2	4.3[b]	2	.014
Laxative use							
Last 30 days (n = 1264)	443 (35.05%)	28.6	32.8	39.8	9.8[a]	2	.007
Ever (n = 1273)	700 (54.99%)	45.5	51.5	62.8	22.3[c]	2	≤.001
Age first used diet pills (years), mean	17.1	18.2	17.3	16.7	8.4[c]	2	≤.001
Diet pill use							
Last 30 days (n = 1270)	551 (43.66%)	22.7	42.1	53.6	55.8[c]	2	≤.001
Ever (n = 1271)	917 (72.15%)	53.5	72.0	79.7	48.9[c]	2	≤.001
Excessive exercise last 30 days (n = 1260)	965 (76.59%)	68.4	74.5	82.3	18.0[c]	2	≤.001
Self-injury							
Last 30 days (n = 1270)	454 (35.75%)	26.3	34.0	41.9	16.9[c]	2	≤.001
Ever (n = 1271)	988 (77.73%)	66.2	76.7	83.0	23.8[c]	2	≤.001
Age first visited a pro-ED site (years), mean	18.9	21.3	19.5	17.5	40.5[c]	2	≤.001
Eating-disordered activity at sites							
Learned about weight loss methods (n = 1289)	789 (61.21%)	61.3	66.5	61.6	3.2	2	.200
Learned about diet pills or laxatives (n = 1289)	621 (48.18%)	35.7	52	53.6	19.8[c]	2	≤.001
Learned about places to purchase new diet pills, laxatives, or weight loss supplements (n = 1289)	294 (22.81%)	15.1	22.5	28.1	14.1[c]	2	.001
Post-website use activity resulting from visiting pro-ED websites							
Eating habits have changed (n = 1230)	671 (54.55%)	40.3	58.2	56.1	19.3[c]	2	≤.001
Used new diet pills, laxatives, or weight loss supplements in past 30 days (n = 1220)	210 (17.21%)	6.2	17.5	20.9	21.8[c]	2	≤.001
Used new diet pills, laxatives, or weight loss supplements ever (n = 1224)	454 (37.09%)	17.9	34.3	46.9	54.0[c]	2	≤.001
Used new weight loss or purging methods in past 30 days (n = 1212)	391 (32.26%)	28.4	34.6	31.3	2.7	2	.253
Used new weight loss or purging methods ever (n = 1229)	748 (60.86%)	45.1	64	63.4	23.6[c]	2	≤.001
Self-harm in past 30 days (n = 1214)	68 (5.60%)	1.6	5.4	7.1	8.1[b]	2	.017
Self-harm ever (n = 1229)	180 (14.65%)	6.6	13	19.1	19.6[c]	2	≤.001
Host own pro-ED website (n = 1230)	212 (17.24%)	3.6	11.9	27.8	76.1[c]	2	≤.001
Completely support pro-ED (n = 1227)	329 (26.81%)	7.3	25	36.1	135.8[c]	8	≤.001

[a] *P* < .01; in analysis of variance (ANOVA) testing: post hoc differences between light and medium and light and heavy.

[b] *P* < .05; in ANOVA testing: post hoc differences between light and heavy for %median body weight and age when first used laxatives to control weight, between medium and heavy for age when eating disorder began; between light and heavy and medium and heavy for age at first purge.

[c] P ≤ .001; in ANOVA testing: post hoc differences between all groups for all Eating Disorder Examination Questionnaire scores, Eating Disorder Quality of Life (EDQOL) psychological, physical, and total scores, age started dieting, and age first used pro-ED sites; between light and medium and light and heavy for age first used diet pills; between heavy and light and heavy and medium for EDQOL financial and work/school subscores, and unhealthy physical, mental, and sum days.

Predictors of Disordered Eating and Quality of Life Impairment

To estimate models predicting EDE-Q and EDQOL global scores, we first examined colinearity among the variables of interest. Age and duration of disease were strongly and significantly correlated ($r = .76$), as were %median body weight and %median BMI at highest ($r = .74$) and lowest ($r = .77$) weights. While other variables were correlated, no others were at the point of colinearity. Therefore, we entered age, %median body weight, level of pro-ED website usage, and EDE-Q and EDQOL global scores into our multivariate models.

Significant predictors of EDE-Q scores were EDQOL global score (beta = .56, $P < .001$) and higher pro-ED usage level (beta = .19, $P < .001$), predicting 40% of the variance ($F_{4,1116} = 185.2$, $P < .001$). Significant predictors of EDQOL scores were EDE-Q scores (beta = .59, $P < .001$) and age (beta = –.07, $P < .005$), while pro-ED usage level was not significant (beta = .04, $P = .13$). This variable set explained 38% of the variance in EDQOL scores ($F_{4,1116} = 167.4$, $P < .001$).

Discussion

This was the largest and most comprehensive study of adult pro-ED website visitors to date. Most are normal-weight young women, who report multiple extreme weight control behaviors, yet have never been in formal eating disorder treatment. This study challenges presumptions about both pro-ED website users and eating disordered individuals in general.

Our results show a clear association between the level of pro-ED website usage and both disordered eating and quality of life. Pro-ED website usage remains an important predictor of EDE-Q scores even when other commonly reported predictors are considered. Heavy users of pro-ED websites differ significantly from light users; of particular concern are those who spend, on average, around 16 hours per week on these websites. More website usage was strongly associated with higher levels of disordered eating on the EDE-Q and more severe impairment on the EDQOL. Usage level was also incrementally associated with younger age at dieting onset, various disordered eating behaviors, and most harmful post-website use activities, such as diet pill use, weight loss techniques, or self-injury. Nearly a third of heavy website users hosted their own pro-ED website and supported pro-ED as a movement. Website usage was also strongly associated with treatment and hospitalization rates. This was a cross-sectional study, so causality cannot be inferred and may simply reflect the degree of illness in heavy website users. However, the reported average age at onset of visiting pro-ED websites was higher than the reported age at onset of other disordered eating behaviors, and years after participants felt their eating disorder had begun. This suggests a disease progression in which website visitation is a later consequence, and not an earlier cause.

These participants differ from the media's typical image of underweight teenaged pro-ED users, as over half of participants had been overweight or obese, and a fifth were overweight or obese at the time of the survey. Participants displayed high levels of eating disorder pathology and impaired quality of life on validated measures, consistent with those previously reported for homogeneous populations of more strictly defined anorexia nervosa patients [14,16,18]. Pro-ED website users self-reported dangerous behaviors including purging, laxative use, compulsive exercise, and diet pill use. They also reported high levels of eating disorder sequelae and comorbidities, with nearly 50% reporting menstrual irregularities in the last year. Psychiatric diagnoses were prevalent, with 59% reporting depression, 42% anxiety, 52% a history of treatment with psychiatric medication, and over 75% with a history of self-injury.

Despite the high level of pathology and the majority self-diagnosing an eating disorder, only one-third of participants had ever received formal care for their disordered eating. These findings suggest inadequate screening and diagnosis of eating disorders, and that pro-ED website users are seeking support online instead of with a traditional health model. This is a complex phenomenon, and it has been suggested that pro-ED websites may provide a safe, nonjudgmental, and possibly therapeutic interactive environment [19]. However, they have also been noted to offer advice on managing eating disorders, and may subvert mainstream medical care systems by failing to portray eating disorders as negative conditions requiring professional treatment [20]. Our finding that treatment was much less common in normal-weight, overweight, or obese participants raises the question of whether disordered eating may be more likely to be missed in normal and overweight populations, and that these individuals in particular may seek online support.

Respondents reported an astonishingly high 21.1 unhealthy days in the past month (a normal value is 6 unhealthy days in the past 30 days, with diabetic people averaging 11, and breast cancer patients averaging 12) [15]. Participants also had severe levels of impairment on a disease-specific measure, the EDQOL. These results also indicate the need for further prospective study on quality of life and disordered eating, as they appear to be strongly associated.

Online survey methods are new and developing. This comprehensive study design successfully obtained representation from a wide spectrum of websites meeting specific inclusion criteria based on pro-ED content. Engaging website designers and maintainers to incorporate dedicated links to our study allowed us to include over 1200 participants over a relatively short study period. We also assessed participant truthfulness, which allowed us to further improve the quality of data analyzed. Finally, studying participants via the Internet not only offers a population more diverse than many previously reported clinical samples, but also circumvents the ethical problem of inadvertently introducing potentially harmful website content to a population previously naïve to pro-ED sites.

This study has some important limitations. First, respondents were queried about activities that began many years prior to the study, rendering the possibility of recall bias. Second, this was a convenience sample of users who chose to participate, potentially resulting in a selection bias. Third, all online surveys are limited by an inability to meet with participants to verify responses, although prior studies have noted that truthfulness is actually increased using online survey tools when discussing sensitive topics [21]. Fourth, our survey design did not involve a method of access control and thus could have had duplicate responses from the same IP address. Finally, because so few respondents had received treatment, it is possible that the participants underreported or had not yet identified some medical outcomes.

These findings highlight the need to consider the Internet more often as a vehicle for intervention and study, as it offers easy accessibility to users. If appropriate and helpful interventions are developed, they have the potential to reach many people beyond traditional treatment center walls. Users of pro-ED content websites visit them not just for motivation for weight loss or specific dieting tips, but also for emotional support. Pro-recovery websites do not resonate as well with these users, indicating that the medical community needs to listen to online health seekers to determine whether there are self-help sites or educational modules that this population would find meaningful.

Websites with pro-ED content may play both supportive and harmful roles for those struggling with disordered eating. While the content of these sites may affect the eating behaviors of website visitors, the extent of website usage appears to have a more central role in eating behaviors, weight concerns, and quality of life. Moreover, these findings confirm that many with significantly disordered eating, medical complications, psychiatric comorbidity, and severely impaired quality of life are not accessing traditional care and do not fit conventional eating disorder models. They seek support from a Web-based peer group, which poses both potential harms and opportunities for interventions within these online communities. It is critical that future studies comprehensively address possibilities for intervention and improved relationships with these forums, in order to advance our treatment and screening procedures into an online age.

Acknowledgments

Access to data: Rebecka Peebles had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. The authors are prepared to provide the data and will cooperate fully in obtaining and providing the data on which the paper is based for examination by the editors or their assignees.

Funding: Rebecka Peebles received funding from the Pediatric Research Fund at Stanford University School of Medicine. Jenny Wilson received funding from the Medical Scholars Research Program through Stanford University School of Medicine for her research. James Lock received funding from the National Institutes of Health K24 mechanism. No sponsor had any role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Previous presentation: The authors certify that the manuscript represents valid work and that neither this manuscript nor one with substantially similar content under their authorship has been published or is being considered for publication elsewhere.

Other contributors: The authors wish to thank all participants and website maintainers who participated in our study. We also wish to thank all research assistants at the Stanford WEIGHT Lab for their assistance in data collection and coding. Finally, we wish to thank Scott Engel, PhD, at the Department of Psychology, North Dakota State University, Fargo, ND, USA, for use of the EDQOL, and Christopher Fairburn, DM, Department of Psychiatry, University of Oxford, England, for use of the EDE-Q.

Authors' Contributions

Rebecka Peebles was involved in all components of this study including conception, design, data acquisition and interpretation, drafting, revising and approval of the manuscript, and the integrity of the work as a whole. Jenny Wilson was involved in the conception, design, data acquisition and interpretation, and drafting, revising, and approval of the manuscript. Dina Borzekowski was involved in the conception and design, analysis and interpretation of data, drafting and critically revising the manuscript, supervision, and approval of the published version. Kristina Hardy was involved in conception and design, analysis and interpretation of data, manuscript revision, and approval of the final version. James Lock was involved in the project conception and design, analysis and interpretation of data, critical revision, supervision, and approval of the final version. Julia Mann was involved in the conception and design, acquisition of data, analysis and interpretation of data, drafting the manuscript, and approval of the final version. Iris Litt was involved in the study conception and design, interpretation and analysis of data, critical revision of the manuscript, supervision, and approval of the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Online survey. Note: The survey was one continuous online page.

[PDF File (Adobe PDF File), 103KB - jmir_v14i5e148_app1.pdf]

References

1. Wilson JL, Peebles R, Hardy KK, Litt IF. Surfing for thinness: a pilot study of pro-eating disorder Web site usage in adolescents with eating disorders. Pediatrics 2006 Dec;118(6):e1635-e1643 [FREE Full text] [doi: 10.1542/peds.2006-1133] [Medline: 17142493]
2. Borzekowski DL, Schenk S, Wilson JL, Peebles R. e-Ana and e-Mia: A content analysis of pro-eating disorder Web sites. Am J Public Health 2010 Aug;100(8):1526-1534. [doi: 10.2105/AJPH.2009.172700] [Medline: 20558807]
3. Harshbarger JL, Ahlers-Schmidt CR, Mayans L, Mayans D, Hawkins JH. Pro-anorexia websites: what a clinician should know. Int J Eat Disord 2009 May;42(4):367-370. [doi: 10.1002/eat.20608] [Medline: 19040264]
4. Borzekowski DL, Bayer AM. Body image and media use among adolescents. Adolesc Med Clin 2005 Jun;16(2):289-313. [doi: 10.1016/j.admecli.2005.02.010] [Medline: 16111619]
5. Norris ML, Boydell KM, Pinhas L, Katzman DK. Ana and the Internet: a review of pro-anorexia websites. Int J Eat Disord 2006 Sep;39(6):443-447. [doi: 10.1002/eat.20305] [Medline: 16721839]
6. Abbate Daga G, Gramaglia C, Pierò A, Fassino S. Eating disorders and the Internet: cure and curse. Eat Weight Disord 2006 Jun;11(2):e68-e71. [Medline: 16809973]
7. Andrist LC. Media images, body dissatisfaction, and disordered eating in adolescent women. MCN Am J Matern Child Nurs 2003;28(2):119-123. [Medline: 12629318]
8. Csipke E, Horne O. Pro-eating disorder websites: users' opinions. Eur Eat Disord Rev 2007 May;15(3):196-206. [doi: 10.1002/erv.789] [Medline: 17676689]
9. Custers K, Van den Bulck J. Viewership of pro-anorexia websites in seventh, ninth and eleventh graders. Eur Eat Disord Rev 2009 May;17(3):214-219. [doi: 10.1002/erv.910] [Medline: 19142974]
10. Jett S, LaPorte DJ, Wanchisn J. Impact of exposure to pro-eating disorder websites on eating behaviour in college women. Eur Eat Disord Rev 2010;18(5):410-416. [doi: 10.1002/erv.1009] [Medline: 20572210]
11. Bardone-Cone AM, Cass KM. What does viewing a pro-anorexia website do? An experimental examination of website exposure and moderating effects. Int J Eat Disord 2007 Sep;40(6):537-548. [doi: 10.1002/eat.20396] [Medline: 17525952]
12. Harper K, Sperry S, Thompson JK. Viewership of pro-eating disorder websites: association with body image and eating disturbances. Int J Eat Disord 2008 Jan;41(1):92-95. [doi: 10.1002/eat.20408] [Medline: 17634964]
13. MacManus R. ReadWriteWeb. 2006 Jan 9. Latest SNS Numbers: MySpace Streaks Ahead URL: http://www.readwriteweb.com/archives/latest_sns_numb.php [accessed 2012-08-23] [WebCite Cache ID 6A8fGJ1en]
14. Fairburn CG, Beglin SJ. Assessment of eating disorders: interview or self-report questionnaire? Int J Eat Disord 1994 Dec;16(4):363-370. [Medline: 7866415]
15. Centers for Disease Control and Prevention. USA.gov. 2011 Mar 15. Health-Related Quality of Life (HRQOL): Key Findings URL: http://www.cdc.gov/hrqol/key_findings.htm [accessed 2011-12-08] [WebCite Cache ID 63msobxLM]
16. Engel SG, Wittrock DA, Crosby RD, Wonderlich SA, Mitchell JE, Kolotkin RL. Development and psychometric validation of an eating disorder-specific health-related quality of life instrument. Int J Eat Disord 2006 Jan;39(1):62-71. [doi: 10.1002/eat.20200] [Medline: 16345055]
17. Centers for Disease Control and Prevention. USA.gov. 2010 Sep 09. Growth Charts URL: http://www.cdc.gov/growthcharts/ [accessed 2011-12-08] [WebCite Cache ID 63mrvqgNQ]
18. Mond JM, Hay PJ, Rodgers B, Owen C, Beumont PJ. Validity of the Eating Disorder Examination Questionnaire (EDE-Q) in screening for eating disorders in community samples. Behav Res Ther 2004 May;42(5):551-567. [doi: 10.1016/S0005-7967(03)00161-X] [Medline: 15033501]
19. Dias K. The Ana Sanctuary: women's pro-anorexia narratives in cyberspace. J Int Womens Stud 2003 Apr;4(2):31-45 [FREE Full text]
20. Fox N, Ward K, O'Rourke A. Pro-anorexia, weight-loss drugs and the internet: an "anti-recovery" explanatory model of anorexia. Sociol Health Illn 2005 Nov;27(7):944-971. [doi: 10.1111/j.1467-9566.2005.00465.x] [Medline: 16313524]
21. Kam LE, Chismar WG. Online self-disclosure: model for the use of internet-based technologies in collecting sensitive health information. Int J Healthc Technol Manag 2006;7(3-4):218-232. [doi: 10.1504/IJHTM.2006.008433]

Abbreviations

BMI: body mass index
CDC: Centers for Disease Control and Prevention
EDE-Q: Eating Disorder Examination Questionnaire
EDQOL: Eating Disorder Quality of Life
pro-ED: pro-eating disorder

Edited by G Eysenbach; submitted 16.06.12; peer-reviewed by L Fernandez-Luque; comments to author 11.07.12; revised version received 13.07.12; accepted 13.07.12; published 25.10.12.

Please cite as:
Peebles R, Wilson JL, Litt IF, Hardy KK, Lock JD, Mann JR, Borzekowski DLG
Disordered Eating in a Digital Age: Eating Behaviors, Health, and Quality of Life in Users of Websites With Pro-Eating Disorder Content
J Med Internet Res 2012;14(5):e148
URL: http://www.jmir.org/2012/5/e148/
doi:10.2196/jmir.2023
PMID:23099628

Original Paper

The Representation of Suicide on the Internet: Implications for Clinicians

Michael Westerlund[1,2], PhD; Gergo Hadlaczky[1]; Danuta Wasserman[1], MD, PhD

[1]National Centre for Suicide Research and Prevention of Mental Ill-Health (NASP), Department of Learning, Informatics, Management and Ethics (LIME), Karolinska Institutet, Stockholm, Sweden

[2]Section for Journalism, Media and Communication (JMK), Department of Media Studies (IMS), Stockholm University, Stockholm, Sweden

Corresponding Author:
Michael Westerlund, PhD
National Centre for Suicide Research and Prevention of Mental Ill-Health (NASP)
Department of Learning, Informatics, Management and Ethics (LIME)
Karolinska Institutet
Granits väg 4
Stockholm, 17177
Sweden
Phone: 46 8 524 870 26
Fax: 46 8 31 11 01
Email: michael.westerlund@ims.su.se

Abstract

Background: Suicide is one of the major causes of death in the world, leading to approximately 1 million deaths per year. While much of what is said about suicide and its causes is still taboo in most contemporary societies and cultures, internet websites and discussion forums have become an important and controversial source of information on the subject. A great deal of ambivalence is discernible as to whether online communication about suicide primarily should be seen as an opportunity or a serious threat.

Objective: To investigate how the subject of suicide is represented on the Internet, based on hits generated by the search engine Google.

Methods: In an exploratory design, Google search results on the target word “suicide”, for the years 2005, 2009, and 2012 respectively, were systematically analyzed and compared.

Results: The study shows that web pages of institutional origin on the subject predominate, that the content provided by these institutions concerns primarily research and prevention, and that the form of communication used by these senders is almost exclusively monological. However, besides these institutional pages there are a substantial number of private senders and pages, often anti-medical and against treatment of depression and other mental problems, characterized by dialogue, confessions and narratives, and to a higher degree, an alternative pro-suicide stance.

Conclusions: To counteract the influence of anti-medical and pro-suicide information, the role of the Internet should be discussed with the patient in clinical practice. Dialogical and confessional communications provide an opportunity for the clinician to gain a deeper perspective into perceptions of patients, regarding both their afflictions and the role of medical treatment in their lives.

(J Med Internet Res 2012;14(5):e122) doi:10.2196/jmir.1979

KEYWORDS

suicide; internet; websites; pro-suicide; suicide-preventive

Introduction

Suicide is one of the major causes of death in the world, leading to approximately 1 million deaths per year. It is estimated that by the year 2020, this figure will have increased to 1.5 million [1,2]. While the topic of suicide is still taboo and stigmatized [3], and also often neglected in most contemporary societies and cultures [4], websites and discussion forums have become an important and sometimes controversial source of information on the subject [5,6]. It estimated that there are more than 2 billion Internet users worldwide, with about half of them living in Asia [7]. The application of the Internet in daily life is progressively enhancing. Socioeconomic gaps in Internet use are decreasing, while Internet access is available in more and more rural and remote areas.

One central aspect of the Internet is that the line between private and public communication is expunged. This also has consequences for the development of how the topic of suicide is handled in society. The Internet is not a pretend-world. It is a world populated by real people with different social and cultural backgrounds. In its virtual social environment people meet, exchange thoughts, feelings and experiences; and just like in the real world, the actions people take on the Internet affect their own and/or other people's lives [8].

A great deal of ambivalence is discernible as to whether online communication about suicide should be seen primarily as an opportunity or a serious threat [9,10]. Some researchers have reflected on the emergence of pro-suicide websites on the Internet [11-14]. It has been pointed out that these sites recommend suicide as a solution to life's problems and contain detailed descriptions of methods yielding the maximum effect, as well as suicide notes and pictures of people who committed suicide. They persuade and exert group pressure to fulfil suicide plans, glorifying those who have committed suicide. These websites have also given rise to a new form of suicide pact—"net suicides" [15-18]. These developments have given grounds for a fear of the Internet's detrimental influence on peoples' understanding and behavior towards suicide. Some authors have put forward that the Internet has a stronger "Werther effect"—ie, a greater potential to induce suicide acts—than other mediated communication forms [19]. Others point to a clear "anti-psychiatric" attitude behind the production of the pro-suicide message [20].

However, the Internet can also be seen as a key resource and a powerful communication tool for understanding and psychologically supporting potentially suicidal individuals [21-24], sometimes referred to as the "Papageno effect" [25]. For example, it may be used to identify people in the risk zone for suicidal acts and then communicate with them, thereby preventing such acts [26].

In April 2005, the word *suicide* generated around 27 million hits on the Google search engine. Four years later, in April 2009, the number of search hits had more than doubled, and in January 2012, the Google search engine generated about 250 million hits on the word *suicide*. The mere existence of so many websites on the suicide subject is not cause for alarm in itself; it means only that the word *suicide* is included in a variety of texts on the Web or that Google's search algorithms are better at finding texts that include the term *suicide*. Without a more detailed study of these websites' content, we cannot know what they stand for.

It has been argued that search engines like Google can be seen as meta-sources for information on various topics and that search results on a certain topic may be representative of Internet content at large: "Google search volumes reflect a large proportion of all available Internet search data on the chosen term, thereby capturing a vast amount of information" [27].

The aim of this quantitative study is to examine how the subject of suicide is represented to users on the Internet by proxy of hits for the word *suicide* generated by the search engine Google.

The specific research questions that the study is designed to answer are:

- Who are the senders of the information about suicide on the Internet?
- How is the subject of suicide communicated? Is it primarily through a monological or a dialogical communication?
- What categories of content on the suicide subject appear on the Internet?
- Which discourse about suicide predominates on the Internet?

Methods

In April 2005, a search was generated on Google (google.com) using the word *suicide.* The first 100 search hits were analyzed systematically. The rationale behind this procedure was to capture a cross-sectional sample of search results. At the heart of Google's search engine system is the software PageRank [28]. PageRank can be described simply as a type of priority scheme in which more "valuable" and frequently visited websites are ranked higher in the search lists than the less visited and less "valuable" pages. Value is defined by terms similar to that used in the academia, ie, topic relevance and number of citations by other publications (in PageRank this corresponds to links by other websites). The first 100 search results for the word *suicide* should therefore be regarded as the most visited and most cited web pages in the world about suicide.

The search procedure and analysis were repeated in April 2009 and in January 2012, in order to investigate possible changes in the representation of suicide on the Internet over time.

One aspect that could be argued to influence the results of this study is changes in search engine algorithms used by Google. Such algorithms are optimized by Google frequently to enhance user experience. Changes in these algorithms may affect the search results in a way that does not reflect a change in tangible suicide-related material on the Internet. However, as the aim of this study is to capture what Internet users find on the web, as opposed to describing the complete range of suicide-related materials on the Internet (including those that cannot be found), it is argued that the representativeness of this sample is not threatened: A change in the algorithms affects the results of this study the same way as it affects the results as experienced by users.

After visiting each website, two researchers filled out a checklist with a number of items relating to four pre-specified characteristics, described below.

Senders: this variable attempted to identify the actor(s) behind the specific text or web page. Authorities associated with a government or a public body were identified, as well as Non-Governmental Organisations (NGOs) and associations, corporations, or private senders.

Form of communication: this refers to the direction of the communication on the given website. Whether users have the opportunity to discuss with each other, comment on/change the contents, or only to receive the contents without interaction.

Content: this refers to the type of information provided on the website, for example: scientific findings versus specific suicide-related events such as the death of a celebrity, or personal “confessions” and discussions about suicide.

Discourse: this variable aimed to capture how the topic of suicide was approached. The current medical-psychiatric view, where suicide is seen as harmful behavior, both for the individual and society, and something that is preventable and treatable was contrasted to views promoting/enticing the act of suicide as valid solution to problems in life, and/or suggesting methods and means to commit suicide.

Exclusion Criteria

Websites in which the word suicide was used metaphorically but that did not otherwise contain any suicide-related content were excluded. Examples of this may be references to members of parliament who have committed “political suicide”, cells in the human body that “commit suicide”. Suicide is a very powerful and symbolically charged word in our culture and is sometimes used in contexts unrelated to the suicidal subject and the act of suicide itself, probably in order to boost the effect of what is expressed [29].

Interestingly, the non-related search hits for the word suicide were especially high for the year 2009 (46%) compared to 2005 and 2012 (13% and 19% respectively) (Table 1). This reflects the volatile nature of Internet-based information content when defined in terms of Google search hits. In this study, the main cause of this increase in non-related search hits in 2009 was firstly an increase in popularity of a pornographic website called *SuicideGirls*, which not even after thorough examination revealed any suicide-related material; and secondly the promotional activities of the electronic punk musical duo called *Suicide* who were releasing a new album as well as starting a tour in 2009. Thus, websites related to both these topics were lacking suicide-related content.

Attrition

As web pages are removed or modified and because web addresses are constantly changing, the “website attrition” that occurred between period of sampling and analysis was noted in our sample. Time is an important factor affecting attrition in studies of the Internet using a cross-sectional methodology [30, 31]. All websites in this study were visited within 2 weeks of the date of sampling. An attempt was made to analyze even those websites that were unavailable at the time of the visit using Internet Archive (www.archive.org). However, due to the exclusion of non-related hits and website attrition, the sample was thus reduced to 84 hits of full material in 2005, to 48 hits in 2009 and to 82 hits in 2012 (Table 1).

Table 1. Included suicide-related content, exclusion due to relevance, and attrition in 2005, 2009 and 2012 respectively. *P*-values based on Fisher’s exact test.

Material	Year			
	2005	2009	2012	*P*-value
	n = 100	n = 100	n = 100	
Included hits	84	48	82	<.01
Exclusion (non-related)	13	46	19	<.01
Exclusion (attrition)	3	6	0	.05

Coding

Fifty randomly selected search hits were coded by two independent coders in order to test inter-rater reliability. The Cohen’s Kappa value of 0.72 was interpreted as a substantial agreement between the coders [32]. The same coders and measurements were used in the 2005, 2009 and 2012 material.

Analysis

Considering that this is not a hypothesis-driven study, but rather an attempt at mapping suicide-related content on the Internet, Fisher’s exact was used to establish an association (or the lack thereof) between time and the study variables regarding suicide-related content. Cochran-Armitage trend analysis was used to identify trends in the changes regarding suicide-related content. No significant trends were identified.

Results

Senders

The 2005 material shows that 65 (77%) of the 84 Google hits on the search term *suicide* come from *institutional senders* and 19 (23%) from *private senders* (Table 2). The frequencies for 2009 and 2012 confirm that *institutional senders* are the largest category. Yet, the 2005, 2009 and 2012 material has a large proportion of *private senders* in the more relevant search positions (the first ten search hits), which indicates that these senders’ websites are relatively visible and highly attended.

Table 2. The share of search hits on Google on the search term suicide distributed on the variable Senders. In percent (and frequencies). *P*-values based on Fisher's exact test for association (and Cochran-Armitage trend statistic in parentheses).

Senders	Year			
	2005	2009	2012	*P*-value
	n = 84	n = 48	n = 82	
Institutional senders	77% (65)	71% (34)	83% (68)	.27 (.41)
Private senders	23% (19)	29% (14)	17% (14)	

Form of Communication

The communication about the subject of suicide on the Internet is clearly dominated by the one-way communication form *monologue*, even though the differences between the categories are somewhat smaller in 2009 and 2012 than in 2005 (Table 3).

Table 3. The share of search hits on Google on the search term suicide distributed on the variable Form of communication. In percent (and frequencies). *P*-values based on Fisher's exact test for association (and Cochran-Armitage trend statistic in parentheses).

Form of communication	Year			
	2005	2009	2012	*P*-value
	n = 84	n = 48	n = 82	
Monologue	93% (78)	88% (42)	85% (70)	.30 (.41)
Dialogue	7% (6)	12% (6)	15% (12)	

Content

Research and suicide-prevention is the most common content in the search hits on suicide in all of the studied years. In the 2005 material, 61 (73%) of the search hits fell into this category; in 2009, it was 23 (48%); and in the 2012 material, 61 (74%) of the search hits were categorized as *research and suicide-prevention* (Table 4).

The relatively small share of personal and intimate statements and conversations about suicide, found in the category *confessions*, has been fairly the same for all the three years (Table 4).

Table 4. The share of search hits on Google on the search term suicide distributed on the variable Content. In percent and (frequencies). In percent (and frequencies). *P*-values based on Fisher's exact test for association (and Cochran-Armitage trend statistic in parentheses).

Content	Year			
	2005	2009	2012	*P*-value
	n = 84	n = 48	n = 82	
Research /suicide prevention	73% (61)	48% (23)	74% (61)	<.01 (.82)
Suicide events	2% (2)	6% (3)	9% (7)	.20 (.08)
General reflections	18% (15)	21% (10)	12% (10)	.39 (.33)
Confessions	6% (5)	6% (3)	4% (3)	.78 (.50)
Fiction	1% (1)	19% (9)	1% (1)	<.01 (.97)

Discourse

The results show that a majority of the search hits for all the studied years have a clear *suicide-preventive* message. In 2005, it was 63 out of 84 hits (75%); in 2009, 27 out of 48 hits (56%); and in 2012, 64 out of 82 hits (78%) (Table 5).

Table 5. The share of search hits on Google on the search term suicide distributed on the variable Discourse. In percent and (frequencies). *P*-values based on Fisher's exact test. *P*-values based on Fisher's exact test for association (and Cochran-Armitage trend statistic in parentheses).

Discourse	Year			
	2005	2009	2012	*P*-values
	n = 84	n = 48	n = 82	
Suicide preventive	75% (63)	56% (27)	78% (64)	.03 (.20)
Pro-suicide	6% (5)	6% (3)	0	<.01 (.053)
Both categories	11% (9)	21% (10)	9% (7)	.12 (.78)
Non-applicable	8% (7)	17% (8)	13% (11)	.33 (.42)

Discussion

Although the *institutional senders* quantitatively dominate the communication about suicide on the Internet, the relatively large proportion of non-institutional *private senders* illustrates how the Internet effects the way that suicide is communicated in today's society.

The analysis of the relation between the variables *senders* and *form of communication* in the 2005, 2009, and 2012 material shows that communication about suicide on the Internet by *institutional senders* is almost entirely in a *monological* form, with very few instances of *dialogue*. In contrast to this, the proportion of the communication form *dialogue* for the *private senders* constitutes 21% in 2005, and in 2009 and 2012, the *dialogue* represents more than a third of the search hits (36%) for these senders. So, today there is a clear distinction between the *institutional senders* and the *private senders* in terms of how to communicate the subject of suicide.

Dialogue is thus a quite common form of communication about suicide for the *private senders*. The Internet, uniquely, consists of several different forms of mass communication, where the interactive multi-way communication is the part that means something qualitatively new to the suicide subject, while the one-way monological communication on the Internet stands for an extension of an already existing institutional communication about the suicide subject. The contents of the dialogic communication are based on the social reality outside the Internet. They are shaped and conventionalized according to the specific factors of the communication that occur in the virtual environment, such as a higher degree of anonymity, a greater spatial distance and, to a large extent, a text-based communication.

The most common content on the subject of suicide on the Internet is about *research and suicide-prevention*. The analysis of the relation between *senders* and *content* in the 2005, 2009 and 2012 material shows that communication about suicide from the *institutional senders* is made up mainly from texts on *research and suicide-prevention* (82%, 62%, 82% respectively). This content category also makes up a substantial proportion for the *private senders* (39%, 14%, 36% respectively), but here the categories *confessions* and *general reflections* also make up a noteworthy share (for *confessions* 22%, 21%, 21% respectively).

The content category *confessions* contain personal and intimate stories and conversations about one's own or other people's suicidal acts and suicidal thoughts. When reading these texts more closely, it is clear that the people writing them are asking questions about the meaning of their existence, that they are crying out for help, that they seek confirmation and understanding, and that they are trying to support and encourage each other the best they can, and in some cases even urge each other to carry out suicidal plans.

According to other studies regarding the representation of suicide on the Internet, pro-suicide websites have shown to be high up on search engines' result lists, making them highly visible and accessible for those who are looking for suicide-related material [33,34]. These findings are congruent with this study's results for the years 2005 and 2009, but it is not as clear in the 2012 material. Another recent study also points out that the preventive websites are more visible and accessible on the Internet today compared to those that are pro-suicide [35]. This does not necessarily mean that pro-suicide websites have declined in numbers; they're just not as prevalent among the first 100 search results, when using the more general search term *suicide*. This could be due to a recent and deliberate strategy (eg, search engine optimization) from suicide-preventive actors to make their websites more reachable and popular, and thereby get a better ranking in the search results lists. Hopefully it is an increased awareness about the risks that pro-suicide websites may have on vulnerable Internet users that lies behind this strategy. These types of websites play a documented role in information on various types of suicide methods, the glorification of individuals who have committed suicide, as well as encouragement to go through with suicide plans [36-40]. Before the development of the Internet, pro-suicide texts were unusual and difficult for audiences to get hold of. The expansion of the Internet has thus enabled pro-suicide messages to spread and can consequently be received by more people today than before. Thus, for an Internet user searching for suicide-related material using a search engine, the likelihood, or risk, of unintentionally finding pro-suicide messages is lower in 2012 than in 2009 and 2005. However, users consciously aiming at finding pro-suicide material will have no problem in doing so [41].

The Challenge for Clinicians

The results show that it is institutional actors, distributing contents of research and suicide-prevention in a monological way, that dominate the communication about suicide on the

Internet. But besides these institutional senders, there is a considerable proportion of private actors for which the Internet has meant the opportunity to publish material and discuss, confess and seek contact on a subject that has always been strongly taboo and therefore "belonged" to only a few voices in public discourse. This opportunity has resulted in both constructive and strongly destructive contributions, such as pro-suicide messages.

Today, it seems a very important task for clinicians to respond to the substantial amount of pro-suicide messages on the Internet and to continue to develop preventive strategies and resources for individuals at risk for suicidal acts [42]. The preventive resources should also be developed using the *dialogic* communication between the mental-health expertise and help-seeking individuals; unidirectional information about mental illness and suicide prevention is not enough [43]. Another important task is to raise awareness among clinicians about the risks it may pose to suicidal individuals to visit pro-suicide websites, which contain descriptions and evaluations of potent suicide methods and promote an attitude that suicide can be an acceptable way to solve life problems. A key instrument could be to establish routines in which clinicians ask their patients about their use of the Internet and also help the patients to find preventive sites with therapeutic resources.

Conclusion

Through the Internet, the dominant understanding of suicide is challenged by new voices, and the battle over definitions of "right and wrong" and perceptions of "true and false" has intensified. The possibility of reaching out to large groups of users is no longer monopolized by institutional senders. These developments brought on by the wide spread use of the Internet have posed obstacles as well as benefits in the field of suicide prevention.

Clinicians taking part in the online dialogical communication, on platforms where patients/laypersons discuss mental health-related topics, could be a helpful tool in the development of cohesion between patient and clinicians [44], and increase the understanding of the patient perspective.

Acknowledgments

The authors would like to thank Professor Jerzy Wasserman for his significant contributions to the intellectual content of the paper.

Conflicts of Interest

None declared.

References

1. Bertolote JM, Fleischmann A. A global perspective on the magnitude of suicide mortality. In: Wasserman D, Wasserman C, editors. Oxford Textbook of Suicidology and Suicide Prevention. Oxford: Oxford University Press; 2009:91-99.
2. Wasserman D. Evaluating suicide prevention: various approaches needed. World Psychiatry 2004 Oct;3(3):153-154. [Medline: 16633481]
3. Joiner T. Why people die by suicide. Cambridge, MA: Harvard University Press; 2005.
4. Anderson M, Jenkins R. The role of the state and legislation in suicide prevention. The five continents perspective. In: Wasserman D, Wasserman C, editors. Oxford Textbook of Suicidology and Suicide Prevention. Oxford: Oxford University Press; 2009:373-381.
5. Westerlund M. The production of pro-suicide content on the Internet: a counter-discourse activity. New Media & Society 2011;14:764. [doi: 10.1177/1461444811425221]
6. Durkee T, Hadlaczky G, Westerlund M, Carli V. Internet pathways in suicidality: a review of the evidence. Int J Environ Res Public Health 2011 Oct;8(10):3938-3952. [doi: 10.3390/ijerph8103938] [Medline: 22073021]
7. Internet World Stats. URL: http://www.internetworldstats.com/stats.htm [accessed 2012-09-06] [WebCite Cache ID 6ATwUasMf]
8. Westerlund M, Wasserman D. The role of the Internet in suicide prevention. In: Wasserman D, Wasserman D, editors. Oxford Textbook of Suicidology and Suicide Prevention. Oxford: Oxford University Press; 2009:525-532.
9. Alao AO, Soderberg M, Pohl EL, Alao AL. Cybersuicide: review of the role of the internet on suicide. Cyberpsychol Behav 2006 Aug;9(4):489-493. [doi: 10.1089/cpb.2006.9.489] [Medline: 16901252]
10. D'Hulster N, Van Heeringen C. [Cyber-suicide: the role of the internet in suicidal behaviour. A case study]. Tijdschr Psychiatr 2006;48(10):803-807 [FREE Full text] [Medline: 17086945]
11. Baume P, Cantor CH, Rolfe A. Cybersuicide: the role of interactive suicide notes on the Internet. Crisis 1997;18(2):73-79. [Medline: 9286130]
12. Thompson S. The Internet and its potential influence on suicide. Psychiatric Bulletin 1999;23:449-451. [doi: 10.1192/pb.23.8.449]
13. Biddle L, Donovan J, Hawton K, Kapur N, Gunnell D. Suicide and the internet. BMJ 2008 Apr 12;336(7648):800-802. [doi: 10.1136/bmj.39525.442674.AD] [Medline: 18403541]

14. Recupero PR, Harms SE, Noble JM. Googling suicide: surfing for suicide information on the Internet. J Clin Psychiatry 2008 Jun;69(6):878-888. [Medline: 18494533]
15. Lee D. Web of despair. Foreign Policy 2003 Sep 01;138:90-91.
16. Rajagopal S. Suicide pacts and the internet. BMJ 2004 Dec 4;329(7478):1298-1299. [doi: 10.1136/bmj.329.7478.1298] [Medline: 15576715]
17. Naito A. Internet suicide in Japan: implications for child and adolescent mental health. Clin Child Psychol Psychiatry 2007 Oct;12(4):583-597. [Medline: 18095539]
18. Ozawa-De Silva C. Shared death: self, sociality and internet group suicide in Japan. Transcult Psychiatry 2010 Jul;47(3):392-418. [doi: 10.1177/1363461510370239] [Medline: 20688797]
19. Baume P, Cantor CH, Rolfe A. Cybersuicide: the role of interactive suicide notes on the Internet. Crisis 1997;18(2):73-79. [Medline: 9286130]
20. Becker K, Schmidt MH. Internet chat rooms and suicide. J Am Acad Child Adolesc Psychiatry 2004 Mar;43(3):246-247. [Medline: 15076254]
21. Wang YD, Phillips-Wren G, Forgionne G. E-delivery of personalised healthcare information to intermediaries for suicide prevention. Int J Electron Healthc 2005;1(4):396-412. [Medline: 18048226]
22. Winkel S, Groen G, Petermann F. [Social support in suicide forums]. Prax Kinderpsychol Kinderpsychiatr 2005 Nov;54(9):714-727. [Medline: 16305019]
23. Gilat I, Shahar G. Emotional first aid for a suicide crisis: comparison between Telephonic hotline and internet. Psychiatry 2007;70(1):12-18. [doi: 10.1521/psyc.2007.70.1.12] [Medline: 17492908]
24. Baker D, Fortune S. Understanding self-harm and suicide websites: a qualitative interview study of young adult website users. Crisis 2008;29(3):118-122. [Medline: 18714907]
25. Niederkrotenthaler T, Voracek M, Herberth A, Till B, Strauss M, Etzersdorfer E, et al. Role of media reports in completed and prevented suicide: Werther v. Papageno effects. Br J Psychiatry 2010 Sep;197(3):234-243 [FREE Full text] [doi: 10.1192/bjp.bp.109.074633] [Medline: 20807970]
26. Barak A. Emotional support and suicide prevention through the Internet: A field project. Computers in Human Behavior 2007;23(2):971-984.
27. McCarthy MJ. Internet monitoring of suicide risk in the population. J Affect Disord 2010 May;122(3):277-279. [doi: 10.1016/j.jad.2009.08.015] [Medline: 19748681]
28. Brin S, Page L. Cite Seer X. 1998. The anatomy of a large-scale hypertextual web search engine URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.96.2493&rep=rep1&type=pdf [accessed 2011-04-24]
29. Westerlund M. Självmord och Internet: kommunikation om ett livsfarligt ämne (Suicide and the Internet: communication about a deadly topic). Stockholm: Stockholms universitet; 2010.
30. Paasonen S. Net years, pioneers, and flat perspectives. Internet Research Annual 2003;2:3-13.
31. Foot K. Web sphere analysis and cybercultural studies. In: Silver D, Massanari A, editors. Critical cyberculture studies. New York and London: New York University Press; 2006:18-38.
32. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
33. Biddle L, Donovan J, Hawton K, Kapur N, Gunnell D. Suicide and the internet. BMJ 2008 Apr 12;336(7648):800-802. [doi: 10.1136/bmj.39525.442674.AD] [Medline: 18403541]
34. Recupero PR, Harms SE, Noble JM. Googling suicide: surfing for suicide information on the Internet. J Clin Psychiatry 2008 Jun;69(6):878-888. [Medline: 18494533]
35. Kemp CG, Collings SC. Hyperlinked suicide: assessing the prominence and accessibility of suicide websites. Crisis 2011;32(3):143-151. [doi: 10.1027/0227-5910/a000068] [Medline: 21616763]
36. Adekola A, Yolles J, Armenta W. Cybersuicide: The Internet and suicide. American Journal of Psychiatry 1999;156(11):1836-1837.
37. Baume P, Cantor CH, Rolfe A. Cybersuicide: the role of interactive suicide notes on the Internet. Crisis 1997;18(2):73-79. [Medline: 9286130]
38. Becker K, Schmidt MH. Internet chat rooms and suicide. J Am Acad Child Adolesc Psychiatry 2004 Mar;43(3):246-247. [Medline: 15076254]
39. Haut F, Morrison A. The Internet and the future of psychiatry. Psychiatric Bulletin 1998;22(10):641-642. [doi: 10.1192/pb.22.10.641-a]
40. Thompson S. The Internet and its potential influence on suicide. Psychiatric Bulletin 1999;23:449-451. [doi: 10.1192/pb.23.8.449]
41. Hagihara A, Miyazaki S, Abe T. Internet suicide searches and the incidence of suicide in young people in Japan. Eur Arch Psychiatry Clin Neurosci 2012 Feb;262(1):39-46. [doi: 10.1007/s00406-011-0212-8] [Medline: 21505949]
42. Yang AC, Tsai SJ, Huang NE, Peng CK. Association of Internet search trends with suicide death in Taipei City, Taiwan, 2004-2009. J Affect Disord 2011 Jul;132(1-2):179-184. [doi: 10.1016/j.jad.2011.01.019] [Medline: 21371755]
43. McMillan SJ. A four-part model of cyber-interactivity. New Media & Society 2002;4:271-291. [doi: 10.1177/146144480200400208]

44. Mohr DC, Duffecy J, Jin L, Ludman EJ, Lewis A, Begale M, et al. Multimodal e-mental health treatment for depression: a feasibility trial. J Med Internet Res 2010;12(5):e48 [FREE Full text] [doi: 10.2196/jmir.1370] [Medline: 21169164]

Edited by G Eysenbach; submitted 28.10.11; peer-reviewed by M Sarchiapone; comments to author 13.01.12; revised version received 14.02.12; accepted 18.06.12; published 26.09.12.

Please cite as:
Westerlund M, Hadlaczky G, Wasserman D
The Representation of Suicide on the Internet: Implications for Clinicians
J Med Internet Res 2012;14(5):e122
URL: http://www.jmir.org/2012/5/e122/
doi:10.2196/jmir.1979
PMID:23010086

Original Paper

Reducing Suicidal Ideation: Cost-Effectiveness Analysis of a Randomized Controlled Trial of Unguided Web-Based Self-help

Bregje A.J van Spijker[1], PhD; M. Cristina Majo[2], PhD; Filip Smit[2,3], PhD; Annemieke van Straten[1], PhD; Ad J.F.M Kerkhof[1], PhD

[1]Department of Clinical Psychology and the EMGO+ Institute for Health and Care Research, Faculty of Psychology and Education, VU University Amsterdam, Amsterdam, Netherlands

[2]Trimbos Institute, Netherlands Institute of Mental Health and Addiction, Utrecht, Netherlands

[3]Department of Epidemiology and Biostatistics, EMGO+ Institute for Health and Care Research, VU University Medical Centre, Amsterdam, Netherlands

Corresponding Author:
Bregje A.J van Spijker, PhD
Department of Clinical Psychology and the EMGO+ Institute for Health and Care Research
Faculty of Psychology and Education
VU University Amsterdam
van der Boechorststraat 1
Amsterdam, 1081 BT
Netherlands
Phone: 31 5988497
Fax: 31 5988758
Email: baj.van.spijker@gmail.com

Abstract

Background: Suicidal ideation is highly prevalent, but often remains untreated. The Internet can be used to provide accessible interventions.

Objective: To evaluate the cost-effectiveness of an online, unguided, self-help intervention for reducing suicidal ideation.

Methods: A total of 236 adults with mild to moderate suicidal thoughts, defined as scores between 1-26 on the Beck Scale for Suicide Ideation (BSS), were recruited in the general population and randomized to the intervention (n = 116) or to a waitlist, information-only, control group (n = 120). The intervention aimed to decrease the frequency and intensity of suicidal ideation and consisted of 6 modules based on cognitive behavioral techniques. Participants in both groups had unrestricted access to care as usual. Assessments took place at baseline and 6 weeks later (post-test). All questionnaires were self-report and administered via the Internet. Treatment response was defined as a clinically significant decrease in suicidal ideation on the BSS. Total per-participant costs encompassed costs of health service uptake, participants' out-of-pocket expenses, costs stemming from production losses, and intervention costs. These were expressed in Euros (€) for the reference year 2009.

Results: At post-test, treatment response was 35.3% and 20.8% in the experimental and control conditions, respectively. The incremental effectiveness was $0.35 - 0.21 = 0.15$ (SE 0.06, $P = .01$). The annualized incremental costs were −€5039 per participant. Therefore, the mean incremental cost-effectiveness ratio (ICER) was estimated to be −€5039/0.15 = −€34,727 after rounding (US −$41,325) for an additional treatment response, indicating annual cost savings per treatment responder.

Conclusions: This is the first trial to indicate that online self-help to reduce suicidal ideation is feasible, effective, and cost saving. Limitations included reliance on self-report and a short timeframe (6 weeks). Therefore, replication with a longer follow-up period is recommended.

(J Med Internet Res 2012;14(5):e141) doi:10.2196/jmir.1966

KEYWORDS

suicidal ideation; randomized controlled trial; cost-effectiveness; Internet; cognitive behavior therapy

Introduction

Suicidal ideation is highly prevalent and causes considerable disease burden [1], but often remains untreated [2]. Frequently reported barriers to seeking help include a preference to handle the problem alone, believing the problem is not severe, and believing treatment will not be effective [2]. For these reasons, an online self-help intervention specifically aimed at reducing suicidal ideation was developed [3]. Self-help can be defined as a standardized psychological treatment that a participant can work through independently. The rationale for using online delivery for this intervention included the reach, accessibility, and anonymity of the Web, thereby facilitating dissemination. Web-based interventions have been found effective for a range of mental disorders (eg, depression, anxiety, and problem drinking) [4-7].

The potential economic advantages of Web-based interventions are among commonly cited motivations for their development [8]. Indeed, promising results have been published for Web-based interventions targeting both somatic [9-10] and psychological problems [11-15]. However, being a relatively young research field, economic evaluations of Web-based interventions are still scarce and often have limitations [8].

For face-to-face psychological treatments targeting suicidality, some empirical evidence is available for cognitive behavior therapy (CBT) [16-17], dialectical behavior therapy (DBT) [18-20], problem-solving treatment (PST) [21-22], and mindfulness-based cognitive therapy (MBCT) [23-25]. Economic evaluations of psychological treatments for suicidality are almost non-existent, which has been attributed to a general lack of unambiguous effectiveness of treatment programs [26-27]. One study comparing manual-assisted cognitive behavior therapy (MACT) with care as usual found indications that MACT was valuable from an economic perspective; however, results were not conclusive [28]. Furthermore, a review of therapies for borderline personality disorder indicated that DBT could potentially be cost-effective [29].

The economic impact on society of not taking preventive measures (ie, the costs of suicide) have been estimated to be well over £1,000,000 per suicide (2005 prices) [27]. Although no similar estimates have been reported regarding suicidal ideation, these are likely to be substantial when considering the economic burden of depression [30-31], a common mental disorder in people with suicidal ideation.

This paper reports the results of an economic evaluation of a randomized controlled trial comparing online self-help for suicidal ideation with a waitlist control condition (Netherlands Trial Register, NTR1689).

Methods

Design and Participants

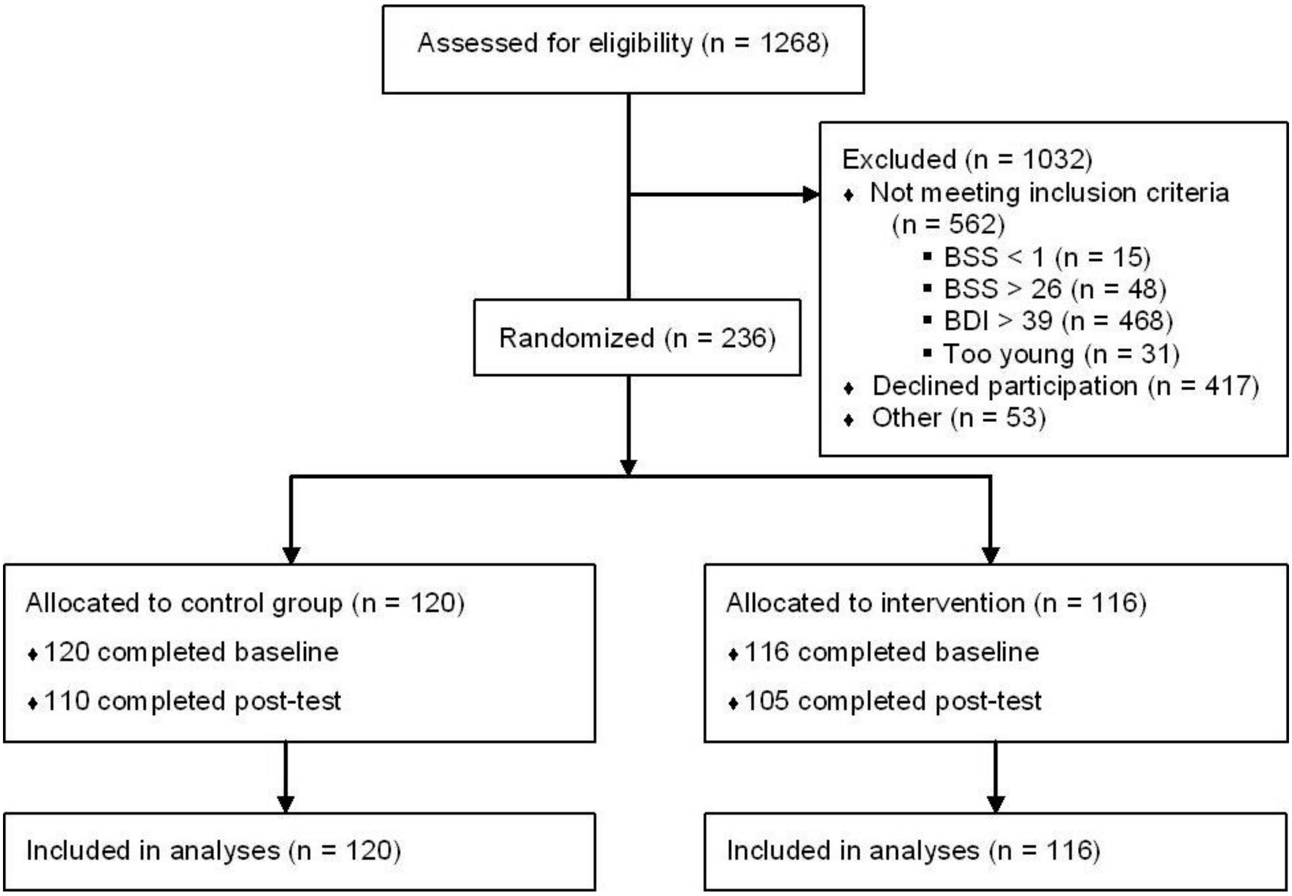
Participants were recruited between October 2009 and November 2010 from the Dutch general population by means of advertisements in newspapers, relevant websites, and Google AdWords. The methods used in this trial have been described in detail elsewhere [3]. To be included, people had to be over 18 years, have access to the Internet and a valid email address, and have a good command of the Dutch language. In addition, they needed to present with a score between 1 and 26 on the Beck Scale for Suicide Ideation (BSS) [32] suggesting mild to moderate suicidal ideation, and a score < 40 on the Beck Depression Inventory (BDI) [33], to avoid including people with severe levels of depression. These criteria were established in consultation with clinical experts.

Eligibility was assessed using an online application procedure. In total, 1268 respondents filled in the screening questionnaires (BSS and BDI). Respondents who exceeded the cutoff scores (562/1268, 44.32%) were referred to other (mental health) services by means of an automated response. Eligible respondents were requested to fill in their email address, after which an information brochure and an informed consent form were emailed to them. A small number did not fill in their email address and were consequently excluded (53/1268, 4.18%). After returning the informed consent form, on which participants had to disclose their identity and that of their family physician, participants (n = 236) were stratified for gender and randomized in blocks of 20 to the intervention (n = 116) or to the waitlisted information-only control condition (n = 120) by an independent researcher using random allocation software. Most of the eligible respondents (417/706, 59.1%) did not return their informed consent form, possibly due to the lack of anonymity when participating (see Figure 1).

After randomization, participants in the intervention group received log-in codes for the self-help intervention. Participants in the control group received a link to an information website and were informed that they would receive access codes to the intervention 6 weeks later. It is worth noting that all participants, in both conditions, had unrestricted access to care as usual (CAU) and were encouraged to make use of this. In the Netherlands, people most commonly go to a family physician who can refer to specialized mental health services if necessary.

Because this study was conducted in a vulnerable population, safety procedures were employed. Each time a participant in either condition exceeded cutoff scores on suicidal ideation or depressive symptoms, a risk assessment was carried out over the phone. If necessary, or if a participant could not be reached, their family physician was contacted [3]. The study was approved by the Medical Ethics Committee of the VU University Medical Centre (registration number 2008/204).

Figure 1. Flowchart of participants in the trial.



Intervention

The experimental group received an online, unguided, self-help intervention aimed at decreasing the frequency and intensity of their suicidal ideation. This intervention is based on CBT (DBT, PST, and MBCT). All of these treatment programs have evidence for their effectiveness in reducing suicidality [18,24,34-35].

The intervention consists of six weekly modules which consecutively focus on (1) the repetitive character of suicidal thoughts [36], (2) dealing with intense emotions, (3) identifying negative automatic thoughts, (4) learning to recognize thinking patterns, (5) reformulating negative thoughts, and (6) relapse prevention. A more detailed description of the intervention has been previously published [3]. The intervention is currently available through 113Online (www.113online.nl), a Dutch online suicide prevention platform [37,38].

Participants were encouraged to follow one module per week and they received an automated weekly motivating email. Approximately half (56.0%, 65/116) of the participants in the intervention group completed at least three modules of the intervention, 21.6% (25/116) completed the whole intervention, and a similar percentage (22.4%, 26/116) did not start the intervention.

If desired, participants were able to ask questions pertaining to the intervention via the website. Questions asked were often about specific exercises (eg, “What thoughts should I tally?” and “Should I continue worry time in the second module?”) or about the other aspects of the website (eg, “How long will my log-in codes be valid for?”). These were answered by the researchers, taking an average of 6 minutes per participant over the entire intervention period.

The control group received a link to a website created for the study that provided information on suicidality, such as prevalence, warning signs, and risk factors. Pretesting indicated that a maximum of 15 minutes was needed to read this information. In addition, links to relevant mental health centers were provided and participants were advised to use these.

Power Analyses

Sample size was based on the expected effect on the primary outcome measure (ie, the reduction of suicidal thoughts). In order to be able to detect an effect size of 0.35 with alpha = .05 and beta = .80, 100 subjects were needed in each condition. Including an expected drop-out attrition rate of 20% to 30% in each group, the sample size was determined at 260.

Outcome Measures

Questionnaires were self-report and administered via the Internet. For the current paper, data from baseline and post-test (6 weeks after baseline) were used.

Primary Clinical Outcome

The primary clinical outcome in this paper is suicidal ideation, assessed using the BSS [32]. This self-report questionnaire consisted of 21 items, each scored on a 0-2 scale. The total score was obtained by adding items 1-19 (range 0-38). The last two items relate to suicide attempts and the intent to die during the most recent attempt. Internal reliability of the BSS is high, with Cronbach alpha ranging between .87 and .97 [39]. In this study, Cronbach alpha = .89 at baseline.

Resource Use and Costing

A societal perspective was adopted in this study; therefore, the costs of health service uptake, patients' out-of-pocket costs, and production losses in paid work were included. Data on health care uptake and production losses were collected using the Trimbos/Institute of Medical Technology Assessment Questionnaire for Costs associated with Psychiatric Illness (TIC-P) [40], a health service receipt questionnaire that is widely used in economic evaluations in the Netherlands. This produces three cost categories: direct medical costs, direct non-medical costs, and indirect non-medical costs. Data were collected for two periods: the 6 weeks prior to baseline and the 6 weeks following baseline.

Direct medical costs relate to the utilization of health care services. To calculate these costs, health service units were multiplied by their standard full economic cost prices as reported in the Dutch guidelines [41] for health economic evaluations for the reference year 2009 (see Table 1). The costs of prescription psychotropic drugs (eg, antidepressants, benzodiazepines, and antipsychotics) were calculated as the price per standard daily dose as reported in the Dutch Pharmacotherapeutic Compass [42], multiplied by the number of prescription days, plus pharmacists' dispensing costs of €14 per prescription.

Table 1. Direct medical and direct non-medical costs by health service type.

Health service type	Direct medical costs		Direct non-medical costs		
	Unit	Unit cost price[a] (€)	Distance (km)[b]	Time (h)[b]	Unit cost price[c] (€)
General practitioner	Contact	28	1.1	1	15.72
Company doctor[d]	Contact	28	17.6	0.5	9.77
Social worker	Contact	65	5	2	29
Private practice psychotherapist, psychiatrist	Contact	90[e]	7	2	29.40
Alcohol and drug consultant (CAD)	Contact	171	10[f]	3	42.50
Regional mental health service	Contact	171	7	3	41.90
Physiotherapist	Contact	36	2.2	2	28.44
Mental hospital	Contact	173	7	4	54.40
Medical specialist general hospital	Contact	72	7	3	41.90
Alternative treatment[g]	Contact	50.70	5	1	16.50
Daycare, mental health treatment	Contact	154	7	4	54.40
Home care	Hour	35	NA	NA	NA
Informal care (family, friends)[h]	Hour	12.50	NA	NA	NA

[a] Integral unit cost prices [41] presented in 2009 €.

[b] Based on average distances (in special tariff taxi and public transport zones) and travel + waiting + treatment times (in hours) for receiving treatment [41].

[c] Costs = (0.2 × km) + 3 + (12.5 × hrs). With €0.20 = cost per km; €3 = 1 h parking time; €12.5 = 1 h time [41].

[d] No parking costs assumed.

[e] Own calculation, valued as average of private practice psychotherapist and psychiatrist [41].

[f] Assumed as CAD were more dispersed than regional mental health services.

[g] Own calculation, valued as average of homoeopath and acupuncturist [41].

[h] Valued as domestic help [41].

Direct non-medical costs encompassed participants' travel expenses to receive professional help and loss of leisure time, the latter valued at €12.50 per hour [41]. Additionally, informal caregivers' (ie, friends, neighbors, and family) use of time (eg, running errands for participants), was valued at €12.50 per hour (see Table 1).

Finally, the costs stemming from production losses in paid work (indirect non-medical costs) were calculated from the number of days absent from work (absenteeism), plus the number of days spent at work with reduced efficiency, corrected for degree of inefficiency (presenteeism). Table 2 reports the age-specific economic costs of each hour of lost productivity for men and women.

Table 2. Productivity costs by gender and age class [41].

Age range	Gender	
	Men (€)[a]	Women (€)[a]
15-19	9.65	8.76
20-24	17.75	17.18
25-29	24.19	23.62
30-34	29.65	27.54
35-39	34.03	29.25
40-44	36.67	29.06
45-49	38.32	28.91
50-54	39.06	29.25
55-59	39.38	29.50
60-64	39.13	28.67
65+	39.13	28.67

[a] Costs are indexed from the Collective Labor Agreement, 2008 (2.8%) and presented in 2009 € values.

The total associated costs (in Euros) can be converted to US dollars using the purchasing power parity (PPP) exchange rates reported by the Organisation for Economic Co-operation and Development (OECD) [43], which converts currency while taking into account the differential buying power across countries. For the reference year 2009, US $1.00 was equivalent to €0.848173.

Per-participant Intervention Costs

In estimating per-participant intervention costs, the average time spent on the intervention was valued at €12.50 per hour (leisure time value [41]) for an average of 10.5 hours per participant over the 6-week intervention period. In addition, psychologist time spent answering questions was included for an average of 6 minutes per participant over the course of the intervention (at €154/hour [41]). Further costs were related to website maintenance, which amounted to €1740 and €10,000 per annum for upgrading and hosting the website, respectively.

Relying on data from a Dutch population survey on suicidality [44] and Statistics Netherlands [45], 91% of the Dutch adult population with suicidal ideation (N = 180,000) has Internet access (n = 163,800). Taking a conservative approach, it was assumed that 40% would search for online help, and 10% of these people would engage in the online self-help intervention. This resulted in an estimated usage by 6552 participants per year.

Based on the above assumptions and data, the per-participant intervention costs were estimated to be €148 (US $176).

Analysis

Statistical Analyses

Analyses were carried out on an intention-to-treat basis. Therefore, all participants were analyzed in the condition to which they were randomized and missing data at post-test for the BSS (n = 21, 8.9%) were imputed using regression imputation as implemented in Stata data analysis and statistical software (StataCorp LP, College Station, Texas, USA) with age, gender, employment status, education, relationship status, nationality, baseline clinical outcomes (ie, suicidal ideation, depression, hopelessness, worry, and anxiety) and randomization status as predictor variables.

For suicidal ideation, a reliable and clinically significant change was calculated to be 6.48 points on the BSS according to the Jacobson and Truax method [46]. Participants were dichotomized according to this criterion into treatment responders and non-responders.

In addition to the primary clinical outcome, the use of the safety procedures is reported (in number of phone calls with participants who exceeded cutoff scores during the trial and referrals to the family physician), as well as the number of suicide attempts.

Cost-Effectiveness Analyses

Missing cost data at post-test (between 1% and 18% depending on the type of costs) were imputed using similar regression imputation as for the BSS.

The mean total costs for each of the conditions were calculated at baseline and post-test. Since mean baseline costs were similar across both conditions (see Results), the incremental costs were calculated as the between-group difference at post-test. For reasons of comparability, annualized costs are presented.

Both incremental costs and incremental effects were used to calculate the incremental cost-effectiveness ratio (ICER). The ICER was calculated as $(C_1 - C_0)/(E_1 - E_0)$, where C is the average annual per-participant cost and E is the proportion of treatment responders in the experimental and control conditions (subscripted 1 and 0, respectively). The ICER describes the incremental costs for gaining one additional treatment response [47-49]. One additional treatment response is defined as 1 participant improving at least 6.48 points on the BSS.

Non-parametric bootstraps were used to simulate 2500 ICERs that were plotted on the cost-effectiveness plane. In this way, the degree of uncertainty associated with the ICER is captured [50]. Each simulated ICER can potentially fall into one of the four quadrants of the ICER plane. The northeast (NE) quadrant represents superior health gains associated with the intervention, but at additional costs relative to routine care. This scenario is typically encountered in economic evaluations—better health is obtained for additional costs. In the northwest (NW) quadrant, health diminishes while costs increase. Clearly, this is the worst possible outcome because the intervention is "dominated" by CAU. In the southwest (SW) quadrant, health diminishes, but there are cost savings. Finally, in the southeast (SE) quadrant, the intervention generates superior health gains (relative to the comparator condition) and does so for lower costs; the intervention "dominates" the comparator condition, which is the best possible outcome.

Use of willingness to pay (WTP) estimates is another method for determining value for money. By assigning hypothetical maximum WTP amounts (ceilings), ranging from €0 to €100,000 per treatment responder, probability estimates for the acceptability of the intervention compared with CAU from a cost-effectiveness point of view, were calculated. The relationship between each assigned WTP ceiling and the probability that the new intervention is viewed as acceptable, can be plotted in an ICER acceptability curve.

Sensitivity Analysis

The estimated per-participant intervention costs are surrounded by some uncertainty. To ascertain the robustness of the overall findings, all analyses were repeated for three alternative scenarios, encompassing 10, 20, or 30 minutes of additional guidance per participant, per module (ie, 1, 2, and 3 hours, respectively, per participant during the intervention). These are relevant scenarios because guidance is often provided with Web-based interventions. It was assumed that guidance would be provided by a clinical psychologist, and conservatively, that more therapist time would not increase clinical effectiveness.

Results

Sample Characteristics

Participants had a mean age of 40.9 years (SD 13.7). The majority of the sample was female (156/236, 66.1%) and born in the Netherlands (218/232, 94.0%). Approximately half of the sample had completed high school or intermediate vocational training (112/236, 47.5%) and 38.1% (90/236) had completed higher vocational or academic training. A minority was living with a partner (95/236, 40.3%) and had children (87/232, 37.5%). Half of the sample was in paid employment (116/232, 50.0%). Mean score for suicidal ideation was 14.9 at baseline (SD 7.1) There were no significant differences in sociodemographic or clinical characteristics between the intervention and control groups, indicating that randomization had resulted in comparable groups (see Table 3).

Table 3. Baseline characteristics of total sample.

Characteristic	Total (n = 236)	Condition		*P*
		Control (n = 120)	Intervention (n = 116)	
Gender, n (%)				
Female	156 (66.1)	80 (66.7)	76 (65.5)	.85
Age, mean (SD)	40.93 (13.71)	41.39 (13.39)	40.46 (14.07)	.60
Education, n (%)				
Lower	19 (8.1)	8 (6.7)	11 (9.5)	
Intermediate	112 (47.5)	52 (43.3)	60 (51.7)	.36
Higher	90 (38.1)	51 (42.5)	39 (33.6)	
Other	15 (6.4)	9 (7.5)	6 (5.2)	
Living with a partner, n (%)	95 (40.3)	54 (45.0)	41 (35.3)	.13
Has children, n (%)[a]	87 (37.5)	50 (42.0)	37 (32.7)	.14
Born in the Netherlands, n (%)[a]	218 (94.0)	111 (93.3)	107 (94.7)	.65
Paid employment, n (%)[a]	116 (50.0)	59 (49.6)	57 (50.4)	.89
Suicidal thoughts, mean (SD)	14.85 (7.08)	14.50 (7.33)	15.20 (6.82)	.44

[a] Missing: n = 4 (1 in control and 3 in intervention group).

In the 6 weeks prior to baseline, the mean per-participant total costs were €1227 (SD 2364) in the intervention group and €1323 (SD 1891) in the control group, indicating that randomization produced evenly distributed costs across the conditions (t_{234} = 0.346, *P* = .73).

Safety Procedures

As part of the safety procedures, 50 participants were called because they exceeded cutoff scores on suicidal ideation and/or depressive symptoms (31 in the control group and 19 in the intervention group). For 12 of them, the family physician was called (9 in the control group and 3 in the intervention group). Participants were not pulled from the study after being called as part of the safety procedures. Furthermore, 11 participants reported a suicide attempt (7 in the control group and 4 in the intervention group). No suicides occurred during the study.

Incremental Costs

The average total annualized per-participant costs were calculated to be €13,303 in the intervention group and €18,343 in the control group. The incremental costs were €13,303 − €18,343 = −€5039 (rounded to the nearest Euro) per participant per year (equivalent to a cost savings of US $5941). Table 4 shows the cost components by condition (control and intervention groups) and time (at baseline and post-test). The main difference between the conditions can be observed in costs associated with productivity losses (ie, costs stemming from absenteeism, presenteeism, and domestic help). There was an increase in costs due to absenteeism and domestic work in the control group between baseline and post-test, whereas these costs decreased in the intervention group.

Table 4. Cost distribution by condition and time.

Condition	Test time costs[a]	
	Baseline, € (SD)	Post-test, € (SD)
Care as usual (n = 120)		
Direct medical costs	441.61 (1005.70)	558.70 (953.17)
Medication costs	19.18 (50.60)	21.43 (57.75)
Direct non-medical costs	142.02 (288.11)	175.86 (282.23)
Presenteeism[b]	342.63 (818.06)	278.06 (559.30)
Absenteeism[b]	337.53 (934.96)	392.50 (1135.25)
Domestic help costs[b]	37.77 (135.47)	69.37 (448.96)
Intervention costs	NA	NA
Total costs[c]	1322.98 (1890.93)	1528.56 (1911.82)
Online self-help (n = 116)		
Direct medical costs	459.31 (1541.31)	459.36 (607.32)
Medication costs	37.55 (84.70)	28.04 (70.84)
Direct non-medical costs	145.37 (475.21)	142.11 (180.75)
Presenteeism[b]	276.56 (620.38)	207.22 (453.04)
Absenteeism[b]	288.13 (1168.70)	251.29 (914.84)
Domestic help costs[b]	19.83 (54.47)	15.52 (44.04)
Intervention costs	NA	148.00 (0)
Total costs[c]	1226.77 (2364.29)	1244.20 (1404.37)

[a] Mean costs on monthly basis [41] presented in 2009 values.

[b] Presenteeism, absenteeism, and domestic help all relate to production losses.

[c] Total costs are the sum of the other cost components. Differences in the totals are due to rounding.

Incremental Effectiveness

In the intervention group, 35.3% (41/116) met the criteria for clinically significant change, compared with 20.8% (25/120) in the control group. The difference in effectiveness was 0.353 − 0.208 = 0.15 (SE 0.06). This difference was evaluated using a linear probability model while taking into account the clustered data structure (z = 2.51, P = .01, 95% CI 0.03 0.26).

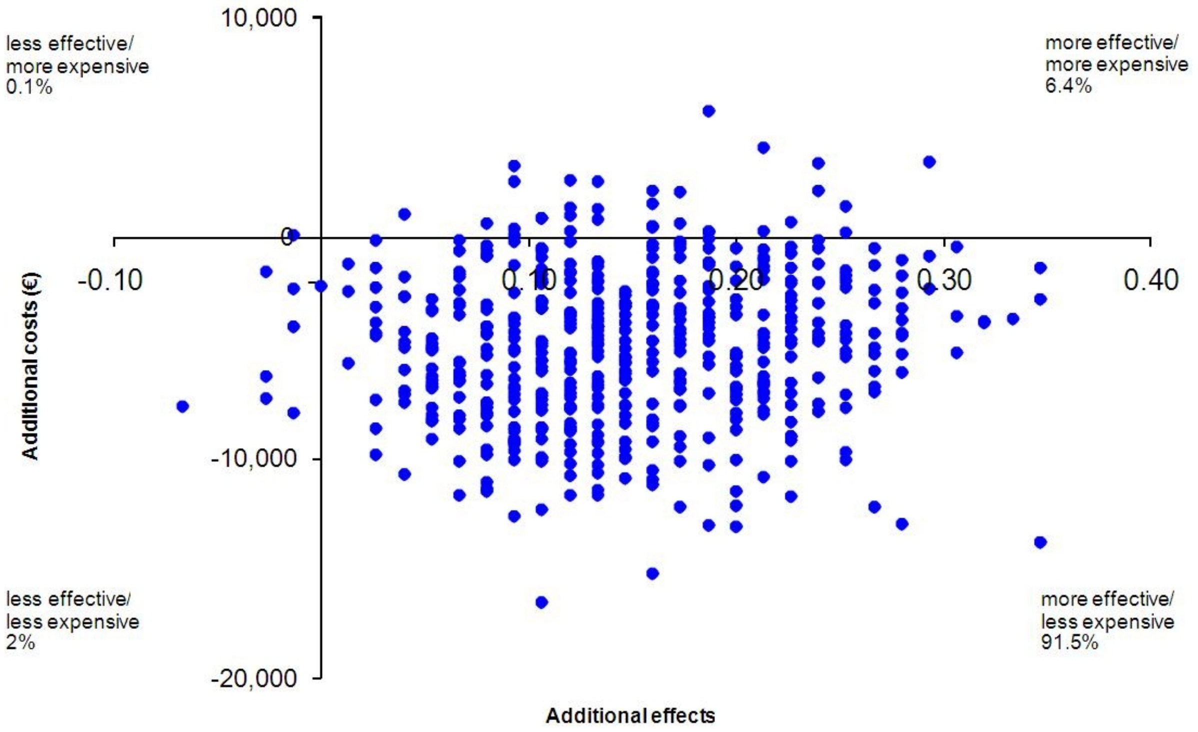
Incremental Cost-Effectiveness

As noted, the incremental costs were −€5039 (negative costs, hence a cost reduction) and the incremental effect was 0.15. Therefore, the mean incremental cost-effectiveness ratio (ICER) was estimated to be −€5039/0.15 = −€34,727 after rounding (a savings of US $41,325) for an additional treatment response. Using the 2500 bootstraps, the median ICER could be estimated as −€31,921 (a savings of US $37,985), essentially conveying the same message.

On the incremental cost-effectiveness plane, each data point represents one simulated ICER. Of these, 91.5% fall into the southeast quadrant, indicating that greater health gains are generated for less cost by the intervention relative to CAU. In addition, 6.4% of the simulated ICERs fall in the northeast quadrant, indicating a probability of 6.4% that by applying the intervention a health gain is produced, but at additional costs. The remainder of the simulated ICERs show up on the west side of the plane, indicating less effectiveness and less cost

(2%), or less effectiveness and more cost (0.1%) (see Figure 2).

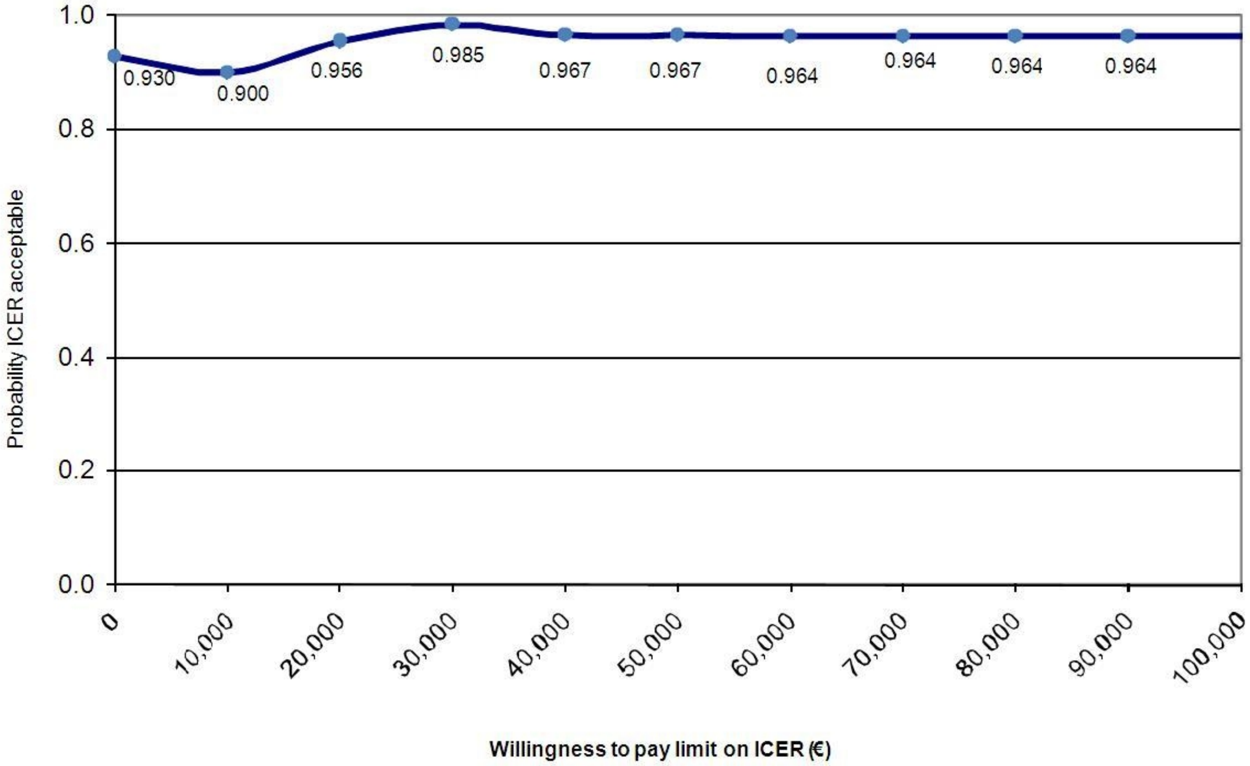
Figure 2. Distribution of bootstrapped incremental cost-effectiveness ratios (ICERs) (n = 2500) on the cost-effectiveness plane, primary analysis.



Acceptability

The incremental cost-effectiveness acceptability curve (Figure 3) suggests that with no willingness to pay for one significantly improved participant, there is a 93% probability that the intervention would be regarded as more cost-effective than CAU. When the willingness to pay for a favorable treatment response is €10,000, €20,000, or €30,000, this probability is 90.4%, 95.6%, and 98.5%, respectively. The minor variations in probabilities between the WTP ceilings imply that the intervention is acceptable from an economic perspective, irrespective of WTP. It can be concluded from this that the intervention can be regarded as acceptable from a cost-effectiveness point of view and that this conclusion is not sensitive to the WTP ceiling used.

Figure 3. Incremental cost-effectiveness ratio (ICER) acceptability curve: Probability that the intervention is acceptable relative to care as usual (y-axis) given varying thresholds for willingness to pay (x-axis) based on 2500 bootstrap replications.



Sensitivity Analysis

Increasing the intervention costs by adding varying amounts of guidance did not affect the overall conclusion that the intervention produces better health outcomes at lower costs, compared with CAU (Table 5). In the first scenario (Scenario A: 1 hour of psychologist support), the median ICER was −€32,342 per treatment responder. The median ICER increased to −€32,708 if the psychologist time was increased by 100%, an ICER that has a 93% probability of falling below the zero willingness to pay threshold (Scenario B: 2 hours of support). Similarly the median ICER increased to −€31,647 if the psychologist time was increased to 180 minutes, an ICER that still has a 93% probability of falling below the zero willingness to pay threshold. The outcomes of these sensitivity analyses are presented in Table 5. These increases in ICERs do not affect the overall conclusion. This indicates that the intervention on top of CAU still produces better health at lower costs, compared with CAU alone.

Table 5. Sensitivity analysis of the incremental cost-effectiveness for different scenarios.

Sensitivity analysis	Standard self-help intervention	Scenario		
		A[a]	B[b]	C[c]
Cost, €[d]	−5039	−4900	−4746	−4592
Effect	0.15	0.15	0.15	0.15
incremental cost-effectiveness ratio, median €[e]	−33,593	−32,342	−32,708	−31,647
Distribution on the cost-effectiveness plane				
1st quadrant (northeast)	0.06	0.06	0.06	0.08
2nd quadrant (inferior: northwest)	0.00	0.00	0.00	0.00
3rd quadrant (southwest)	0.02	0.01	0.02	0.01
4th quadrant (dominant: southeast)	0.91	0.92	0.90	0.89
Willingness to pay ceiling, %				
€0	93	93	93	93
€10,000	90	90	91	88
€20,000	96	95	96	94
€30,000	98	99	99	99

[a] Intervention guided by a clinical psychologist for 10 minutes per module, per participant (ie, 1 hour guidance per participant for the whole intervention).

[b] Guidance by psychologist for 20 minutes per module, per participant (ie, 2 hours guidance per participant for the whole intervention).

[c] Guidance by psychologist for 30 minutes per module, per participant (ie, 3 hours guidance per participant for the whole intervention).

[d] Cost per disease-free year at 2009 prices.

[e] Median is 50th percentile of 2500 bootstrap replications of the ICER.

Discussion

Main Findings

The aim of this paper was to determine whether online self-help for suicidal thoughts would be cost-effective, using data from the first randomized controlled trial comparing online self-help for suicidal ideation on top of CAU to CAU alone. The proportion of participants that showed clinically significant change in suicidal ideation was significantly higher in the intervention group: 35% compared with 21% in the control group. For each significantly improved participant, €34,727 (US $41,325) of societal costs were saved relative to CAU. The finding that different willingness to pay ceilings only minimally affects cost-effectiveness probability estimates also demonstrates that it is a preferable option from a health economic point of view. Sensitivity analyses confirmed the robustness of these findings.

In general, these results add to the observation that Web-based interventions can be favorable from an economic perspective for a range of disorders [8]. However, because no previous cost-effectiveness analyses have been reported for online self-help for suicidal ideation, the obtained results cannot be directly compared. Therefore, it may be more appropriate to compare the results with previous cost-effectiveness studies of face-to-face interventions targeting suicidality, although these are also scarce [26-27]. At best, it can be concluded that the obtained results are in line with the finding that MACT is more effective and cheaper in generating a 1% reduction in the proportion of patients with a self-harm episode than CAU [28]. As MACT is a brief and manual-based treatment, it may be more comparable to self-help than regular face-to-face treatment. Furthermore, a review of therapies for borderline personality disorder (in which suicidality is common) assessed cost-effectiveness in terms of costs per “parasuicide event avoided.” Although results indicated that DBT could potentially be cost-effective, the mixture of results, high levels of uncertainty, and other limitations prevented clear supportive conclusions [29]. It is important to keep in mind that the outcomes in the above comparisons related to suicidal behavior, whereas our study was aimed at suicidal ideation.

Because comparison with previous cost-effectiveness studies targeting suicidality is limited, our results may also be compared with interventions for depression, a common mental disorder in people with suicidal ideation. In this respect, economic evaluations have shown that guided online self-help for depression has a high probability (91%) of being cost-effective compared to CAU [11]. Also, unguided online self-help and therapist-delivered online CBT have been found to be more efficient than CAU [51-52].

Strengths and Limitations

The findings reported here should be interpreted with caution. Firstly, because of the relatively short time period of 6 weeks, it is unknown how the cost-effectiveness of online self-help is affected after a longer follow-up period. Some economic costs may not have been incurred in this period. Secondly, data on health care consumption and production loss in this study were based on self-report, which may have introduced recall bias. For example, self-report of health care uptake may have been underestimated or overestimated, depending upon the health resource [53-55]. However, because participants were randomized, such a bias is expected to occur in both groups. Finally, several assumptions and estimates were made when calculating the per-participant intervention costs, in particular the number of people who would engage in online self-help for suicidal ideation and costs related to website maintenance. Because these estimates could not be based entirely on actual data, they would need “real world” verification. Unforeseen variations may arise after implementation.

Strengths of this study pertain to its randomized design and the large sample size. Furthermore, because health care utilization data and data on production losses were available, it was possible to study the cost-effectiveness of the intervention from a societal perspective. However, possible spillover effects onto third parties not involved in this study (eg, effects on family and friends) could not be taken into account because data on these were not collected or available.

Implications

From a research perspective, it is evident that this study needs replication to verify results both within and outside the Netherlands. The latter may be challenging because this would require designing safety procedures that match both general and local ethical and legal considerations, for which no ready-made recipe is available. Moreover, this holds true for implementation of the intervention in practice as well.

From a clinical perspective, it is important to keep in mind that suicidal ideation was the primary focus of this study. It was not designed to detect differences at the level of attempted suicide, so it is unknown whether these could be decreased by online self-help. Evidently, the same is true regarding suicide. However, it still seems a remarkable result that suicidal ideation can be reduced in a cost-effective way, especially given that all participants were encouraged to engage in CAU. Moreover, the control group made more use of this than the intervention group, further strengthening the findings. Similarly, the fact that participants in the control group were called more often due to exceeding the cutoff score supports this.

Finally, it is important to note that this online self-help program was not meant to replace face-to-face contact, but was designed for people who are reluctant to seek face-to-face care. Ideally, people who struggle with suicidality should be seen in person by mental health professionals. In this respect, the intervention studied here can also serve as a complement to face-to-face treatment.

Conclusion

With respect to psychological interventions targeting suicidality, economic evaluations are practically non-existent. Findings suggest that offering an online intervention on top of CAU increases the likelihood of a favorable outcome. Moreover, these improved clinical outcomes are achieved at lower cost. However, more studies with longer follow-up periods are needed to further substantiate these findings. Therefore, results of this economic evaluation may best be regarded as initial tentative proof of a promising online self-help intervention for suicidal thoughts.

Acknowledgments

The study was supported by a grant from the Netherlands Health Research Council, The Hague (ZonMw) (project number 120510003). This funding organization was not involved in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; or the preparation, review, and approval of the manuscript.

Authors' Contributions

BvS, AvS, and AK designed the study. BvS carried out the study and collected the data. CM and FS analyzed the data. All authors contributed to writing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-eHealth Checklist V1.6 [56].

[PDF File (Adobe PDF File), 572KB - jmir_v14i5e141_app1.pdf]

References

1. van Spijker BA, van Straten A, Kerkhof AJ, Hoeymans N, Smit F. Disability weights for suicidal thoughts and non-fatal suicide attempts. J Affect Disord 2011 Nov;134(1-3):341-347. [doi: 10.1016/j.jad.2011.05.020] [Medline: 21652085]
2. Bruffaerts R, Demyttenaere K, Hwang I, Chiu WT, Sampson N, Kessler RC, et al. Treatment of suicidal people around the world. Br J Psychiatry 2011 Jul;199(1):64-70 [FREE Full text] [doi: 10.1192/bjp.bp.110.084129] [Medline: 21263012]

3. van Spijker BA, van Straten A, Kerkhof AJ. The effectiveness of a web-based self-help intervention to reduce suicidal thoughts: a randomized controlled trial. Trials 2010;11:25 [FREE Full text] [doi: 10.1186/1745-6215-11-25] [Medline: 20214777]
4. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
5. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009 Dec;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
6. Cuijpers P, Marks IM, van Straten A, Cavanagh K, Gega L, Andersson G. Computer-aided psychotherapy for anxiety disorders: a meta-analytic review. Cogn Behav Ther 2009 Jun;38(2):66-82. [doi: 10.1080/16506070802694776] [Medline: 20183688]
7. 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]
8. Tate DF, Finkelstein EA, Khavjou O, Gustafson A. Cost effectiveness of internet interventions: review and recommendations. Ann Behav Med 2009 Aug;38(1):40-45. [doi: 10.1007/s12160-009-9131-6] [Medline: 19834778]
9. Andersson E, Ljótsson B, Smit F, Paxling B, Hedman E, Lindefors N, et al. Cost-effectiveness of internet-based cognitive behavior therapy for irritable bowel syndrome: results from a randomized controlled trial. BMC Public Health 2011;11:215 [FREE Full text] [doi: 10.1186/1471-2458-11-215] [Medline: 21473754]
10. Kaldo V, Levin S, Widarsson J, Buhrman M, Larsen HC, Andersson G. Internet versus group cognitive-behavioral treatment of distress associated with tinnitus: a randomized controlled trial. Behav Ther 2008 Dec;39(4):348-359. [doi: 10.1016/j.beth.2007.10.003] [Medline: 19027431]
11. Warmerdam L, Smit F, van Straten A, Riper H, Cuijpers P. Cost-utility and cost-effectiveness of internet-based treatment for adults with depressive symptoms: randomized trial. J Med Internet Res 2010;12(5):e53 [FREE Full text] [doi: 10.2196/jmir.1436] [Medline: 21169166]
12. Cunningham JA, Khadjesari Z, Bewick BM, Riper H. Internet-based interventions for problem drinkers: From efficacy trials to implementation. Drug Alcohol Rev 2010 Nov;29(6):617-622. [doi: 10.1111/j.1465-3362.2010.00201.x] [Medline: 20973846]
13. Mihalopoulos C, Kiropoulos L, Shih ST, Gunn J, Blashki G, Meadows G. Exploratory economic analyses of two primary care mental health projects: implications for sustainability. Med J Aust 2005 Nov 21;183(10 Suppl):S73-S76. [Medline: 16296957]
14. Gerhards SA, 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]
15. McCrone P, Knapp M, Proudfoot J, Ryden C, Cavanagh K, Shapiro DA, et al. Cost-effectiveness of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. Br J Psychiatry 2004 Jul;185:55-62 [FREE Full text] [Medline: 15231556]
16. Tarrier N, Taylor K, Gooding P. Cognitive-behavioral interventions to reduce suicide behavior: a systematic review and meta-analysis. Behav Modif 2008 Jan;32(1):77-108. [doi: 10.1177/0145445507304728] [Medline: 18096973]
17. Wenzel A, Brown GK, Beck AT. Cognitive Therapy for Suicidal Patients: Scientific and Clinical Applications. Washington, DC: American Psychological Association (APA); 2009.
18. Linehan MM, Comtois KA, Murray AM, Brown MZ, Gallop RJ, Heard HL, et al. Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. Arch Gen Psychiatry 2006 Jul;63(7):757-766. [doi: 10.1001/archpsyc.63.7.757] [Medline: 16818865]
19. van den Bosch LM, Koeter MW, Stijnen T, Verheul R, van den Brink W. Sustained efficacy of dialectical behaviour therapy for borderline personality disorder. Behav Res Ther 2005 Sep;43(9):1231-1241. [doi: 10.1016/j.brat.2004.09.008] [Medline: 16005708]
20. Verheul R, Van Den Bosch LM, Koeter MW, De Ridder MA, Stijnen T, Van Den Brink W. Dialectical behaviour therapy for women with borderline personality disorder: 12-month, randomised clinical trial in The Netherlands. Br J Psychiatry 2003 Feb;182:135-140 [FREE Full text] [Medline: 12562741]
21. Hawton K, Townsend E, Arensman E, Gunnell D, Hazell P, House A, et al. Psychosocial versus pharmacological treatments for deliberate self harm. Cochrane Database Syst Rev 2000(2):CD001764. [doi: 10.1002/14651858.CD001764] [Medline: 10796818]
22. Townsend E, Hawton K, Altman DG, Arensman E, Gunnell D, Hazell P, et al. The efficacy of problem-solving treatments after deliberate self-harm: meta-analysis of randomized controlled trials with respect to depression, hopelessness and improvement in problems. Psychol Med 2001 Aug;31(6):979-988. [Medline: 11513383]
23. Barnhofer T, Crane C, Hargus E, Amarasinghe M, Winder R, Williams JM. Mindfulness-based cognitive therapy as a treatment for chronic depression: A preliminary study. Behav Res Ther 2009 May;47(5):366-373. [doi: 10.1016/j.brat.2009.01.019] [Medline: 19249017]

24. Williams JM, Duggan DS, Crane C, Fennell MJ. Mindfulness-based cognitive therapy for prevention of recurrence of suicidal behavior. J Clin Psychol 2006 Feb;62(2):201-210. [doi: 10.1002/jclp.20223] [Medline: 16342287]
25. Williams JM, Swales M. The use of mindfulness-based approaches for suicidal patients. Arch Suicide Res 2004;8(4):315-329. [doi: 10.1080/13811110490476671] [Medline: 16081399]
26. Leitner M, Barr W, Hobby L. Effectiveness of interventions to prevent suicide and suicidal behaviour: A systematic review. Edinburgh, UK: Scottish Government Social Research; 2008.
27. McDaid D, Kennelly B. An economic perspective on suicide across the five continents. In: Oxford Textbook of Suicidology and Suicide Prevention: A Global Perspective. Oxford: Oxford University Press; 2009:359-367.
28. Byford S, Knapp M, Greenshields J, Ukoumunne OC, Jones V, Thompson S, POPMACT Group. Cost-effectiveness of brief cognitive behaviour therapy versus treatment as usual in recurrent deliberate self-harm: a decision-making approach. Psychol Med 2003 Aug;33(6):977-986. [Medline: 12946082]
29. Brazier J, Tumur I, Holmes M, Ferriter M, Parry G, Dent-Brown K, et al. Psychological therapies including dialectical behaviour therapy for borderline personality disorder: a systematic review and preliminary economic evaluation. Health Technol Assess 2006 Sep;10(35):iii, ix-xii, 1 [FREE Full text] [Medline: 16959171]
30. Donohue JM, Pincus HA. Reducing the societal burden of depression: a review of economic costs, quality of care and effects of treatment. Pharmacoeconomics 2007;25(1):7-24. [Medline: 17192115]
31. Sobocki P, Jönsson B, Angst J, Rehnberg C. Cost of depression in Europe. J Ment Health Policy Econ 2006 Jun;9(2):87-98. [Medline: 17007486]
32. Beck AT, Steer RA. Manual for the Beck Scale for Suicide Ideation. San Antonio: The Psychological Corporation; 1991.
33. van der Does AJW. The Dutch version of the Beck Depression Inventory - second edition. Lisse: Swets Zeitlinger; 2002.
34. Chang NA, Stanley B, Brown GK, Cunningham A. Treating the suicidal patient: Cognitive therapydialectical behaviour therapy. In: O'Connor R, Platt S, Gordon J, editors. International Handbook of Suicide Prevention: Research, Policy and Practice. Chichester: Wiley; 2011:369-382.
35. Brown GK, Ten Have T, Henriques GR, Xie SX, Hollander JE, Beck AT. Cognitive therapy for the prevention of suicide attempts: a randomized controlled trial. JAMA 2005 Aug 3;294(5):563-570. [doi: 10.1001/jama.294.5.563] [Medline: 16077050]
36. Kerkhof AJFM, van Spijker BAJ. Worryingrumination as proximal risk factors for suicidal behaviour. In: O'Connor R, Platt S, Gordon J, editors. International Handbook of Suicide Prevention: Research, Policy and Practice: Wiley:199-209.
37. 113Online. URL: https://www.113online.nl/zelfmoord-preventie-aanbod/zelfhulpcursus [accessed 2011-07-15] [WebCite Cache ID 6AT3UkRwr]
38. Mokkenstorm JK, Huisman A, Kerkhof AJ. [Suicide prevention via the internet and the telephone: 113Online]. Tijdschr Psychiatr 2012;54(4):341-348 [FREE Full text] [Medline: 22508352]
39. Brown GK. A Review of Suicide Assessment Measures for Intervention Research with Adults and Older Adults. Pennsylvania: University of Pennsylvania; 2000.
40. Hakkaart-van Roijen L, van Straten A, Donker M, Tiemens B. Manual Trimbos/iMTA Questionnaire for Costs Associated with Psychiatric Illness (TiC-P). Rotterdam: Institute for Medical Technology Assessment, Erasmus MC Rotterdam; 2002.
41. Hakkaart- van Roijen L, Tan SS, Bouwmans CAM. Manual for Cost Research. Methods and Standard Cost Prizes for Economic Evaluations in Health Care (in Dutch). Rotterdam: Institute for Medical Technology Assessment, Erasmus University; 2010.
42. College voor Zorgverzekeringen. Pharmacotherapeutic Compass (in Dutch). Diemen: College voor Zorgverzekeringen; 2009.
43. The Organisation for Economic Co-operation and Development (OECD). URL: http://www.oecd.org [accessed 2011-07-06] [WebCite Cache ID 6AT5l1Eiy]
44. ten Have M, van Dorsselaer S, Tuithof M, de Graaf R. New data on suicidality in the general population. Results from the Netherlands Mental Health Survey and Incidence Study-2 (NEMESIS-2). Utrecht: Trimbos-instituut; 2011.
45. StatLine Database.: Statistics Netherlands URL: http://statline.cbs.nl/statweb/ [accessed 2011-06-08] [WebCite Cache ID 6AT6tWcgK]
46. 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]
47. Langley PC. The November 1995 revised Australian guidelines for the economic evaluation of pharmaceuticals. Pharmacoeconomics 1996 Apr;9(4):341-352. [Medline: 10160108]
48. Siegel JE, Torrance GW, Russell LB, Luce BR, Weinstein MC, Gold MR. Guidelines for pharmacoeconomic studies. Recommendations from the panel on cost effectiveness in health and medicine. Panel on cost Effectiveness in Health and Medicine. Pharmacoeconomics 1997 Feb;11(2):159-168. [Medline: 10172935]
49. Torrance GW, Blaker D, Detsky A, Kennedy W, Schubert F, Menon D, et al. Canadian guidelines for economic evaluation of pharmaceuticals. Canadian Collaborative Workshop for Pharmacoeconomics. Pharmacoeconomics 1996 Jun;9(6):535-559. [Medline: 10160481]
50. van Hout BA, Al MJ, Gordon GS, Rutten FF. Costs, effects and C/E-ratios alongside a clinical trial. Health Econ 1994;3(5):309-319. [Medline: 7827647]

51. Gerhards SA, de Graaf LE, Jacobs LE, Severens JL, Huibers MJ, Arntz A, et al. Economic evaluation of online computerised cognitive-behavioural therapy without support for depression in primary care: randomised trial. Br J Psychiatry 2010 Apr;196(4):310-318 [FREE Full text] [doi: 10.1192/bjp.bp.109.065748] [Medline: 20357309]
52. Hollinghurst S, Peters TJ, Kaur S, Wiles N, Lewis G, Kessler D. Cost-effectiveness of therapist-delivered online cognitive-behavioural therapy for depression: randomised controlled trial. Br J Psychiatry 2010 Oct;197(4):297-304 [FREE Full text] [doi: 10.1192/bjp.bp.109.073080] [Medline: 20884953]
53. van den Brink M, van den Hout WB, Stiggelbout AM, van de Velde CJ, Kievit J. Cost measurement in economic evaluations of health care: whom to ask? Med Care 2004 Aug;42(8):740-746. [Medline: 15258475]
54. Wolinsky FD, Miller TR, An H, Geweke JF, Wallace RB, Wright KB, et al. Hospital episodes and physician visits: the concordance between self-reports and medicare claims. Med Care 2007 Apr;45(4):300-307. [doi: 10.1097/01.mlr.0000254576.26353.09] [Medline: 17496713]
55. Zuvekas SH, Olin GL. Validating household reports of health care use in the medical expenditure panel survey. Health Serv Res 2009 Oct;44(5 Pt 1):1679-1700. [doi: 10.1111/j.1475-6773.2009.00995.x] [Medline: 19619249]
56. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

BDI: Beck Depression Inventory
BSS: Beck Scale for Suicide Ideation
CAU: care as usual
CBT: cognitive behavior therapy
DBT: dialectical behavior therapy
ICER: incremental cost-effectiveness ratio
MACT: manual-assisted cognitive behavior therapy
MBCT: mindfulness-based cognitive therapy
PPP: purchasing power parity
WTP: willingness to pay

Edited by G Eysenbach; submitted 18.10.11; peer-reviewed by C Knapp; comments to author 15.01.12; revised version received 08.05.12; accepted 24.05.12; published 26.10.12.

Please cite as:
van Spijker BA, Majo MC, Smit F, van Straten A, Kerkhof AJ
Reducing Suicidal Ideation: Cost-Effectiveness Analysis of a Randomized Controlled Trial of Unguided Web-Based Self-help
J Med Internet Res 2012;14(5):e141
URL: http://www.jmir.org/2012/5/e141/
doi:10.2196/jmir.1966
PMID:23103835

Original Paper

FluBreaks: Early Epidemic Detection from Google Flu Trends

Fahad Pervaiz[1], BS; Mansoor Pervaiz[1], BS; Nabeel Abdur Rehman[1], BS(Current); Umar Saif[1], BS, PhD, Postdoc

[1] School of Science and Engineering, Computer Science Department, Lahore University of Management Sciences, Lahore, Pakistan

Corresponding Author:
Fahad Pervaiz, BS
School of Science and Engineering
Computer Science Department
Lahore University of Management Sciences
Opposite Sector-U, D.H.A.
Lahore, 54792
Pakistan
Phone: 92 333 443 8347
Fax: 92 42 3560 8303
Email: fahadp@lums.edu.pk

Abstract

Background: The Google Flu Trends service was launched in 2008 to track changes in the volume of online search queries related to flu-like symptoms. Over the last few years, the trend data produced by this service has shown a consistent relationship with the actual number of flu reports collected by the US Centers for Disease Control and Prevention (CDC), often identifying increases in flu cases weeks in advance of CDC records. However, contrary to popular belief, Google Flu Trends is not an early epidemic detection system. Instead, it is designed as a baseline indicator of the trend, or changes, in the number of disease cases.

Objective: To evaluate whether these trends can be used as a basis for an early warning system for epidemics.

Methods: We present the first detailed algorithmic analysis of how Google Flu Trends can be used as a basis for building a fully automated system for early warning of epidemics in advance of methods used by the CDC. Based on our work, we present a novel early epidemic detection system, called FluBreaks (dritte.org/flubreaks), based on Google Flu Trends data. We compared the accuracy and practicality of three types of algorithms: normal distribution algorithms, Poisson distribution algorithms, and negative binomial distribution algorithms. We explored the relative merits of these methods, and related our findings to changes in Internet penetration and population size for the regions in Google Flu Trends providing data.

Results: Across our performance metrics of percentage true-positives (RTP), percentage false-positives (RFP), percentage overlap (OT), and percentage early alarms (EA), Poisson- and negative binomial-based algorithms performed better in all except RFP. Poisson-based algorithms had average values of 99%, 28%, 71%, and 76% for RTP, RFP, OT, and EA, respectively, whereas negative binomial-based algorithms had average values of 97.8%, 17.8%, 60%, and 55% for RTP, RFP, OT, and EA, respectively. Moreover, the EA was also affected by the region's population size. Regions with larger populations (regions 4 and 6) had higher values of EA than region 10 (which had the smallest population) for negative binomial- and Poisson-based algorithms. The difference was 12.5% and 13.5% on average in negative binomial- and Poisson-based algorithms, respectively.

Conclusions: We present the first detailed comparative analysis of popular early epidemic detection algorithms on Google Flu Trends data. We note that realizing this opportunity requires moving beyond the cumulative sum and historical limits method-based normal distribution approaches, traditionally employed by the CDC, to negative binomial- and Poisson-based algorithms to deal with potentially noisy search query data from regions with varying population and Internet penetrations. Based on our work, we have developed FluBreaks, an early warning system for flu epidemics using Google Flu Trends.

(J Med Internet Res 2012;14(5):e125) doi:10.2196/jmir.2102

KEYWORDS

Influenza; public health; epidemics; statistical distributions; early response

Introduction

Infodemiology introduced the use of nontraditional data sources for the detection of disease trends and outbreaks [1]. These data sources include search queries, social media, Web articles, and blogs posts, which are now being used for real-time disease surveillance [1-4]. In terms of search queries as a source, interest in using these to predict epidemics has been growing recently [5-8]. Most notably, the Google Flu Trends [9] service was launched in 2008 as a way to track changes in the volume of online search queries related to flu-like symptoms [5]. Google Flu Trends provides search query trend data that are real-time and reported on a daily basis, and have been shown to predict the actual cases of a disease such as flu at least 2 weeks ahead of the US Centers for Disease Control and Prevention (CDC).

In the absence of other real-time disease surveillance mechanisms, services such as Google Flu Trends are vitally important for the early detection of epidemics. Existing research on using Google Flu Trends for epidemic detection has focused on addressing this need by collecting data related to the volume of queries for disease symptoms. This work demonstrates that Google search query trends closely follow the actual disease cases reported by the CDC. While these results provide strong support for the potential use of Google Flu Trends data as a basis for an early warning system for epidemics, existing research needs to be advanced along two essential directions to realize this opportunity. First, there is a need to rigorously explore and evolve algorithms for higher-level inference from the Google Flu Trends data that can generate alerts at early stages of epidemics. In particular, the ability of existing approaches to collect raw search volume data needs to be supplemented with computational intelligence to translate these data into actionable information. Second, there is also a need to develop a more detailed appreciation of how changes in population size and Internet penetration affect the ability of a system based on Google Flu Trends data to provide accurate and actionable information.

In this study, we aimed to provide new insights related to these opportunities. We built upon Google Flu Trends data and compared the accuracy and practicality of widely used algorithms for early epidemic detection. These algorithms are classified into three categories based on the type of data distribution they expect. The classifications in question are normal distribution algorithms, Poisson distribution algorithms, and negative binomial distribution algorithms. For normal distribution algorithms, we used cumulative sum (CUSUM) [10-12], the historical limits method (HLM) [10,13], and historical CUSUM (HCusum) [14,15]. For Poisson distribution algorithms, we used Poisson outbreak detection (POD) [16], SaTScan [17], and Poisson CUSUM (PSC) [18,19]. For negative binomial distribution algorithms, we used negative binomial CUSUM (NBC) [20,21] and historical NBC. Some of these algorithms have already been compared on Ross River disease data [22]. Our choice of some of the algorithms (CUSUM, HLM, POD, SaTScan, and NBC) and parameters was also based on this work [22]. However, our work performed the comparison on Google Flu Trends data. We quantitatively compared the accuracy, specificity, and sensitivity of these algorithms, as well as the impact of leveraging information in baseline training periods, seasonal changes, population sizes, and Internet penetrations, on their suitability for detecting epidemics.

Methods

Data Sources

Google Flu Trends

Traditional disease surveillance networks such as the CDC take up to 2 weeks to collect, process, and report disease cases registered at health centers [23].

Google Flu Trends [9], on the other hand, provides near real-time data on disease cases by benefiting from the likelihood that many patients with flu symptoms search online for their symptoms and remedies before visiting a doctor.

Google Flu Trends compares the popularity of the 50 million most common Google search queries in the United States with flu-like illness rates reported by the CDC’s national surveillance program. The Flu Trends data are derived from a pool of 45 search terms that relate to symptoms, remedies, and complications of flu and generate a trend that closely correlates with CDC data on influenza-like illnesses.

In our experiments, we used Google Flu Trends data from the 9 years between 2003 and 2011.

CDC‘s Outpatient Illness Surveillance

Information on patient visits to health care providers in the United States for influenza-like illness is collected through the Outpatient Influenza-like Illness Surveillance Network (ILINet). ILINet consists of more than 3000 health care providers in all 50 states, reporting over 25 million patient visits each year. Each week, approximately 1800 outpatient care sites around the United States report data to the CDC on the total number of patients seen and the number of those patients with influenza-like illnesses. For this system, an influenza-like illness is defined as fever (temperature of 100°F [37.8°C] or greater) and a cough or a sore throat in the absence of a known cause other than influenza. Sites with electronic records use an equivalent definition as determined by state public health authorities. The percentage of patient visits to health care providers for influenza-like illnesses reported each week is weighted on the basis of a state’s population. This percentage is compared each week with the national baseline of 2.5%. The baseline is the mean percentage of patient visits for influenza-like illnesses during noninfluenza weeks for the previous three seasons plus 2 standard deviations [24].

In our experiments, much like Google Flu Trends data, we used CDC influenza-like illness data from the 9 years between 2003 and 2011. Though the CDC has missing data in the nonflu season between 2009 and 2010, we believe this had a minimal effect on our quantitative comparison.

Outbreak

For determining periods of outbreaks, the starting points in the time and the duration, we consulted two epidemiologists from different institutes. The first was from the Institute of Public Health, Lahore, Pakistan (responsible for informing the provincial health ministry about disease outbreaks) and the second was from Quaid-e-Azam Medical College, Bahawalpur, Pakistan. These original outbreaks were marked on CDC influenza-like illness data [23].

Outbreak Detection Algorithms

The early epidemic detection algorithms that we have used are divided into three categories, based on the expected distribution in data: (1) normal distribution algorithms: these expect normal distribution in the data, (2) Poisson distribution algorithms: these expect a Poisson distribution, and (3) negative binomial distribution algorithms: these expect a negative binomial distribution.

Normal Distribution Algorithms

The algorithms classified in this category are the Early Aberration Reporting System (EARS) algorithm (CUSUM), HLM, and HCusum.

Early Aberration Reporting System Algorithms

EARS was developed and used by the CDC. EARS comprises three syndromic surveillance early event detection methods called C1, C2, and C3 [11], which are the Shewhart variations of the CUSUM method. These methods use a moving average and standard deviation to standardize the number of occurrences in the historical data. In our analysis, C1 used the 4 weeks prior to the current week of observation for calculating the average and standard deviation. The value of average and standard deviation is used to determine the C1 score (Figure 1, parts a, b, and c). C2 is similar to C1 but used the 4 weeks after a 1-week lag. It means that it used week 2 to week 5 for calculating the average and standard deviation (Figure 1, parts d, e, and f). C3 used the C2 score for the previous 3 weeks to calculate the C3 score, as shown in Figure 1 (part g).

The C1, C2, and C3 EARS algorithms require a baseline (training period) and cut-off (threshold) as parameters. In our experiments, we used both 4 weeks and 8 weeks as the baseline. A shorter training period (baseline) has been shown to insulate CUSUM from seasonal changes [15]. For each of these baseline periods, we compared the algorithms for four cut-off values: 2, 4, 6, and 8. This means that we declared an outbreak if the observed value exceeded the mean value more than a standard deviation of 3, 5, 7, and 9, respectively. A higher cut-off point makes the algorithm less responsive to transient changes in the rate of occurrence of disease cases. In our analysis, we excluded C1 with cut-offs of 6 and 8 for both baselines because it rarely raised an outbreak alarm for the 9 years of data.

Since CUSUM uses mean and standard deviation for raising alarms, it is best for outbreaks with respect to normal distribution of data. It means that the algorithm is very sensitive to a sudden rise, which generates an early alarm. In addition, it expects a constant rise in data for an outbreak to continue because the start of a rise becomes part of historical data, consequently also raising the mean and standard deviation for the algorithm.

Figure 1. Early Aberration Reporting System (EARS) algorithm equations. C_1 = cumulative sum (CUSUM) score of C1 algorithm, C_2 = CUSUM score of C2 algorithm, C_3 = CUSUM score of C3 algorithm, sigma = standard deviation, X-bar = mean number of cases, X_n = number of cases in current time interval. Subscripts refer to a specific variable being linked to either of the three algorithms.

$$C_1(n) = \frac{X_n - \bar{X}_1(n)}{\sigma_1(n)} \qquad (a)$$

$$\bar{X}_1(n) = \frac{1}{8}\sum_{i=n-1}^{n-8} X_i \qquad (b)$$

$$\sigma_1{}^2\,(n) = \frac{1}{7}\sum_{i=n-1}^{n-8} [X_i - \bar{X}_1(n)]^2 \qquad (c)$$

$$C_2(n) = \frac{X_n - \bar{X}_3(n)}{\sigma_3(n)} \qquad (d)$$

$$\bar{X}_3(n) = \frac{1}{8}\sum_{i=n-2}^{n-9} X_i \qquad (e)$$

$$\sigma_3{}^2\,(n) = \frac{1}{7}\sum_{i=n-2}^{n-9} [X_i - \bar{X}_3(n)]^2 \qquad (f)$$

$$C_3(n) = \sum_{i=n}^{n-2} max[0, C_2(i) - 1] \qquad (g)$$

Historical Limits Method

The CUSUM methods used in EARS do not account for seasonality by design; however, the HLM incorporates historical data. In HLM an outbreak is signaled when the identity in Figure 2 is true.

In the HLM, the system determines the expected value of a week by (1) using 3 consecutive weeks for every year in the historical data, which is the current week, the preceding week, and the subsequent week (entitled HLM-3), and (2) using 5 consecutive weeks for every year, in the past years: the current week, the preceding 2 weeks, and the subsequent 2 weeks (HLM-5) in the historical data (Figure 3).

The above two variations (HLM-3, which uses 15 baseline points, and HLM-5, which uses 25 baseline points) are recommended by Pelecanos et al [22].

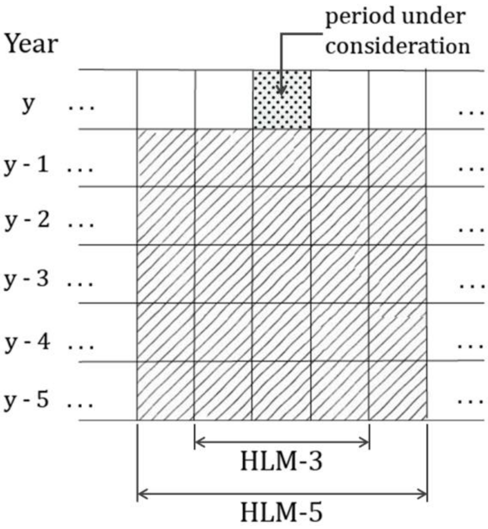
We used both HLM-3 and HLM-5, in which the training period comprised 5 years, starting from 2003 and ending in 2008. For determining outbreaks within the training period, we removed 1 year at a time from the timeline between 2003 and 2008 (both years inclusive). Then we assumed that the remaining years were consecutive and determined outbreaks during the omitted year by using the remaining 4 years. This process was repeated for every year of the training period.

Just like EARS, HLM runs on the mean and standard deviation of the data. Therefore, the definition of outbreak is to expect a normal distribution and to mark any outlier, according to normal distribution, as an outbreak.

Figure 2. Historical limits method (HLM) equation. Sigma = standard deviation, X = number of reported cases in the current period, X-bar = mean.

$$\frac{X}{\bar{X}} > 1 + \frac{2\sigma}{\bar{X}}$$

Figure 3. Historical data of the historical limits methods (HLM). HLM-3 = 3 consecutive weeks in the historical data, HLM-5 = 5 consecutive weeks in the historical data.



Historical CUSUM

HCusum is a seasonally adjusted CUSUM [15]. It considers the same period during previous years before creating an alert. This ignores the regular trend of rises in count data and signals only when an anomaly happens. Hence, the baseline data of our calculation was the patient count for the same week number during the preceding 5 years. The mean (X-bar) gives a reference value of what an expected count will be. Sigma gives an insight into how much variation there has been in the values used to calculate the expected value [14] (Figure 4, parts a and b).

An outbreak is declared if the identity in (c) is true.

Figure 4. Historical cumulative sum (HCUSUM) equations. Sigma = standard deviation, X_n = number of cases in current time interval, X-bar = mean number of cases. N = 5, as the baseline period is the preceding 5 years.

$$\bar{X} = \frac{1}{N}\sum_{k=1}^{N} X_{n-k} \qquad (a)$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{k=1}^{N} (X_{n-k} - \bar{X})^2} \qquad (b)$$

$$\frac{X_n - \bar{X}}{2} > \sigma \qquad (c)$$

Poisson Distribution Algorithms

The algorithms classified in this category are POD, SaTScan, and PSC.

Poisson Outbreak Detection Method

The POD method assumes that the number of cases follow a Poisson distribution. The POD method [16] uses 10 years of historical data for calculating the incidence rate of a disease. This 10-year period is used for accommodating high variability and skewed distribution of the seasonal incidence rates. To accommodate variability in the population of the various regions, if the population of a certain region is less than 2500, the crude incidence rate is used for determining outbreaks. If the population size is larger than 2500, then the crude incidence rate is used if either the maximum number of notifications is less than 5 or the crude incidence rate and trimmed incidence rate differ by less than 20%. (Trimmed incidence rate is calculated by omitting the years with the maximum or minimum number of cases). Otherwise, the median incidence rate is used. An outbreak is considered if the chance of the actual number of cases occurring was less than 1%. For POD, a year is divided into seasons (winter, spring, summer, and autumn) and the IRs are calculated for each season rather than the whole year. This is how POD caters for seasonality. Since it is Poisson-based algorithm, it is the best fit when the outbreak data’s variance to mean ratio (VMR) equals 1. This value of VMR implies that the data follow a Poisson distribution.

We followed certain suggestions from Pelecanos et al [22] and increased the percentage chance from 1% to 5%. This is because we did not have 10 years’ worth of historical data to train the system. Therefore, this change in percentage helped reduce the sensitivity of the algorithm. We used the first 5 years as a training period and then added every subsequent year for further outbreak detection.

Purely Temporal SaTScan

The SaTScan algorithm can be used for spatial analysis, temporal analysis, and spatiotemporal analysis. We used only the temporal analysis for our outbreak detection, since spatial mapping is already fixed to a CDC-defined region. We used a Poisson permutation, which works best for data following a Poisson distribution. This is the case when the data’s VMR is equal to 1.

Temporal SaTScan creates 1-dimensional clusters by sliding and scaling a window within an interval of 60 days. We relied on the Poisson permutation to determine the clusters with the highest likelihood ratio.

The equation (Figure 5) calculates the log likelihood ratio for the selection of the cluster.

Once we had the best cluster within an interval, the algorithm calculated the *P* value of the cluster using Monte Carlo testing. A *P* value less than .001 determines, with high significance, that there is an outbreak in the cluster.

SaTScan does not accommodate seasonality. Therefore, to adjust SaTScan for seasonality we scaled the population size of the region under analysis on a weekly basis. SaTScan uses population size as one of the parameters, so for every week the population is scaled. The factor for scaling the population size is dependent on the incidence rate for each week and the annual population: *populationi = annual population * (incidence rate for week i / total incidence rate)*, where *populationi* is the scaled population for a particular week, *annual population* is the population of the year under observation, the *incidence rate for week i* is the average incidence rate of a particular week over the past years, and the *total incidence rate* is the average of incidence rates for all weeks through the year.

Moreover, as CDC and Google data are reported on a weekly cycle, we parameterized SaTScan on a weekly time unit. We set the *P* value cut-off at .001 (to avoid detecting smaller clusters in response to seasonal changes) and set the number of iterations to 15 (since our data comprised 8 flu seasons). To detect a new cluster in each iteration, we set the iterative scan to adjust for more likely clusters. We did not change the default value of the maximum Monte Carlo replication (999).

Figure 5. SaTScan equation. C = total number of cases, C_z = observed number of cases in window z, LLR = likelihood ratio, n_z = expected number of cases or population in window z.

$$LLR(z) = \left(\frac{c_z}{n_z}\right)^{c_z}\left(\frac{C - c_z}{C - n_z}\right)^{C - c_z}$$

Poisson CUSUM

PSC is an algorithm that detects the anomalies efficiently in data that follow a Poisson distribution [18,19]. It tests a null hypothesis that the current value is in control and the alternative hypothesis that the value is out of control. As a Poisson distribution can be defined by only one parameter (mean), the reference values for both hypotheses are taken as the mean. The reference for the null hypothesis is the average value (X_a-bar) of the data in the baseline window. The baseline window is the period of the past 7 weeks from the current week of analysis, with a 1-week guard band in between. For the alternative hypothesis, the mean value (X_d-bar) is the sum of the average and 2 times the standard deviation of the baseline period. (X_a-bar) and (X_d-bar) are used in calculating k^+ of the PSC, as shown in Figure 6 (part a), which also shows the equations used to calculate CUSUM (parts b and c).

An outbreak is signaled when the computed CUSUM score is higher than the threshold *h*. The threshold *h* is equal to *t* * *k* [19]. We did our analysis with *t* = 1 and *t* = 1.5.

Figure 6. Poisson cumulative sum (CUSUM) equations. k = reference value, S_n = CUSUM score, X_n = number of cases in current time interval, X-bar_a = null hypotheses mean, X-bar_d = alternative hypotheses mean, + superscript refers to values always being positive.

$$k^+ = \frac{\bar{X}_d - \bar{X}_a}{\ln(\bar{X}_d) - \ln(\bar{X}_a)} \qquad (a)$$

$$S_0 = 0 \qquad (b)$$

$$S_n^+ = max(0, S_{n-1}^+ + X_n - k^+) \qquad (c)$$

Negative Binomial Distribution Algorithms

This category comprises NBC and historical NBC.

Negative Binomial CUSUM

Static Thresholds

We selected NBC [20,21] because of its property of handling inaccuracies due to overdispersion in data. Overdispersion in data causes VMR to be higher than 1. This generally happens during seasonal increase. Two parameters, (*r*) and (c_0), are used to describe negative binomial distribution. Equations in Figure 7 (parts a and b) are used to determine the value of these parameters based on mean (X-bar) and variance (sigma^2), which are derived from the baseline period. The decision interval is given by equations in Figure 7 (parts c, d, and e) through looking for changes in c from an in-control c_0 to an out-of-control c_1, where $c_1 > c_0$ [20].

The out-of-control level c_1 is determined by adding 2 times the standard deviation of the baseline period to the mean of the baseline. We kept a baseline interval of 7 weeks and a guard band of 1 week. The guard band prevents the most recent data from being included in baseline calculations. Therefore, the baseline period and current week will have a gap of 1 week as a guard band. The CUSUM score is compared with the threshold value *h*. An outbreak is declared if the CUSUM score (S_n^+) > *h*. The results were calculated using static cut-off (threshold) values of 8 and 15 [22].

Figure 7. Negative binomial cumulative sum (CUSUM) equations. k = reference value, (r,c) = parameters of negative binomial distribution, S_n = CUSUM score, sigma =standard deviation, X_n = number of cases in current time interval, X-bar = mean number of cases, + superscript refers to values always being positive.

$$C_0 = \bar{X}/(\sigma^2 - \bar{X}) \qquad (a)$$

$$r = \bar{X}^2/(\sigma^2 - \bar{X}) \qquad (b)$$

$$S_0 = 0 \qquad (c)$$

$$S_n^+ = max(0, S_{n-1}^+ + X_n - k^+) \qquad (d)$$

$$k^+ = r.\frac{\ln[c_0(1 + c_1)/c_1(1 + c_0)]}{\ln[(1 + c_0)/(1 + c_1)]} \qquad (e)$$

Variable Thresholds

NBC with static threshold, although it catches the longevity of the outbreak, is sensitive in raising early alarms. To cater for this sensitivity we introduce variable thresholds for NBC. A new parameter, *hv*, is calculated, which is used as the threshold for the CUSUM score. The calculation of the rest of the parameters is based on equations in Figure 7. The variable threshold *hv* is calculated by the equation *hv* = *t* * *k*, where *t* is a constant. We performed our analysis with values of *t* of both 1 and 1.5. Involving *k* in the threshold calculation changes the cut-off values with the variation in the count data of the baseline window. This reduces the sensitivity of the CUSUM [19].

Historical Negative Binomial CUSUM

Historical NBC is a seasonally adjusted negative binomial CUSUM [20,21]. It calculates c_0, r, and k^+using equations in Figure 7 (parts a, b, and e, respectively). The baseline data are patient case counts of the current period during the past 5 years. The calculation of mean (X-bar) and variance (sigma^2) is based on the given baseline period of the past years. The CUSUM score is given by equation shown in Figure 8.

An outbreak is declared if $S_n^+ > h$, where h is the maximum limit for the results to remain in a tolerable state. We used h = 15 [18,22] for our analysis. As an outbreak can exist on the very first calculation, generally a shorter period of 5 years is used as a baseline [18,19,22].

Figure 8. Historical negative binomial cumulative sum (CUSUM) equation. k = reference value, S_n = CUSUM score, X_n = the case count of the current week, + superscript refers to values always being positive.

$$S_n^+ = max(0, X_n - k^+)$$

Performance Metrics

To understand how Google Flu Trends data can be used to build an early epidemic detection system, we compared the results of 24 variants of 8 base algorithms (from three categories of algorithms) across three regions in the United States. To the best of our knowledge, this paper presents the first comparative analysis of epidemic detection algorithms for Google Flu Trends data.

For our base algorithms, we used EARS CUSUM, HCusum, HLM, POD, SaTScan, PSC, NBC, and HNBC. The characteristics of these algorithms afford a degree of diversity in our analysis: EARS CUSUM and NBC were designed for rapid detection of outbreaks; HCusum, HNBC, HLM, and POD incorporate seasonal changes but require a substantial training period; and SaTScan requires minimal training and offers flexibility in detecting statistically significant disease clusters.

We chose the target regions, as divided by the CDC, to compare the sensitivity of the various algorithms to population size and Internet penetration. Table 1 shows the population size and Internet penetration across HHS regions used in our experiments. Figure 9 maps the regions to US states. We calculated the population of each region by adding the populations of the states in that region. The population in 2009 was used for this calculation [25]. Internet use was taken from a report of the Current Population Survey of Internet Use 2009, published by the National Telecommunications and Information Administration [26] and in the 2009 census published by the US Census Bureau [27].

For our comparison with respect to population size, we focused on region 4 (with the largest population) and region 10 (smallest population). For evaluating the impact of Internet penetration, we focused on region 6 (lowest Internet penetration) and region 10 (highest Internet penetration). Results from region 10 are of particular interest, since it has the lowest population and highest Internet penetration. We expect that the results from region 10 could serve as a benchmark of how accurately Google Flu Trends data can be used as a basis for detecting epidemics. Furthermore, the weather in regions 4 and 6 was similar but very different from that in region 10.

In our analysis, we evaluated each algorithm by comparing its results using Google Flu Trends data with the disease cases reported by CDC. We compared the performance of the algorithms on the following key metrics.

Percentage True-Positives

Percentage true-positives (RTP) measures the percentage of time an epidemic signaled in the CDC data is also detected by the target algorithm on Google Flu Trends data. This percentage is calculated by the number of outbreak intervals when the signal was raised divided by the total number of outbreak intervals, with the result multiplied by 100.

Percentage False-Positives

Percentage false-positives (RFP) measures the percentage of time an epidemic not signaled in the CDC data is detected as an epidemic by the target algorithm on Google Flu Trends data. This percentage is calculated by the number of nonoutbreak weeks when a signal was raised divided by the total number of weeks with no outbreak, with the result multiplied by 100.

Percentage Overlap Time

Percentage overlap (OT) measures the percentage of the time an epidemic detected by an algorithm overlaps with the epidemic signaled in CDC data. Any part of a signal that does not overlap with the original outbreak is not considered in OT.

Percentage Early Alarms

Percentage early alarms (EA) measures the percentage of time an algorithm raises an alarm on Google Flu Trends before it is signaled as an epidemic by the CDC data. The early alarm period is limited to the 2 weeks before the start of the original outbreak. Part of a signal starting before this 2-week time period is considered false-positive.

These four metrics capture different aspects of the detection algorithms. RTP measures the sensitivity of an algorithm to outbreaks. At the same time, an overly sensitive algorithm generates a higher number of RFPs.

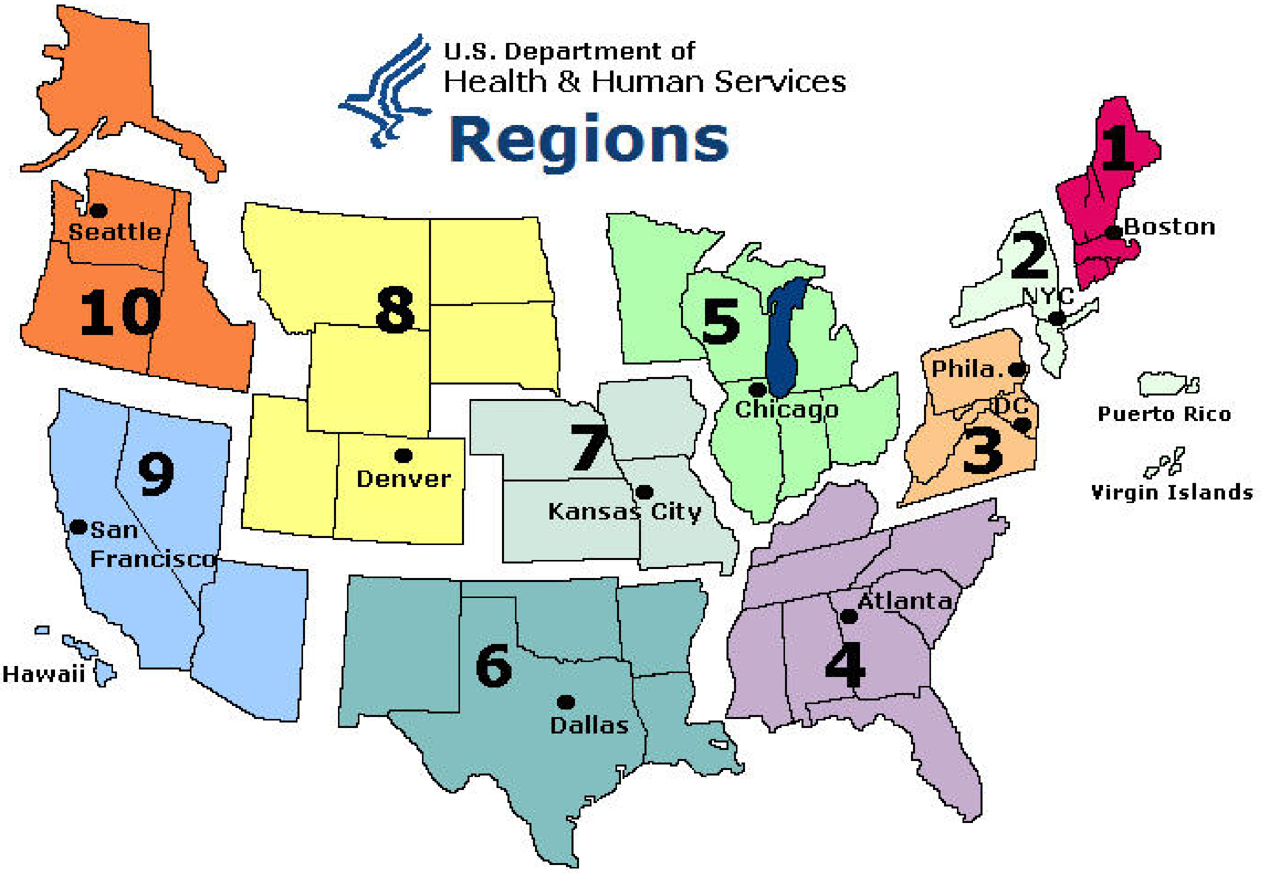
The average overlap time captures the stability of an algorithm to transient changes in the rate of disease cases. Algorithms that signal the entire period of an epidemic are more desirable than those that raise short, sporadic signals.

Finally, algorithms that signal an epidemic ahead of other algorithms are more suited for early epidemic detection. However, this metric must be viewed in conjunction with an algorithm's RFP to discount algorithms that generate spurious signals. For our analysis, we counted a signal as an early alarm if its fell within a 2-week window preceding the signal in the CDC data, so long as it was not a continuation of a previous alarm.

Table 1. Population and percentage Internet use by US Department of Health and Human Services (HHS) region.

HHS region	Population (2009 census)	% Internet use	States
1	14,412,684	74.07	CT, ME, MA, NH, RI, VT
2	28,224,114	70.20	NJ, NY
3	29,479,361	69.30	DE, DC, MD, PA, VA, WV
4	60,088,178	63.25	AL, FL, GA, KY, MS, NC, SC, TN
5	51,745,410	71.42	IL, IN, MI, MN, OH, WI
6	37,860,549	61.56	AR, LA, NM, OK, TX
7	31,840,178	71.68	IA, KS, MO, NE
8	20,802,785	72.13	CO, MT, ND, SD, UT, WY
9	46,453,010	67.95	AZ, CA, HI, NV
10	6,691,325	76.93	AK, ID, OR, WA

Figure 9. US Department of Health and Human Services regions.



Results

Figure 10, Figure 11, and Figure 12 compare all of the algorithms in our study on a 9-year timescale between 2003 and 2011. Details for these figures are presented in Multimedia Appendix 1, Multimedia Appendix 2, and Multimedia Appendix 3, which compare the algorithms according to our four comparison metrics: RTP, RFP, OT, and EA across our three target regions [12,13,22].

In each Multimedia Appendix there is a sorted column (overall position of algorithm). In this column the algorithms are sorted based on their median across the four performance metrics. We chose the median to cater for extreme values in the performance metrics.

Although we have divided the algorithms into three categories, namely Poisson, negative binomial, and normal distribution algorithms, another subcategory called historical algorithms surfaced during our analysis. This is a subset of both negative binomial and normal distribution categories, as it has algorithms in both. HNBC from the negative binomial and HLM, and

HCusum from the normal distribution showed a similar pattern of results across the four performance metrics. Therefore, for the remainder of the discussion, we will add the classification of historical algorithm and analyze its results independently.

In Table 2, for the first performance metric, RTP, all the categories had high average values (96.4%, 99.0%, and 98.8% for normal, NBC, and Poisson distribution algorithms, respectively), with the only exception being the historical algorithms (64%). Moreover, among the algorithms showing a high RTP percentage, there were no significant differences between the values.

In the second performance metric, RFP, values go the other way round, with the historical algorithms showing remarkably optimal values (average 3.3%, where lower is better), whereas normal, NBC, and Poisson distribution algorithms show percentages of 11.4%, 28.3%, and 17.5%, respectively. Clearly the historical algorithms and normal distribution algorithms led in this metric.

In the third metric, OT, negative binomial distribution algorithms led, with an OT of 71.3%, followed by Poisson distribution (60.3%), historical algorithms (30.8%), and normal distribution algorithms (16.4%). In this metric, NBC and Poisson distribution led by a major difference, ahead of historical and normal distribution algorithms.

In the fourth and last metric, EA, negative binomial, on average, led with an EA value of 75.8%, followed by Poisson distribution (55.1%), normal distribution (36.8%) and historical algorithms (22.3%).

For some performance metrics, certain categories did not perform consistently, and values of these categories varied over a large range. In normal distribution algorithms, the values of EA varied from 0% to 75%. In Poisson distribution algorithms, EA varied from 13% to 75%. Therefore, in these cases the average value of that particular metric could not be considered representative, and we needed to examine the algorithms (or variations of algorithms) for suitability.

When we looked at EA values in normal distribution algorithms, the C3 variations of EARS showed a high EA value for only one region. Otherwise, the next best values were barely in the optimal range. Moreover, the OT of C3 at best was 34, which is very low and made this algorithm not suitable.

In case of the EA values in Poisson distributions, the SaTScan algorithm pulled the average of Poisson distribution algorithms down in EA. Therefore, if we considered the average EA value of Poisson distribution algorithms without SaTScan, it actually rose from 55.1 to 66.7.

Overall, negative binomial and Poisson distribution algorithms performed much better than normal distribution algorithms. This is mainly because of the data distribution that these algorithms expect. The VMR of seasonal influenza-like illness data was greater than 1, most of the time (Figure 13). Therefore, the data followed a negative binomial distribution [28]. Moreover, the Poisson distribution was an approximation of the negative binomial distribution [29,30]. Therefore, the overall percentages of both Poisson-based and negative binomial-based algorithms turned out to be high.

Historical algorithms performed poorly because they considered data during the same period in past years to declare the outbreak. They did not consider the distribution of data during the current year. This made them robust in terms of false-positives, but the performance across other metrics lagged by a substantial difference.

Furthermore, to understand the impact of population variation and change in Internet penetration across regions, we picked the top two algorithms from the negative binomial distribution and Poisson distribution algorithms and applied them to all the regions (instead of just three). Table 3, Table 4, Table 5, and Table 6 present the results of the algorithms applied.

The result of this analysis showed that in regions of high Internet penetration the RFP and OT were high.

Table 2. Average percentages of various performance metrics for various categories of algorithms.

Metric	Normal	Negative binomial	Poisson	Historical
RTP[a]	96.4	99.0	98.8	64.0
RFP[b]	11.4	28.3	17.5	3.3
OT[c]	16.4	71.3	60.3	30.8
EA[d]	36.8	75.8	55.1	22.3

[a] Percentage true-positives.

[b] Percentage false-positives.

[c] Percentage overlap time.

[d] Percentage early alarms.

Table 3. Result of negative binomial cumulative sum (cut-off = 15), for all performance metrics across all Department of Health and Human Services (HHS) regions of the United States.

HHS region	RTP[a]	RFP[b]	OT[c]	EA[d]
1	100	45	98	87.5
2	100	40	85	77.7
3	100	40	88	87.5
4	100	30	81	88
5	100	40	95	87.5
6	100	40	76	88
7	100	40	95	87.5
8	87.5	50	83	75
9	90	40	71	80
10	100	40	82	71

[a] Percentage true-positives.

[b] Percentage false-positives.

[c] Percentage overlap time.

[d] Percentage early alarms.

Table 4. Result of negative binomial cumulative sum (threshold = 1 * k) for all performance metrics across all Department of Health and Human Services (HHS) regions of the United States.

HHS region	RTP[a]	RFP[b]	OT[c]	EA[d]
1	100	35	87	87.5
2	100	27	74	66.7
3	100	20	81	75
4	100	20	70	75
5	100	30	86	75
6	100	20	63	75
7	100	30	87	75
8	87.5	40	71	75
9	90	30	64	70
10	100	30	68	71

[a] Percentage true-positives.

[b] Percentage false-positives.

[c] Percentage overlap time.

[d] Percentage early alarms.

Table 5. Result of Poisson cumulative sum (threshold = 1 * k) for all performance metrics across all Department of Health and Human Services (HHS) regions of the United States.

HHS region	RTP[a]	RFP[b]	OT[c]	EA[d]
1	100	35	83	87.5
2	100	27	71	66.7
3	100	20	80	75
4	100	20	70	75
5	100	30	84	75
6	100	20	62	75
7	100	30	84	75
8	87.5	40	67	75
9	90	30	64	70
10	100	30	68	57

[a] Percentage true-positives.

[b] Percentage false-positives.

[c] Percentage overlap time.

[d] Percentage early alarms.

Table 6. Result of Poisson outbreak detection for all performance metrics across all Department of Health and Human Services (HHS) regions of the United States.

HHS region	RTP[a]	RFP[b]	OT[c]	EA[d]
1	100	35	77	33
2	100	20	70	40
3	100	30	69	50
4	100	20	58	75
5	100	40	72	50
6	100	20	50	75
7	100	30	72	75
8	87.5	30	74	75
9	90	20	57	40
10	100	20	68	57

[a] Percentage true-positives.

[b] Percentage false-positives.

[c] Percentage overlap time.

[d] Percentage early alarms.

Figure 10. US Department of Health and Human Services region 4. The x-axis plots the Google Flu Trends and Centers for Disease Control and Prevention (CDC) data. The horizontal bars indicate where each method detected an epidemic. Cut indicates the cut-off point (more is less sensitive) and b indicates baseline data (training window). The thick horizontal bars at the bottom show the actual outbreak. HCusum = historical cumulative sum, HLM = historical limits method, HNBC = historical negative binomial cumulative sum, ILI = influenza-like illnesses, k = reference value for threshold, NBC = negative binomial cumulative sum, POD = Poisson outbreak detection, PSC = Poisson cumulative sum.

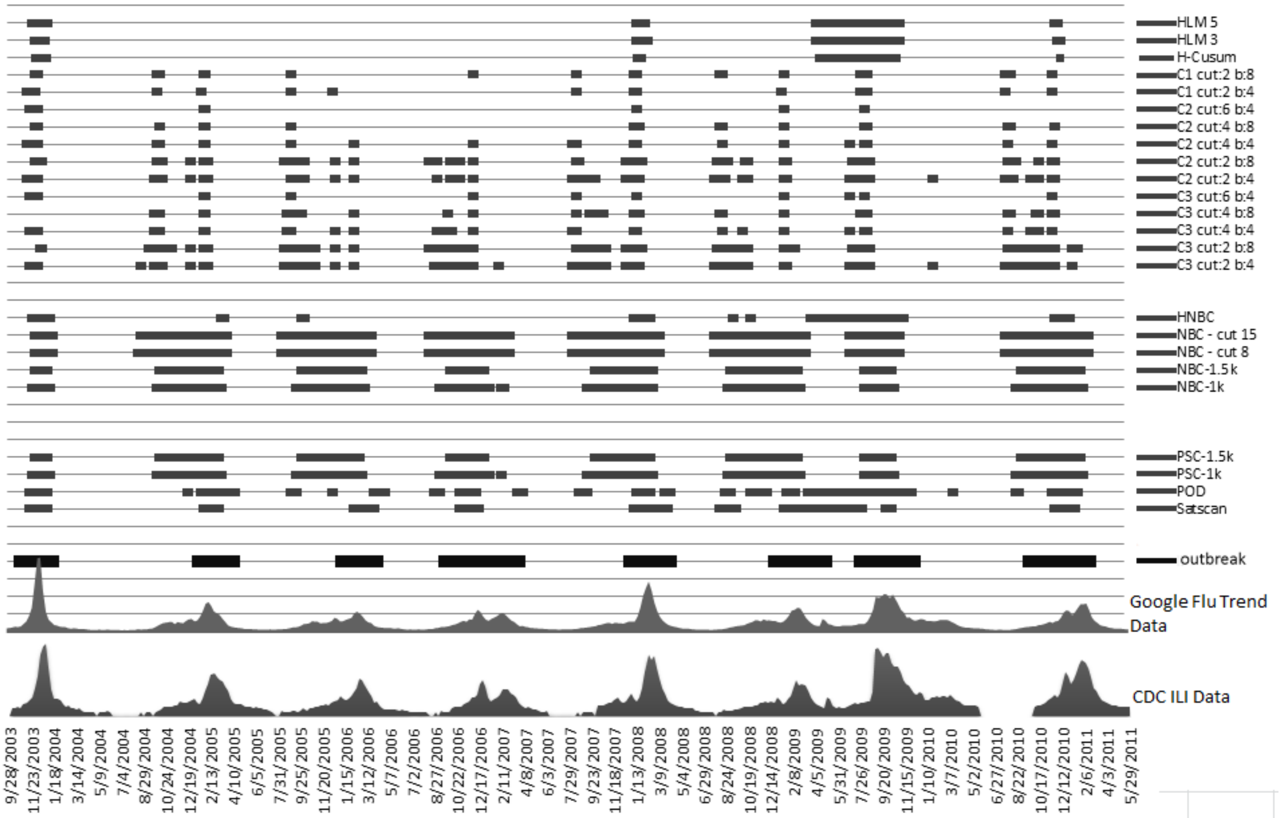


Figure 11. US Department of Health and Human Services region 6. The x-axis plots the Google Flu Trends and Centers for Disease Control and Prevention (CDC) data. The horizontal bars indicate where each method detected an epidemic. Cut indicates cut-off point (more is less sensitive) and b indicates baseline data (training window). The thick horizontal bar at the bottom shows the actual outbreak. HCusum = historical cumulative sum, HLM = historical limits method, HNBC = historical negative binomial cumulative sum, ILI = influenza-like illnesses, k = reference value for threshold, NBC = negative binomial cumulative sum, POD = Poisson outbreak detection, PSC = Poisson cumulative sum.

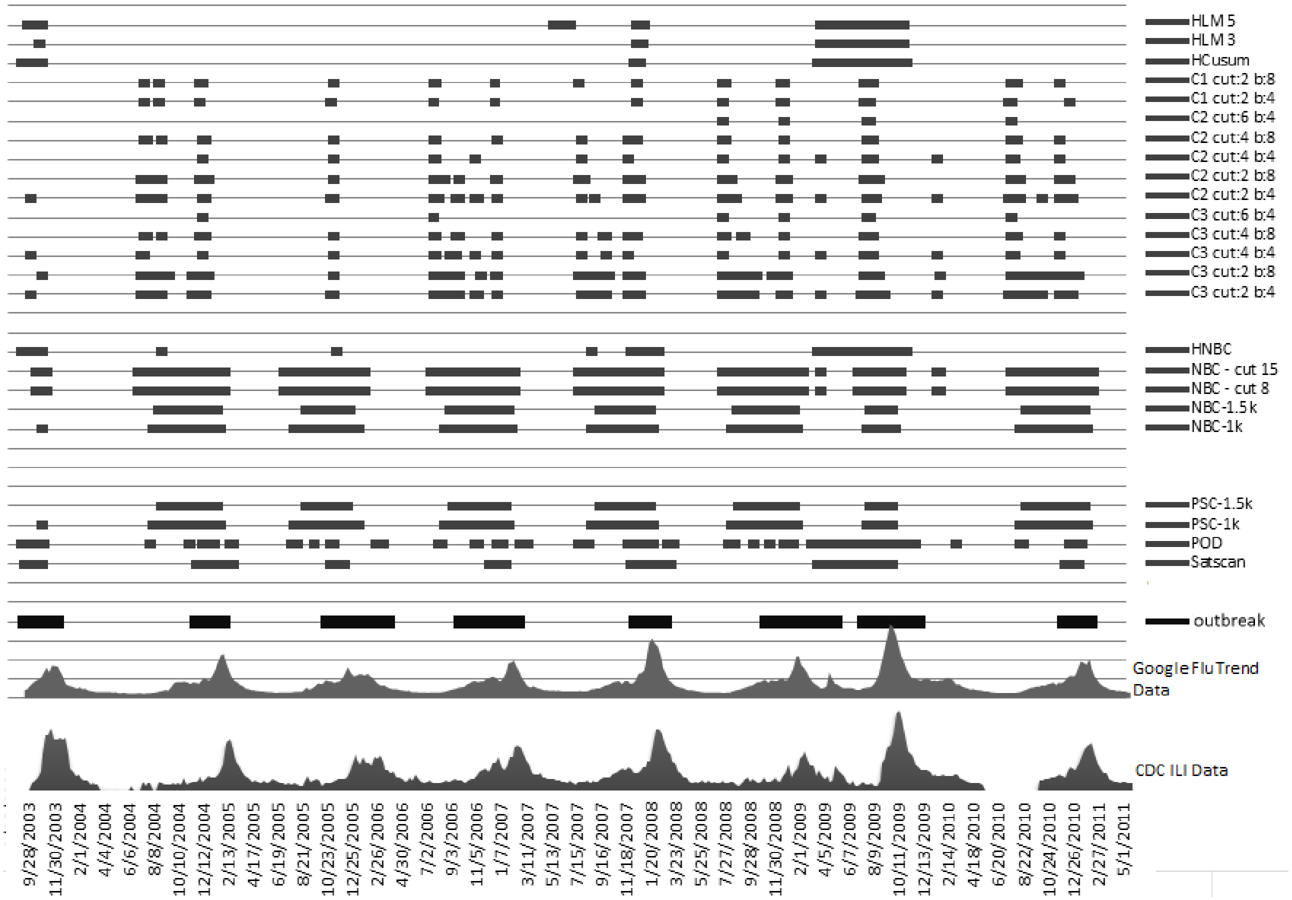


Figure 12. US Department of Health and Human Services region 10. The x-axis plots the Google Flu Trends and Centers for Disease Control and Prevention (CDC) data. The horizontal bars indicate where each method detected an epidemic. Cut indicates cut-off point (more is less sensitive) and b indicates baseline data (training window). The thick horizontal bar at the bottom shows actual outbreak. HCusum = historical cumulative sum, HLM = historical limits method, HNBC = historical negative binomial cumulative sum, ILI = influenza-like illnesses, k = reference value for threshold, NBC = negative binomial cumulative sum, POD = Poisson outbreak detection, PSC = Poisson cumulative sum.

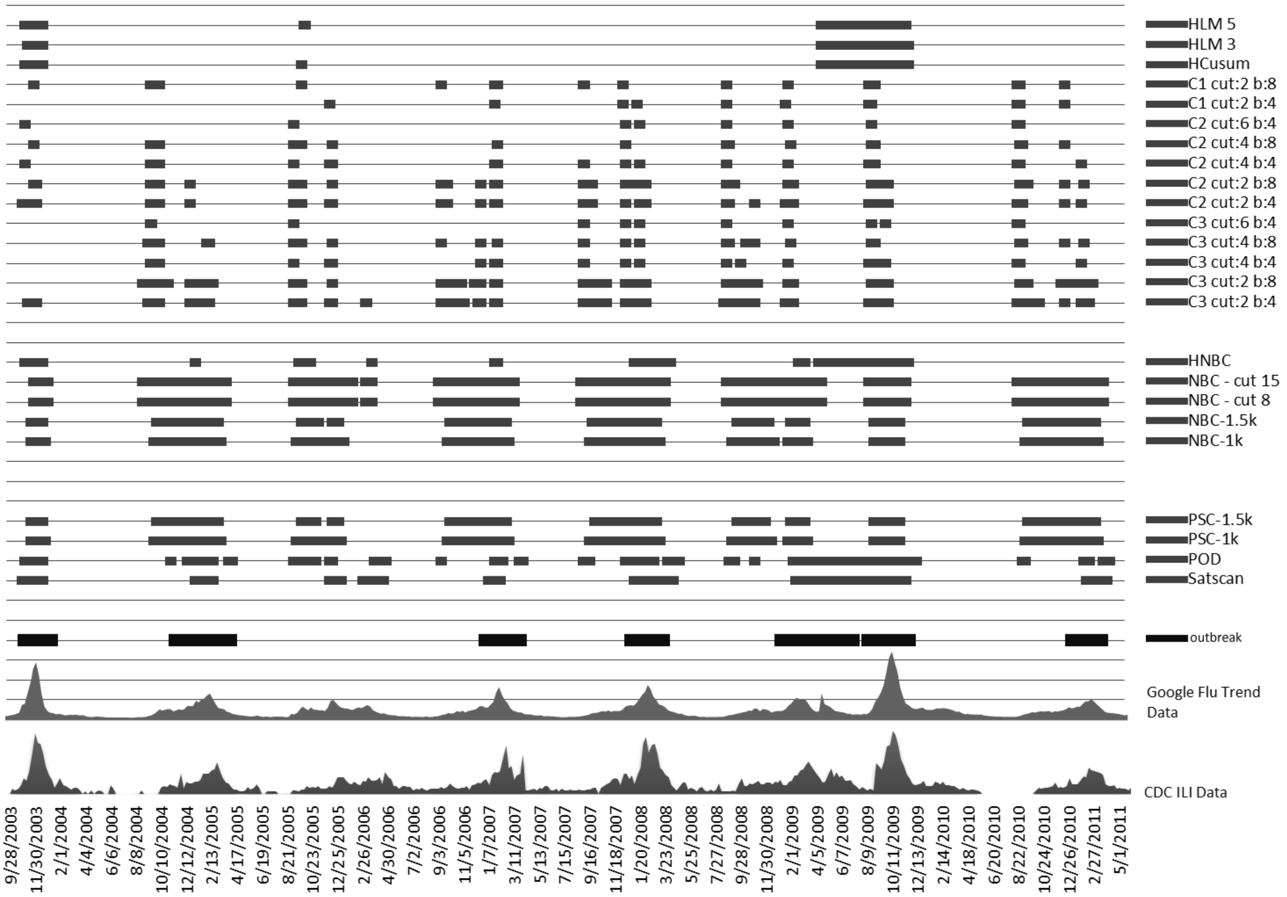
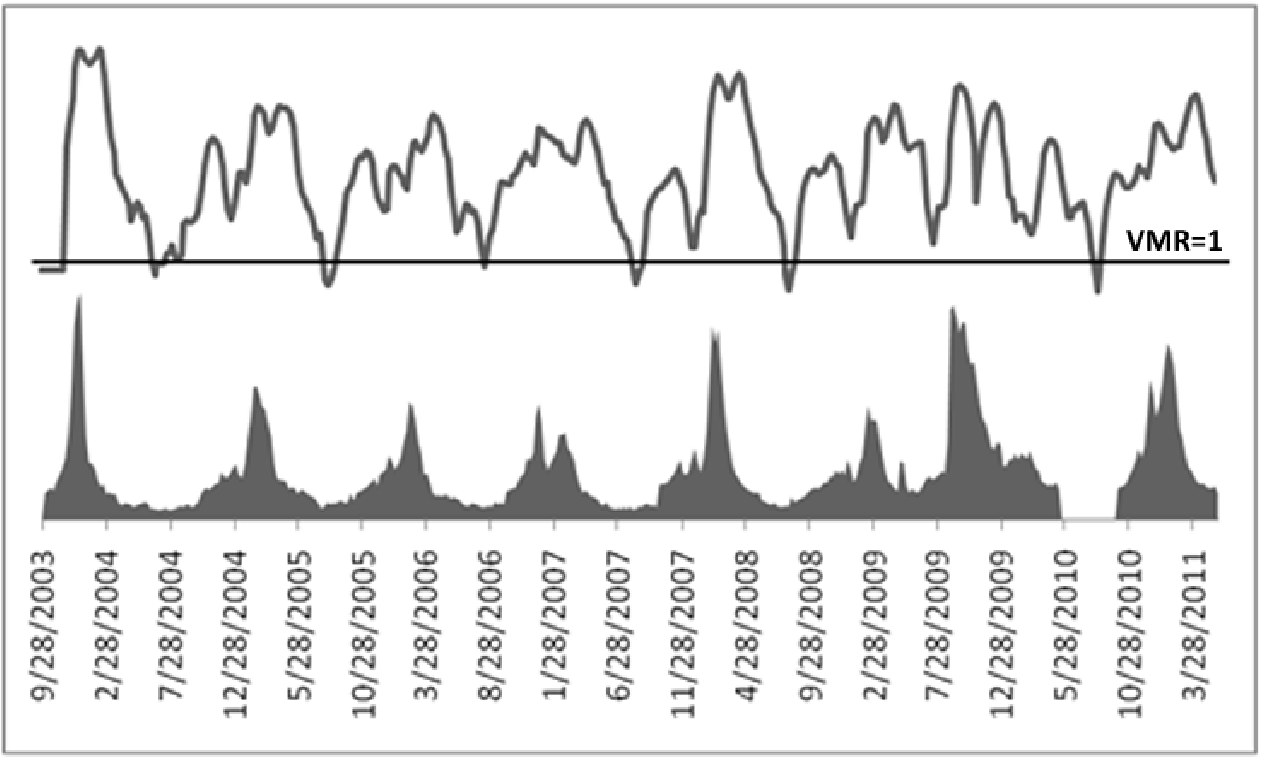


Figure 13. US Centers for Disease Control and Prevention data with the variance to mean ratio (VMR) line above, along the VMR = 1 mark.



Discussion

In this study, we augmented the capabilities of Google Flu Trends by evaluating various algorithms to translate the raw search query volume produced by this service into actionable alerts. We focused, in particular, on leveraging the ability of Google Flu Trends to provide a near real-time alternative to conventional disease surveillance networks and to explore the practicality of building an early epidemic detection system using these data. This paper presents the first detailed comparative analysis of popular early epidemic detection algorithms on Google Flu Trends. We explored the relative merits of these methods and considered the effects of changing Internet prevalence and population sizes on the ability of these methods to predict epidemics. In these evaluations, we drew upon data collected by the CDC and assessed the ability of each algorithm within a consistent experimental framework to predict changes in measured CDC case frequencies from the Internet search query volume.

Our analysis showed that adding a layer of computational intelligence to Google Flu Trends data provides the opportunity for a reliable early epidemic detection system that can predict disease outbreaks with high accuracy in advance of the existing systems used by the CDC. However, we note that realizing this opportunity requires moving beyond the CUSUM- and HLM-based normal distribution approaches traditionally employed by the CDC. In particular, while we did not find a single best method to apply to Google Flu Trends data, the results of our study strongly support negative binomial- and Poisson-based algorithms being more useful when dealing with potentially noisy search query data from regions with varying Internet penetrations. For such data, we found that normal distribution algorithms did not perform as well as the negative binomial and Poisson distribution algorithms.

Furthermore, our analysis showed that the patient data of a disease follows different distributions throughout the year. Therefore, when VMR of data is equal to 1, it is ideally following a Poisson distribution and could be handled by a Poisson-based algorithm. As the increase in variance raises VMR above 1, the data become overdispersed. Poisson-based algorithms can handle this overdispersion, up to a limit [29]. When VMR is very high, an algorithm is needed that considers the variance as a parameter and raises alarms accordingly. Since negative binomial distribution-based algorithms consider the variance [29], such algorithms perform better in similar scenarios. For instance, NBC is accurate in raising an alarm for overdispersed count data [29]. To get better results, we propose an approach, based on the above discussion, of changing the distribution expectation of an algorithm along with the rise and fall in VMR. This area should be explored in more depth to produce algorithms that adapt according to the data's distribution type.

Our research is the first attempt of its kind to relate epidemic prediction using Google Flu Trends data to Internet penetration and the size of the population being assessed. We believe that understanding how these factors affect algorithms to predict epidemics is an integral question for scaling a search query-based system to a broad range of geographical regions and communities. In our investigations, we observed that both Internet penetration and population size had a definite impact on algorithm performance. SaTScan performs better when applied to data from regions with high Internet penetration and small population size, while POD and NBC achieves better results when Internet penetration is low and population size is large. CUSUM performs best in regions with a large population. While the availability of search query data and measured (ie, CDC) case records restrict our analyses to the United States, we believe many of these insights may be useful in developing an early epidemic prediction system for other regions, including communities in the developing world.

In conclusion, we present an early investigation of algorithms to translate data from services such as Google Flu Trends into a fully automated system for generating alerts when the likelihood of epidemics is quite high. Our research augments the ability to detect disease outbreaks at early stages, when many of the conditions that impose an immense burden globally can be treated with better outcomes and in a more cost-effective manner. In addition, the ability to respond early to imminent conditions allows for more proactive restriction of the size of any potential outbreak. Together, the findings of our study provide a means to convert raw data collected over the Internet into more fine-grained information that can guide effective policy in countering the spread of diseases.

Based on our work, we have developed FluBreaks (dritte.org/flubreaks), an early warning system for flu epidemics using Google Flu Trends.

Acknowledgments

We acknowledge Dr Zeeshan Syed of the University of Michigan for his valuable feedback and intellectual contribution. We thank Dr Farkanda Kokab, Professor of Epidemiology, Institute of Public Health, Pakistan, and Dr Ijaz Shah, Professor and Head of Department of Community Medicine, Quaid-e-Azam Medical College, Pakistan, for marking our outbreaks and providing us with valuable feedback. We also thank Dr Lakshminarayanan Subramanian of New York University for reviewing our paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Ranking of algorithms in different parameters of evaluation for HSS Region 4 (Highest Population).

[PDF File (Adobe PDF File), 48KB - jmir_v14i5e125_app1.pdf]

Multimedia Appendix 2

Ranking of algorithms in different parameters of evaluation for HSS Region 6 (Lowest Percent Internet Use).

[PDF File (Adobe PDF File), 48KB - jmir_v14i5e125_app2.pdf]

Multimedia Appendix 3

Ranking of algorithms in different parameters of evaluation for HSS Region 10 (Lowest Population and Highest Percent Internet Use).

[PDF File (Adobe PDF File), 48KB - jmir_v14i5e125_app3.pdf]

References

1. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
2. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006:244-248. [Medline: 17238340]
3. Bennett GG, Glasgow RE. The delivery of public health interventions via the Internet: actualizing their potential. Annu Rev Public Health 2009;30:273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]
4. Castillo-Salgado C. Trends and directions of global public health surveillance. Epidemiol Rev 2010 Apr;32(1):93-109. [doi: 10.1093/epirev/mxq008] [Medline: 20534776]
5. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
6. Wilson K, Brownstein JS. Early detection of disease outbreaks using the Internet. CMAJ 2009 Apr 14;180(8):829-831. [doi: 10.1503/cmaj.090215] [Medline: 19364791]
7. Seifter A, Schwarzwalder A, Geis K, Aucott J. The utility of "Google Trends" for epidemiological research: Lyme disease as an example. Geospat Health 2010 May;4(2):135-137 [FREE Full text] [Medline: 20503183]
8. Breyer BN, Sen S, Aaronson DS, Stoller ML, Erickson BA, Eisenberg ML. Use of Google Insights for Search to track seasonal and geographic kidney stone incidence in the United States. Urology 2011 Aug;78(2):267-271. [doi: 10.1016/j.urology.2011.01.010] [Medline: 21459414]
9. Google. 2012 Mar 6. Flu Trends: United States 2011-2012 URL: http://www.google.org/flutrends/us/ [accessed 2012-03-08] [WebCite Cache ID 660ZkIfcu]
10. Hutwagner L, Thompson W, Seeman GM, Treadwell T. The bioterrorism preparedness and response Early Aberration Reporting System (EARS). J Urban Health 2003 Jun;80(2 Suppl 1):i89-i96. [Medline: 12791783]
11. Fricker RD, Hegler BL, Dunfee DA. Comparing syndromic surveillance detection methods: EARS' versus a CUSUM-based methodology. Stat Med 2008 Jul 30;27(17):3407-3429. [doi: 10.1002/sim.3197] [Medline: 18240128]
12. Hutwagner LC, Thompson WW, Seeman GM, Treadwell T. A simulation model for assessing aberration detection methods used in public health surveillance for systems with limited baselines. Stat Med 2005 Feb 28;24(4):543-550. [doi: 10.1002/sim.2034] [Medline: 15678442]
13. Stroup DF, Williamson GD, Herndon JL, Karon JM. Detection of aberrations in the occurrence of notifiable diseases surveillance data. Stat Med 1989 Mar;8(3):323-9; discussion 331. [Medline: 2540519]
14. Hutwagner LC, Maloney EK, Bean NH, Slutsker L, Martin SM. Using laboratory-based surveillance data for prevention: an algorithm for detecting Salmonella outbreaks. Emerg Infect Dis 1997;3(3):395-400. [doi: 10.3201/eid0303.970322] [Medline: 9284390]
15. Hutwagner L, Browne T, Seeman GM, Fleischauer AT. Comparing aberration detection methods with simulated data. Emerg Infect Dis 2005 Feb;11(2):314-316 [FREE Full text] [doi: 10.3201/eid1102.040587] [Medline: 15752454]
16. Gatton ML, Kelly-Hope LA, Kay BH, Ryan PA. Spatial-temporal analysis of Ross River virus disease patterns in Queensland, Australia. Am J Trop Med Hyg 2004 Nov;71(5):629-635 [FREE Full text] [Medline: 15569796]
17. Kulldorff M, Information Management Services Inc. satscan.org. 2005. SaTScan: Software for the Spatial, Temporal, and Space-Time Scan Statistics URL: http://www.satscan.org/ [accessed 2012-03-08] [WebCite Cache ID 660ZfnlJK]
18. Unkel S, Farrington CP, Garthwaite PH, Robertson C, Andrews N. Statistical methods for the prospective detection of infectious disease outbreaks: a review. J R Stat Soc 2012;175(1):49-82. [doi: 10.1111/j.1467-985X.2011.00714.x]
19. Lucas JM. Counted data CUSUM's. Technometrics 1985;27(2):129-144 [FREE Full text]
20. Hawkins DM, Olwell DH. Cumulative Sum Charts and Charting for Quality Improvement. New York, NY: Springer; 1998.

21. Watkins RE, Eagleson S, Veenendaal B, Wright G, Plant AJ. Applying cusum-based methods for the detection of outbreaks of Ross River virus disease in Western Australia. BMC Med Inform Decis Mak 2008;8:37 [FREE Full text] [doi: 10.1186/1472-6947-8-37] [Medline: 18700044]
22. Pelecanos AM, Ryan PA, Gatton ML. Outbreak detection algorithms for seasonal disease data: a case study using Ross River virus disease. BMC Med Inform Decis Mak 2010;10:74 [FREE Full text] [doi: 10.1186/1472-6947-10-74] [Medline: 21106104]
23. Centers for Disease Control and Prevention. 2012 Mar 2. Seasonal Influenza (Flu): 2011-2012 Influenza Season Week 8 Ending February 25, 2012 URL: http://www.cdc.gov/flu/weekly/ [accessed 2012-03-08] [WebCite Cache ID 660YMRZYi]
24. Centers for Disease Control and Prevention. 2011 Oct 7. Seasonal Flu (Influenza): Overview of Influenza Surveillance in the United States URL: http://www.cdc.gov/flu/weekly/overview.htm [accessed 2012-05-15] [WebCite Cache ID 67gSoQJVS]
25. Fact Monster. Pearson Education. 2012. U.S. Population by State, 1790 to 2010 URL: http://www.factmonster.com/ipka/A0004986.html [accessed 2012-03-08] [WebCite Cache ID 660XsTmos]
26. US Department of Commerce, National Telecommunications & Information Administration. 2009. Current Population Survey (CPS) Internet Use 2009 URL: http://www.ntia.doc.gov/data/CPS2009_Tables.html [accessed 2012-03-08] [WebCite Cache ID 660Y9BiVr]
27. US Department of Commerce, National Telecommunications and Information Administration. US Census Bureau, Statistical Abstract of the United States. 2011. Household Internet Usage by Type of Internet Connection and State: 2009 URL: http://www.census.gov/compendia/statab/2011/tables/11s1155.pdf [accessed 2012-03-08] [WebCite Cache ID 660YHNY66]
28. Cox DR, Lewis PAW. The Statistical Analysis of Series of Events. London: Chapman and Hall; 1966.
29. McCullagh P, Nelder JA. Generalized Linear Models. 2nd edition. London: Chapman and Hall; 1989.
30. Cameron AC, Trivedi. Regression Analysis of Count Data. Cambridge, UK: Cambridge University Press; 1998.

Abbreviations

CDC: Centers for Disease Control and Prevention
CUSUM: cumulative sum
EA: percentage early alarms
EARS: Early Aberration Reporting System
HCusum: historical cumulative sum
HLM: historical limits method
ILINet: Outpatient Influenza-like Illness Surveillance Network
NBC: negative binomial cumulative sum
OT: percentage overlap
POD: Poisson outbreak detection
PSC: Poisson cumulative sum
RFP: percentages false-positives
RTP: percentage true-positives
VMR: variance to mean ratio

Edited by G Eysenbach; submitted 08.03.12; peer-reviewed by M Gatton; comments to author 29.03.12; revised version received 18.05.12; accepted 10.07.12; published 04.10.12.

Please cite as:
Pervaiz F, Pervaiz M, Abdur Rehman N, Saif U
FluBreaks: Early Epidemic Detection from Google Flu Trends
J Med Internet Res 2012;14(5):e125
URL: http://www.jmir.org/2012/5/e125/
doi:10.2196/jmir.2102
PMID:23037553

Original Paper

Development of a Health Information Technology Acceptance Model Using Consumers’ Health Behavior Intention

Jeongeun Kim[1,2,3*], RN, INS, PhD; Hyeoun-Ae Park[1,2,3*], RN, FAAN, PhD

[1]College of Nursing, Seoul National University, Seoul, Korea, Republic Of
[2]Research Institute of Nursing Science, Seoul National University, Seoul, Korea, Republic Of
[3]Interdisciplinary Program of Medical Informatics, Seoul National University, Seoul, Korea, Republic Of
[*]all authors contributed equally

Corresponding Author:
Jeongeun Kim, RN, INS, PhD
College of Nursing
Seoul National University
103 Daehak-ro Jongno-gu
Seoul, 110-799
Korea, Republic Of
Phone: 82 7408483
Fax: 82 7654103
Email: kim0424@snu.ac.kr

Abstract

Background: For effective health promotion using health information technology (HIT), it is mandatory that health consumers have the behavioral intention to measure, store, and manage their own health data. Understanding health consumers’ intention and behavior is needed to develop and implement effective and efficient strategies.

Objective: To develop and verify the extended Technology Acceptance Model (TAM) in health care by describing health consumers’ behavioral intention of using HIT.

Methods: This study used a cross-sectional descriptive correlational design. We extended TAM by adding more antecedents and mediating variables to enhance the model’s explanatory power and to make it more applicable to health consumers’ behavioral intention. Additional antecedents and mediating variables were added to the hypothetical model, based on their theoretical relevance, from the Health Belief Model and theory of planned behavior, along with the TAM. We undertook structural equation analysis to examine the specific nature of the relationship involved in understanding consumers’ use of HIT. Study participants were 728 members recruited from three Internet health portals in Korea. Data were collected by a Web-based survey using a structured self-administered questionnaire.

Results: The overall fitness indices for the model developed in this study indicated an acceptable fit of the model. All path coefficients were statistically significant. This study showed that perceived threat, perceived usefulness, and perceived ease of use significantly affected health consumers’ attitude and behavioral intention. Health consumers’ health status, health belief and concerns, subjective norm, HIT characteristics, and HIT self-efficacy had a strong indirect impact on attitude and behavioral intention through the mediators of perceived threat, perceived usefulness, and perceived ease of use.

Conclusions: An extended TAM in the HIT arena was found to be valid to describe health consumers’ behavioral intention. We categorized the concepts in the extended TAM into 3 domains: health zone, information zone, and technology zone.

(J Med Internet Res 2012;14(5):e133) doi:10.2196/jmir.2143

KEYWORDS

Technology Acceptance Model; health behavior; intention; consumer

Introduction

Since the advent of the information age, the use of diverse health information technology (HIT) has become widespread in chronic disease management, disease prevention, and health promotion. HIT is “the application of information processing involving both computer hardware and software that deals with the storage, retrieval, sharing, and use of health care information, data, and knowledge for communication and decision making" [1]. Informatics, which is another integral aspect of HIT, refers to the science of information, the practice of information processing, and the engineering of information systems. Health informatics refers to the intersection of information science, computer science, and health care, and deals with the resources, devices, and methods required for optimizing the acquisition, storage, retrieval, and use of information in health and biomedicine. Health informatics tools include not only computers but also clinical guidelines, formal medical terminologies, and information and communication systems [2]. As such, the scope of HIT is very broad and encompasses many aspects of hardware and software, including the computer, smartphone, Internet, and social network services. This advance, alongside the advancement of genomics, digitization of data collection, and storage, has inspired the concept of the health avatar [3], which is expected to cause a paradigm shift in the field of health care. The health avatar, promising active use of health and disease management technology, is enabled by life logging, self-tracking, quantified self, and more. In fact, electronics, Web service programs, and other smartphone apps that track, analyze, and provide feedback based on a person's diet, exercise, sleeping pattern, and activity are already in public use [4].

For effective use of collected health-related data in HIT, it is crucial that health consumers have the behavioral intention to measure, store, and manage their own data. The effort put forth by health consumers to measure, store, and manage their own data strongly determines the quality of the data. Therefore, until data collection and storage is fully automated, consumers' behavioral intention will be the predominant deciding factor in the accuracy and usefulness of such data. Additionally, a better understanding of health consumers' intention and behavior would aid the development and implementation of effective and efficient strategies. Thus, identifying the factors influencing health consumers' intention and behavior to measure, store, and manage their own health-related data would enable the development of a theoretical model to successfully describe their intentions and actions. Developing a model requires determining the interrelationships among the factors of health information by integrating various behavior and information technology theories.

First, regarding health behavior theories, health behavior includes any activity undertaken by an individual, regardless of actual or perceived health status, for the purpose of promoting, protecting, or maintaining health, whether or not such behavior is objectively effective in obtaining the intended results [5]. The best way to design programs to achieve positive changes in health behavior is to understand why people behave as they do and what might motivate them to change. Thus, theories and models of health behavior help explain consumers' health behavior and guide the development of more effective ways to influence and change their behavior. Health behavior is, however, far too complex to be explained by a single, unified theory, and no single theory or model dominates the research or practice in health-related behavior [6]. Instead, past research devised models that drew on several theories to help understand a specific problem in a particular setting or context.

Next, regarding the theories of information technology, the Technology Acceptance Model (TAM) is the most widely applied model to describe consumer acceptability of information technology. TAM can be adapted by applying various factors involving consumers' behavior in the context of HIT [7,8]. TAM has been continuously expanded to TAM2, the unified theory of acceptance and use of technology, and TAM3, and each of these expansions was motivated by the need to predict the usability of new information technology and to identify and stimulate the use of the technology [9-12]. TAM considers cultural trends and social context as the main factors, and focuses on what attributes of a given technology increase consumers' acceptance of the technology. Specifically, TAM suggests that the acceptance of a technology by consumers can be increased if efforts to improve the technology are directed by how the technology is perceived by consumers [7,8,13]. Therefore, TAM is a useful model for developing strategies to increase the acceptance of an information technology, as it provides a direct relationship between acceptance of the technology the technology's perceived usability and ease of use. Furthermore, the applicability of TAM has been expanded by implementing theories of health behavior with the advance of HIT [14].

The Health Belief Model (HBM) was one of the first and one of the best-known social cognition models to explain health behavior change [15]. Originally, the model was designed to predict behavioral response to the treatment received by acutely or chronically ill patients, but in more recent years the model has been used to predict more general health behaviors [6]. The HBM suggests that belief in a personal threat, together with belief in the effectiveness of the proposed behavior, predicts the likelihood of engaging in that behavior [16]. The theory of reasoned action is a model for the prediction of behavioral intention, spanning predictions of attitude and predictions of behavior. The subsequent separation of behavioral intention from behavior allows for the explanation of limiting factors on attitudinal influence. The theory of reasoned action was "born largely out of frustration with traditional attitude-behavior research, much of which found weak correlations between attitude measures and performance of volitional behavior" [17]. In psychology, the theory of planned behavior (TPB) is a theory about the link between attitude and behavior. The concept was developed to improve on the predictive power of the theory of reasoned action by including perceived behavioral control. It is one of the most predictive persuasion theories [18].

Several key concepts, such as attitude, behavioral intention, and behavior, from these theories are the same and the paths are very similar; thus, it is possible to synthesize a model by extracting the common key concepts and building the logical paths between them. In 2008, Yun [14] built an integrated model with concepts from TAM [7,8], HBM [15], theory of reasoned action [19], and TPB [20]. The study demonstrated that consumers' cognitive factors such as health concerns, perceived threat, and Internet self-efficacy affect their actions in seeking health information on the Internet through perceived usefulness and perceived ease of use. However, the model is limited to the pursuit of health information from the Internet, and therefore cannot comprehensively describe health consumers' behaviors using various HITs in different environments. Therefore, the model needs to be built upon further and modified to develop TAM in the context of HIT so that it can describe consumers' health behavior when using various HITs.

Objectives

The objective of this study was to develop and test a model describing the behavioral intention and the health behavior of consumers of various HITs, including the Internet, smartphones, and social network services. This model, the HIT-driven extended TAM, is called the Health Information Technology Acceptance Model (HITAM).

Methods

Participants and Data Collection

We recruited study participants from the three largest online health portals in South Korea. They are KorMedi.com, HiDOC, and Kunkang-In (which means Healthy People, from the National Health Insurance Corporation, Republic of Korea). Members of these popular online health portals who used health information through the Internet, smartphones, or social network services were contacted for participation in the study.

Data were collected using an online survey method developed by a private company specializing in online survey research. A survey was posted on these portals from October 21 to December 8, 2011, and in total 728 members replied.

We sought and obtained approval from the institutional review board of the College of Nursing, Seoul National University before collecting the data.

Measurement Tool

For this study, we developed a structured, self-administered questionnaire titled "We would like to know whether you use Smart HIT in your health management." The questionnaire was composed of 50 items, measuring general characteristics of the research participants and variables in 10 categories. General characteristics were measured with 6 items, health belief and concerns with 5 items, subjective norm with 5 items, perceived susceptibility with 3 items, perceived seriousness with 4 items, HIT self-efficacy with 6 items, HIT reliability with 5 items, perceived ease of use with 5 items, perceived usefulness with 5 items, attitude with 3 items, and behavioral intention with 3 items. Health status was measured by asking if they had any diseases or comorbidity. (Refer to Table 4 for Cronbach alpha of each category.)

Attitude and behavioral intention are two outcome categories. Attitude measures the positive perception of and satisfaction with the use of HIT. Behavioral intention measures the intent and willingness to use HIT. Participants were asked to rate their agreement with the following 3 statements to measure attitude: (1) I am positive about using HIT to manage my health and to search for reliable health information, (2) I think it is beneficial to manage my health and search for reliable health information using HIT, and (3) I am satisfied by and large with the use of HIT to manage my health and search for reliable health information using HIT. The following 3 items measured behavioral intention: (1) I will continue to use HIT to manage my health and to search for reliable health information, (2) I will regularly use HIT to manage my health and to search for reliable health information, and (3) I will recommend use of HIT to other people to manage their health and to search for reliable health information.

We derived items measuring perceived susceptibility, perceived seriousness, perceived threat, and behavioral intention from HBM. Items measuring HIT self-efficacy, HIT reliability, perceived ease of use, perceived usefulness, attitude, and behavioral intention were derived from TAM3. Items measuring health belief and concerns, subjective norm, attitude, and behavioral intention were derived from TPB. We adapted categories of the questionnaire from Yun's study on the development of a consumer health information-seeking behavior model from 2008 [14]. Table 1 compares Yun's study with the present study. We made these modifications to extend and change the scope from a narrow Internet focus to broader HIT to reflect the trend of HIT adoption. Specifically, we extended the scope of the questionnaire from Internet health information use to HIT in order to better probe health consumers' behavior.

Table 1. Comparison of variables in Yun [14] and the present study.

Variables in Yun's model	Variables in proposed model	Reasons for inclusion, exclusion, or modification
Health concerns	Health belief and concerns	
Perceived susceptibility	Perceived susceptibility	
Perceived seriousness	Perceived seriousness	
Subjective health-related knowledge		Excluded because it is for the knowledge level only, which is too specific
	Subjective norm	Added to strengthen the normative beliefs such as the motivation to comply in the theory of planned behavior
Internet self-efficacy	HIT[a] self-efficacy	Modified to adapt to HIT because it is the broader term
Perceived ease of use	Perceived ease of use	
Perceived usefulness	Perceived usefulness	
Perceived credibility	HIT reliability	Perceived credibility in Yun's model is the same concept as the current model's HIT reliability
Attitude	Attitude	
Behavioral intention	Behavioral intention	

[a] Health information technology.

The measurement tool was a 7-point Likert-type scale ranging from 1 for strongly disagree to 7 for strongly agree. The questionnaire was originally created in Korean, so it was not necessary to translate it. The reliability of the original instrument was indicated by Cronbach alpha = .853. The content of the questionnaire was independently confirmed by a group of HIT experts.

Path Diagram and Structural Equation Model

To describe health consumers' behavioral intention toward HIT services that use computers, the Internet, and smartphones, we developed a structural equation model based on previous research and a literature review. We also analyzed other theories relevant to HITAM synthesis, such as HBM, TPB, TAM, Extended TAM (TAM2, unified theory of acceptance and use of technology, TAM3), and Consumer's Health Information Seeking Behavior Model.

Figure 1 summarizes the relationships among the models revealed by our analysis. Based on these relationships, various endogenous and exogenous variables that explain the behavioral intentions for HIT and potential variables were derived to construct a hypothetical model. We borrowed many of the concepts using their original phrasing, but we modified a few for clarification. For example, we took age and disease from the demographic category as measurement variables in the health status category that affect perceived threat. Perceived susceptibility and perceived seriousness from HBM were included as exogenous variables that affect perceived threat as well. Finally, we included behavioral belief from TPB in our health belief and concerns category.

Extensive monitoring of behavioral change requires a long-term observational study and prohibitively large sample groups. Also, self-reported behavioral intention and users' action itself are hard to measure. Finally, cross-sectional research methodology provides a challenge, as the correlation between behavioral intention and the actual taking of action has been shown to be weak [21,22]. Therefore, the model does not verify the previously studied relationship between health behavioral intention and health action. Rather, the influential factors and behavioral intention were the primary focus of model verification.

Our structural equation model relied on findings from previous research regarding consumers' behavioral intention to use HIT. Figure 2 shows the theoretical model for the analysis. Arrows with solid lines indicate relationships between the specified concepts. We hypothesized that consumers with worse health status and stronger health beliefs and concerns would perceive greater threat. In addition, we expected that consumers with higher subjective norm, HIT reliability, and HIT self-efficacy would perceive HIT to be more useful. We assumed that consumers with higher HIT reliability and HIT self-efficacy would perceive HIT to be easier to use. We also assumed that consumers with higher perceived threat and perceived ease of use would perceive HIT to be more useful. We hypothesized that consumers who perceive HIT to be more useful and easier to use would have a more positive attitude toward HIT. Finally, we hypothesized that consumers with a more positive attitude would show more positive behavioral intention to accept HIT.

Figure 1. Conceptual relationships of the relevant models for development of the Health Information Technology Acceptance Model (HITAM). HBM = Health Belief Model, TAM = Technology Acceptance Model, TPB = theory of planned behavior.

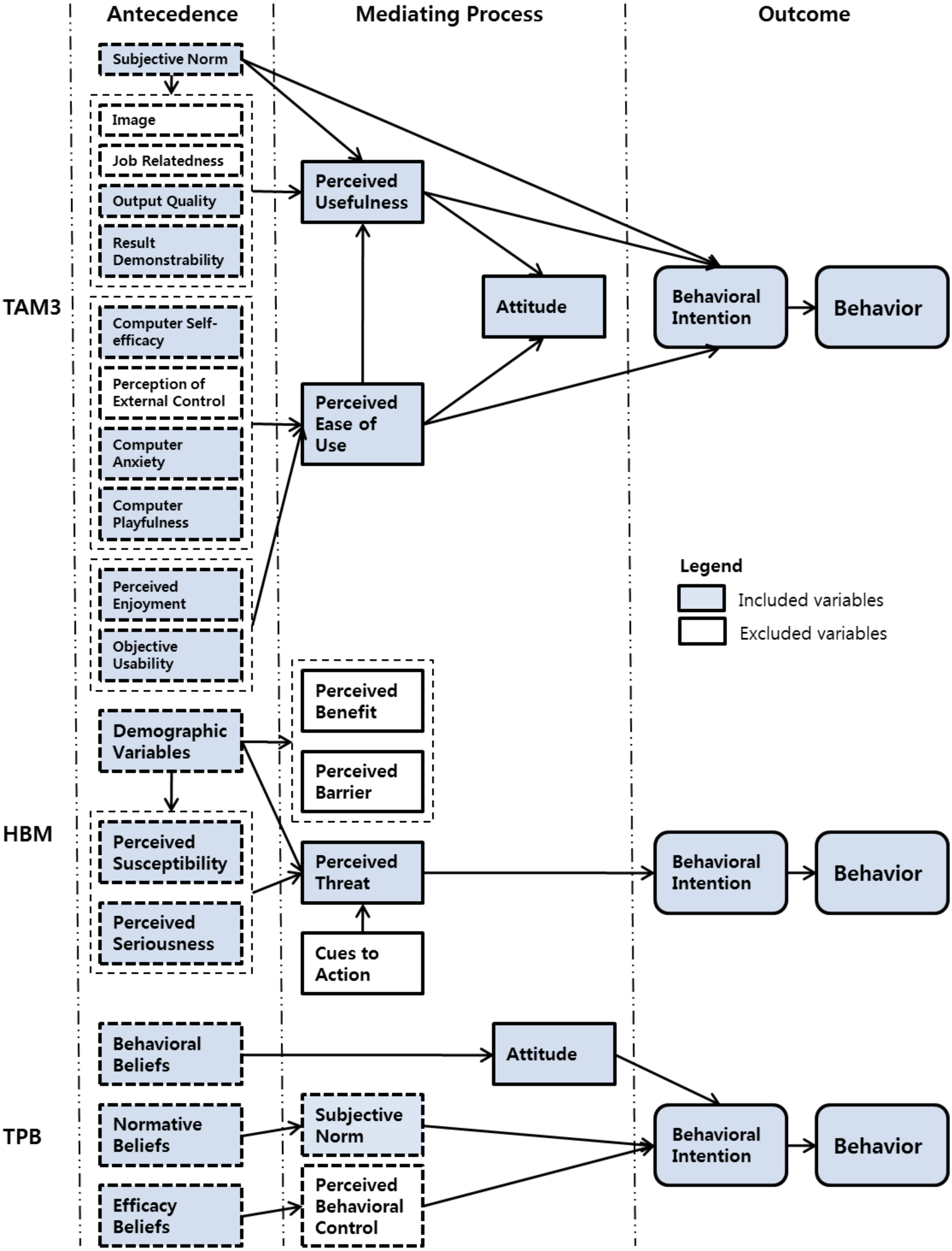
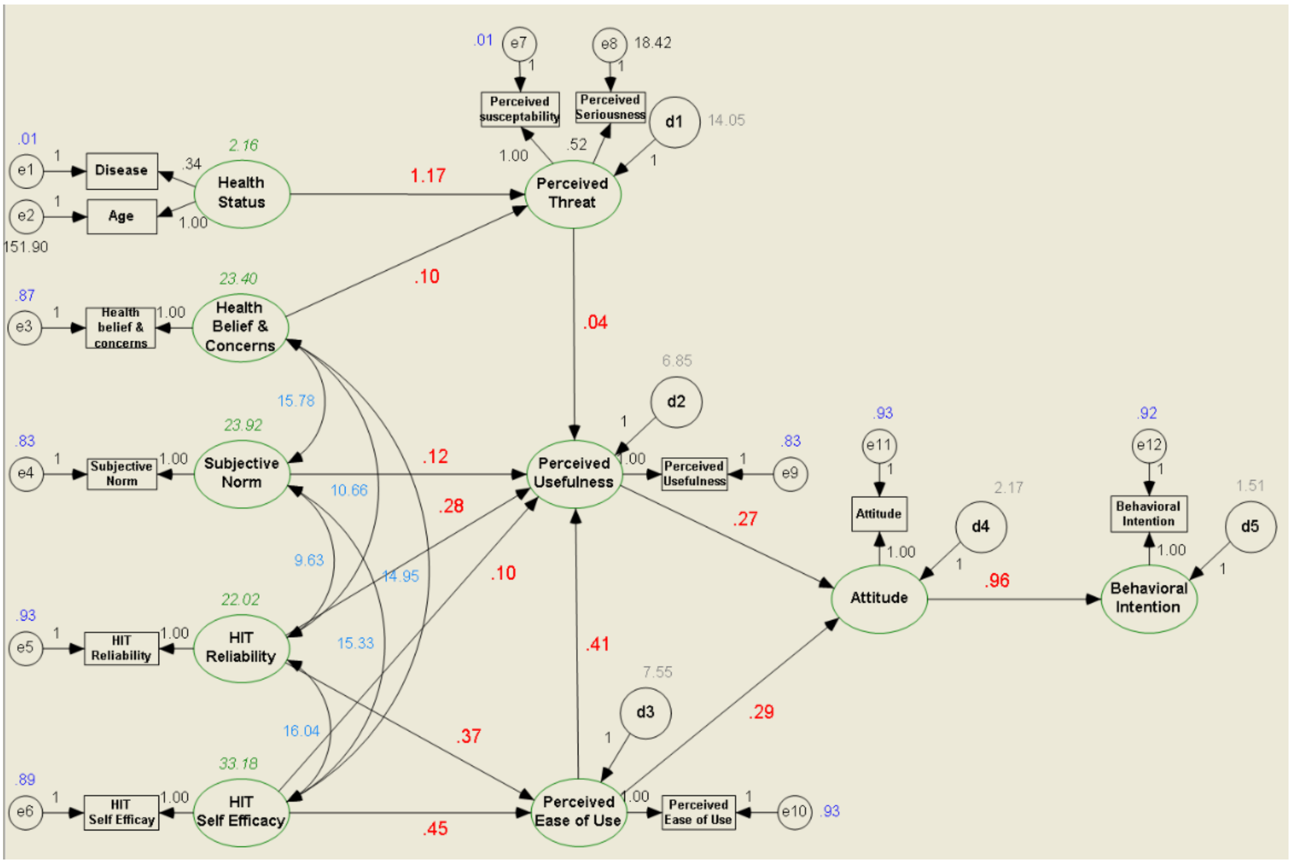


Figure 2. Path diagram of the Health Information Technology Acceptance Model (HITAM) Model. HIT = health information technology.



Statistical Analysis

Descriptive statistics of the participants’ general characteristics were analyzed with IBM SPSS version 19 (IBM Corporation, Somers, NY, USA). The structural equation model was fitted with maximum likelihood estimation routines in IBM SPSS Amos 16.

Results

The general characteristics of the 728 research participants are shown in Table 2. About an equal number of men and women participated in the study. A large fraction of the group were in their 30s, with a majority of the group having a college education or higher. The ratio of participants with chronic illness to those without was about one.

Table 2. General characteristics of the participants (N = 728).

Characteristic	n	%
Gender		
Male	372	51.1%
Female	356	48.9%
Age (years)		
<19	10	1.4%
20–29	144	19.8%
30–39	258	35.4%
40–49	150	20.6%
50–59	101	13.9%
>60	65	8.9%
Occupation		
Clerical	223	30.6%
Professional	113	15.5%
Homemaker	86	11.8%
Student	75	10.3%
Self-employed	58	8.0%
Manufacturing	33	4.5%
Government official	24	3.3%
Other	116	15.9%
Education		
< Middle school	10	1.4%
High school	177	24.3%
College	135	18.5%
University	299	41.1%
> Graduate school	107	14.7%
Chronic diseases		
Yes	361	49.6%
No	367	50.4%
Monthly income (US $)		
>1000	37	(5.1)
1001–2000	107	(14.7)
2001–3000	186	(25.5)
3001–4000	159	(21.8)
4001–5000	118	(16.2)
<5000	121	(16.6)

The descriptive statistics of the variable scores and the reliability coefficients of the measuring tool in each category are provided in Table 3.

Table 3. Descriptive statistics of the latent variables and the reliability coefficients (N = 728 for all variables).

Variable	Minimum	Maximum	Mean	SD	Kurtosis		Skewness		Cronbach alpha	No. of items
Health belief and concerns	5	35	27.04	4.930	–1.028	0.091	2.405	0.181	0.867	5
Subjective norm	5	35	23.68	4.978	–0.318	0.091	0.289	0.181	0.826	5
Perceived susceptibility	3	21	12.94	4.157	–0.291	0.091	–0.448	0.181	0.751	3
Perceived seriousness	4	28	20.91	4.801	–0.844	0.091	0.811	0.181	0.907	4
HIT[a] self-efficacy	6	42	31.25	5.841	–0.412	0.091	0.693	0.181	0.888	6
HIT reliability	5	35	25.55	4.795	–0.485	0.091	0.521	0.181	0.934	5
Perceived ease of use	5	35	27.74	4.847	–0.716	0.091	0.734	0.181	0.925	5
Perceived usefulness	5	35	25.99	4.710	–0.466	0.091	0.364	0.181	0.826	5
Attitude	3	21	16.76	3.024	–0.872	0.091	1.088	0.181	0.93	3
Behavioral intention	3	21	17.07	3.146	–0.921	0.091	0.956	0.181	0.919	3

[a] Health information technology.

We found that each measured variable satisfied the assumption of the univariate normality. Table 4 presents correlation coefficients for the measured variables. The correlations between the measured variables ranged from .001 to .817.

Table 4. Correlation coefficients between measured variables.

	Age	Diseases	Health belief and concerns	Subjective norm	Perceived susceptibility	Perceived seriousness	HIT[a] self-efficacy	HIT reliability	Perceived ease of use	Perceived usefulness	Attitude	Behavioral intention
Age	1											
Diseases	–.117**	1										
Health belief and concerns	.144**	–.023	1									
Subjective norm	.049	.025	.644**	1								
Perceived susceptibility	.032	–.411**	.119**	.151**	1							
Perceived seriousness	.057	–.156**	.340**	.207**	.447**	1						
HIT self-efficacy	–.053	.056	.517**	.529**	.155**	.288**	1					
HIT reliability	.048	.031	.449**	.405**	.143**	.316**	.574**	1				
Perceived ease of use	.002	.024	.488**	.403**	.139**	.350**	.733**	.655**	1			
Perceived usefulness	–.009	.026	.476**	.479**	.175**	.370**	.664**	.672**	.740**	1		
Attitude	.097**	.001	.520**	.387**	.126**	.358**	.640**	.694**	.740**	.737**	1	
Behavioral intention	.145**	.014	.496**	.361**	.110**	.350**	.635**	.604**	.705**	.686**	.817**	1

[a] Health information technology.

*$P < .05$, **$P < .01$.

Figure 2 presents the path diagram of the fitted model, and Table 5 shows unstandardized and standardized estimates of the model. This model exhibited an excellent fit to the data (χ^2_{46} = 375.6, $P < .001$, goodness of fit index = .923, root mean square error of approximation = .099). The model accounts for 83% of the variance in behavioral intention, 73% of the variance in attitude, 18% of the variance in perceived threat, 67% of the variance in perceived ease of use, and 68% of the variance in perceived usefulness. The path from perceived threat to behavioral intention was proposed in the hypothetical model. However, we removed this path in the final model because it was not significant.

Table 5. Standardized estimates of the Health Information Technology Acceptance Model (HITAM).

Endogenous variable	Exogenous variable	Regression weights (SE)	Standardized estimate (beta)	CR[a] (*t* value)	*P* value	SMC[b]
Perceived threat	Health status	1.167 (0.379)	.413	3.078	.002	.184
	Health belief and concerns	0.100 (0.029)	.117	3.415	<.001	
	HIT[c] reliability	0.372 (0.030)	.367	12.305	<.001	
Perceived ease of use	HIT self-efficacy	0.446 (0.024)	.542	18.275	<.001	.665
Perceived usefulness	Subjective norm	0.118 (0.026)	.126	4.475	<.001	.677
	HIT reliability	0.276 (0.033)	.281	8.381	<.001	
	Perceived ease of use	0.409 (0.041)	.422	9.996	<.001	
	Perceived threat	0.045 (0.025)	.040	1.777	.08	
	HIT self-efficacy	0.102 (0.032)	.127	3.218	.001	
Attitude	Perceived usefulness	0.268 (0.024)	.432	11.363	<.001	.734
	Perceived ease of use	0.288 (0.023)	.479	12.588	<.001	
Behavioral intention	Attitude	0.955 (0.025)	.912	38.669	<.001	.831

[a] Critical ratio.

[b] Squared multiple correlation.

[c] Health information technology.

Table 6 displays standardized measures of direct, indirect, and total effects of exogenous variables on endogenous variables in the model. All direct, indirect, and total effects were statistically significant except for the direct effect of perceived threat on perceived usefulness, the indirect effect of perceived threat on attitude, and the indirect effect of perceived threat on behavioral intention.

Table 6. Effects of exogenous variables on endogenous variables in the Health Information Technology Acceptance Model (HITAM).

Endogenous variable	Exogenous variable	Standardized direct effect	*P* value	Standardized indirect effect	*P* value	Standardized total effect	*P* value
Perceived threat	Health status	.413	.01	0		.413	.01
	Health belief and concerns	.117	.006	0		.117	.006
	HIT[a] self-efficacy	.542	.007	0		.542	.007
Perceived ease of use	HIT reliability	.367	.01	0		.367	.01
Perceived usefulness	HIT self-efficacy	.127	.009	.228	.02	.356	.009
	HIT reliability	.281	.009	.155	.009	.436	.008
	Subjective norm	.126	.01	0		.126	.01
	Perceived threat	.040	.06	0		.040	.06
	Perceive ease of use	.422	.008	0		.422	.008
	Health belief and concerns			.005	.04	.005	.04
	Health status			.017	.05	.017	.05
Attitude	Perceived ease of use	.479	.02	.182	.005	.661	.02
	Perceived usefulness	.432	.006	0		.432	.006
	HIT reliability			.364	.03	.364	.03
	HIT self-efficacy			.413	.01	.413	.01
	Subjective norm			.054	.006	.054	.006
	Health belief and concerns			.002	.03	.002	.03
	Health status			.007	.046	.007	.046
	Perceived threat			.017	.05	.017	.05
Behavioral intention	Attitude	.912	.01	0		.912	.01
	HIT reliability			.332	.03	.332	.03
	HIT self-efficacy			.377	.01	.377	.01
	Subjective norm			.049	.006	.049	.006
	Health belief and concerns			.002	.04	.002	.04
	Health status			.007	.04	.007	.04
	Perceived threat			.016	.05	.016	.05
	Perceived ease of use			.603	.01	.603	.01
	Perceived usefulness			.394	.005	.394	.005

[a] Health information technology.

Discussion

We categorized the influential factors affecting the behavioral intention to measure, store, and manage health-related data into three domains called the health zone, information zone, and technology zone. In each zone, the factors follow different mediating processes that lead to behavioral intention in health customers. In the first domain, the health zone, there is a cascade of effects starting from health status (age, disease, etc), to perceived threat, to perceived usefulness, to attitude and, finally, to behavioral intention. This further verifies the description of behavioral intention in the HBM, where an intermediate variable, attitude, leads to behavioral intention [23,24]. Additionally, we showed that health belief and concerns have an indirect effect on the behavioral intention to use HIT. This suggests that psychological incentives such as health belief and concerns may motivate people to take action toward their health management. For example, outcome valuation has a nonnegligible effect on the health belief and concerns, triggering behavioral intention in health consumers.

Consumers' perceived threat, which is measured by perceived susceptibility and perceived seriousness, had a somewhat smaller impact on perceived usefulness than in previous studies. However, perceived threat had a statistically significant indirect effect on attitude through perceived usefulness, subsequently increasing the behavioral intention to use HIT. This indirect relationship is exemplified by the tendency of individuals to actively use HIT when they perceive a potential threat to their health. This result is consistent with the model reported by Yun [14], where perceived threat leads to information-seeking behavior. One possible explanation for the relatively reduced impact of perceived threat on perceived usefulness found in this study may be that some research participants considered the issue of susceptibility to and seriousness of disease not to apply to themselves, owing to their good health. Another explanation may be that perceived susceptibility and perceived seriousness are concerned with disease, but perceived usefulness is concerned with the usefulness of HIT in health promotion.

In the second domain, the information zone, the factors have a similar cascade effect to that in the health zone. Perceived usefulness is significantly sensitive to subjective norms, such as social pressure or community competition, resulting in consumers forming positive attitude. Such a formed attitude has a consequence on behavioral intention. According to the Ajzen' TPB [18], behavioral intention is a direct determinant of behaviors, arguing that attitude, subjective norm, and perceived behavioral control of the action are the most powerful predicting factors for behavioral intention. Therefore, the result suggests that when using HIT to induce health behaviors in consumers, using a social network service or network to form an environment for users to compete with others online can increase efficiency.

The last domain, the technology zone, has factors with the following characteristics. HIT use forms specific HIT reliability, such as output quality and result demonstrability, and these affect perceived usefulness. More interestingly, HIT reliability also affects perceived ease of use, also affecting perceived usefulness. Similarly, HIT self-efficacy, such as HIT anxiety, playfulness, perceived enjoyment, and objective usability, also affect perceived usefulness and perceived ease of use. These two factors affect attitude and, finally, form behavioral intention. Our study revealed that HIT self-efficacy is a significant factor influencing HIT use. This suggests that consumers enjoying the use of HIT and gaining confidence in their ability to use HIT develop increased tendencies to use HIT. Specifically, the greater the self-efficacy, the greater the perceived ease of use, which is consistent with the finding reported by Yun [14].

Ajzen developed the TPB model that explains various human actions by integrating behavioral belief, normative belief, efficacy belief, attitude, subjective norm, and perceived behavioral control [18]. In this study, we demonstrated that behavioral belief, normative belief, and efficacy belief all serve as antecedents to a mediating process that results in behavioral intention. The similarity between these two studies suggests that the TPB is applicable to further research in HIT.

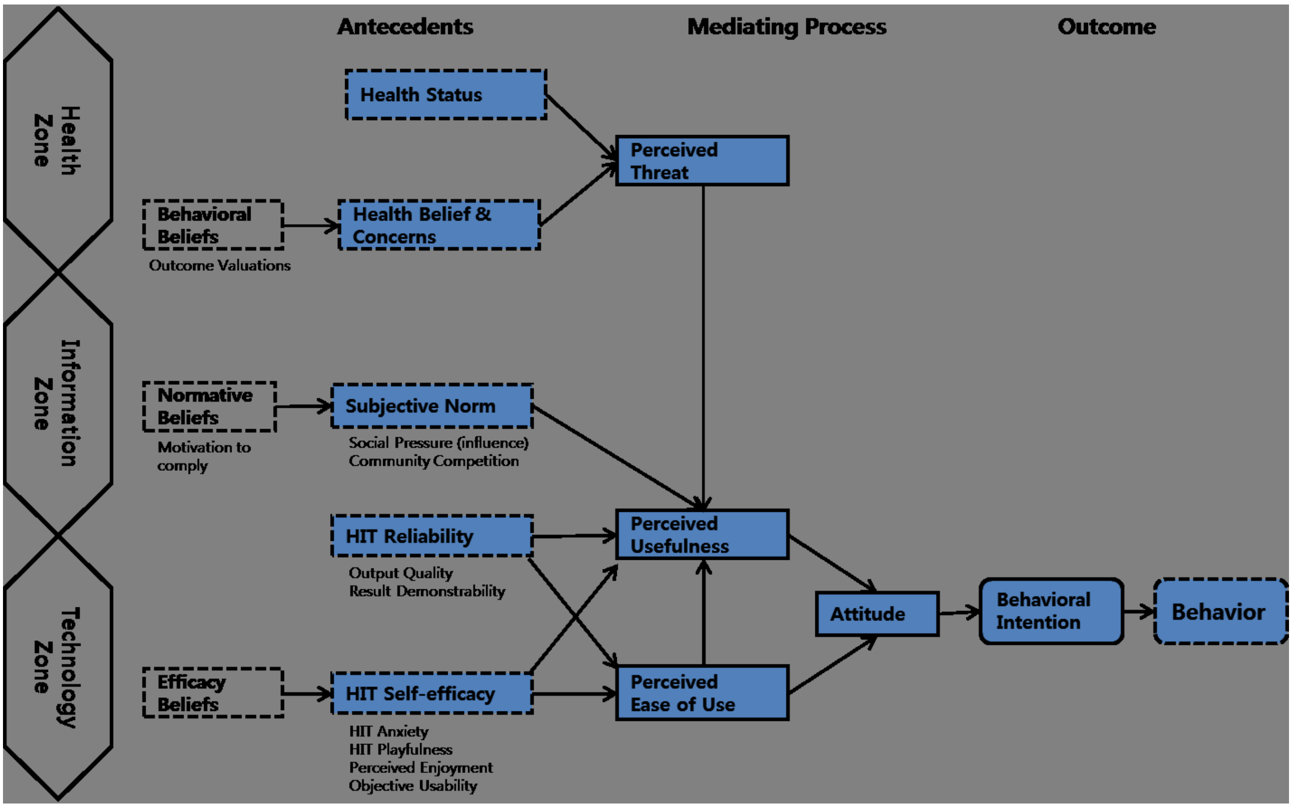
The key factors identified within the three zones—health status and health belief and concerns in the health zone; subjective norms and HIT reliability in the information zone; and HIT self-efficacy in the technology zone—are the predicting factors that form the HITAM with varying ranges of significance and directional relationships. By identifying the core factors that have the largest impact, HITAM is a succinct and powerful model that reevaluates and reorganizes previous findings in the field.

The survey revealed that, although TAM has expanded its utility into various areas and has been successfully implemented, its implementation in the HIT arena has been limited. Furthermore, even in the rapid advancement of information technology and its subsequent impact on health management, a model that captures and predicts various aspects of consumer acceptance is lacking. To address this gap, our research has tried to provide a robust foundation for future HIT research. Yun's similar effort in integrating health information-seeking activities with TAM was limited to the use of the Internet and did not capture the real-life application of HIT. Therefore, by considering the latest developments in HIT, now used more ubiquitously, the model in this work possesses increased real-life applicability and predictive capability.

Figure 3 shows the overall structure of HITAM, which describes health consumers' attitude and behavioral intention when encountering HIT. We discuss five distinct factors in this work: health status, health belief and concerns, subjective norm, HIT reliability and HIT self-efficacy. The first two factors lead to a mediating process of perceived threat toward consumers' health. When consumers consider the possibility of deteriorating health, they may seek the use of HIT to better manage their health. This leads to the consumer determining the perceived usefulness of HIT, and in the case of positive interaction, it results in positive attitude, which ultimately leads to their behavioral intention. The third factor, subjective norm, concerns social pressure or competition within the community. When health consumers discover a given HIT over social network services or other means of online interaction, they are likely to perceive the HIT as useful. This encounter, depending on their experience, leads to positive attitude, and in turn leads to their behavioral intention. The fourth antecedent, HIT reliability, is unique in that it lies in both the information zone and the technology zone. HIT reliability includes quality of output and demonstrability of result, which suggests that it can come from either a direct experience with the technology or an indirect experience, gained through information gathered through other consumers. On the other hand, it also leads to perceived ease of use. This particular mediating process can lead immediately through perceived usefulness, because a user who finds a given information technology easy to use will generally perceive it to be useful as well. The final antecedent, HIT self-efficacy, follows a similar path toward perceived usefulness and perceived ease of use, which can arise from health consumers' confidence in using a given technology. The mediating processes lead to positive attitude, which in the HITAM is a crucial prerequisite to forming behavioral intentions in health consumers.

Using this model, many aspects of health behaviors using HIT can be explained. Thus, in the wake of the exponentially increasing presence of HIT such as the Internet and smartphone apps, the HITAM provides a valuable model of how different interactions with HIT form behavioral intention in health consumers.

Figure 3. The Health Information Technology Acceptance Model (HITAM). HIT = health information technology.



Acknowledgments

This work was supported by a National Research Foundation of Korea (NRF) grant, funded by the Korea government (MEST) (No.2012-0000998).

Conflicts of Interest

None declared.

References

1. Thompson TG, Brailer DJ. The Decade of Health Informaton Technology: Delivering Consumer-Centric and informaton-Rich Health Care: Framework for Strategic Action. Washington, DC: US Department of Health and Human Services; 2004 Jul 21. URL: http://www.providersedge.com/ehdocs/ehr_articles/the_decade_of_hit-delivering_customer-centric_and_info-rich_hc.pdf [accessed 2012-09-26] [WebCite Cache ID 6AxzYlVuB]
2. Wikipedia. 2012 May 16. Health Information Technology URL: http://en.wikipedia.org/wiki/Health_information_technology [WebCite Cache ID 68wZdByXG]

3. Pivovarov R, Crawford M, Kos P, Patil P, Tonellato P. F1000Posters. 2012. Clinical Avatars for Breast Cancer Risk Prediction Creation and Preliminary Bayesian Network Simulations URL: http://f1000.com/posters/browse/summary/401 [WebCite Cache ID 67B1dpg5f]
4. Greene K. Technology Review, Massachusetts Institute of Technology. 2008 Sep 10. Self Surveillance: A New Device Tracks Activity and Sleep Patterns 24-7 URL: http://www.technologyreview.com/communications/21361/page1/ [accessed 2012-04-24] [WebCite Cache ID 67B1mJzA5]
5. Glantz K, Maddock J. eNotes.com. 2010. Encyclopedia of Public Health: Behavior, Health-Related URL: http://www.enotes.com/public-health-encyclopedia/behavior-health-related [WebCite Cache ID 5vAsca9QT]
6. DefinitionOfWellness.com. Definition of Health Behavior URL: http://www.definitionofwellness.com/dictionary/health-behavior.html [WebCite Cache ID 5vArScXtS]
7. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. Manag Inf Syst Q 1989 Sep;13(3):319-339.
8. Davis FD. User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. Int J Man Mach Stud 1993 Mar;38(3):475-487. [doi: 10.1006/imms.1993.1022]
9. Venkatesh V. Determinants of perceived ease of use: integrating control, intrinsic motivation, and emotion into the technology acceptance model. Inf Syst Res 2000 Dec;11(4):342-365.
10. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
11. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. Manag Inf Syst Q 2003 Sep;27(3):425-478.
12. Venkatesh V, Bala H. Technology Acceptance Model 3 and a research agenda on interventions. Decis Sci 2008 May;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
13. Davis FD. A Technology Acceptance Model for Empirically Testing New End-User Information System: Theory and Result [doctoral dissertation]. Boston, MA: Sloan School of Management, Massachusetts Institute of Technology; 1986.
14. Yun EK. Development and Testing of a Model of Consumer's Health Information Seeking Behavior on the Internet [doctoral thesis]. Seoul: Seoul National University; 2008.
15. Becker MH, Rosenstock IM. Advances in Health Education and Promotion: A Research Annual, 1991 (Advances in Health Education and Promotion). Greenwich, CT: Jessica Kingsley Pub; 1987.
16. Wikipedia. 2012 May 8. Health Belief Model URL: http://en.wikipedia.org/wiki/Health_belief_model [WebCite Cache ID 680f0A7ha]
17. Wikipedia. 2012 Apr 23. Theory of Reasoned Action URL: http://en.wikipedia.org/wiki/Theory_of_reasoned_action [WebCite Cache ID 680fPxwj2]
18. Wikipedia. 2012 May 15. Theory of Planned Behavior URL: http://en.wikipedia.org/wiki/Theory_of_planned_behavior [WebCite Cache ID 680fe7mZv]
19. Fishbein M. A theory of reasoned action: some applications and implications. Nebr Symp Motiv 1980;27:65-116. [Medline: 7242751]
20. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
21. Straub D, Limayem M, Karahanna-Evaristo E. Measuring system usage: implications for IS theory testing. Manag Sci 1995;41(8):1328-1342. [doi: 10.1287/mnsc.41.8.1328]
22. Szajna B. Empirical evaluation of the revised technology acceptance model. Manag Sci 1996;42:85-92. [doi: 10.1287/mnsc.42.1.85]
23. Rosenstock IM. Historical origins of the health belief model. In: Becker MH, editor. The Health Belief Model and Personal Health Behavior. Thorofare, NJ: C. B. Slack; 1974:1-8.
24. Rosenstock IM. The health belief model and preventive health behavior. Health Educ Monogr 1974;2(4):354-386.

Abbreviations

HBM: Health Belief Model
HIT: health information technology
HITAM: Health Information Technology Acceptance Model
TAM: Technology Acceptance Model
TPB: theory of planned behavior

Edited by G Eysenbach; submitted 25.04.12; peer-reviewed by T Chiu; comments to author 21.05.12; revised version received 06.06.12; accepted 30.07.12; published 01.10.12.

Please cite as:
Kim J, Park HA
Development of a Health Information Technology Acceptance Model Using Consumers' Health Behavior Intention
J Med Internet Res 2012;14(5):e133
URL: http://www.jmir.org/2012/5/e133/
doi:10.2196/jmir.2143
PMID:23026508

Original Paper

Feasibility of a Wiki as a Participatory Tool for Patients in Clinical Guideline Development

Elvira ME den Breejen[1], MD; Willianne LDM Nelen[1], MD, PhD; Jose M.L Knijnenburg[2]; Jako S Burgers[3], MD, PhD; Rosella PMG Hermens[4], PhD; Jan AM Kremer[1], MD, PhD

[1]Radboud University Nijmegen Medical Center, Department of Obstetrics and Gynecology, Radboud University Nijmegen, Nijmegen, Netherlands
[2]Dutch patient association for people with fertility problems Freya, Wijchen, Netherlands
[3]Dutch college of General Practitioners, Utrecht, Netherlands
[4]Radboud University Nijmegen Medical Center, Department of IQ Healthcare, Radboud University Nijmegen, Nijmegen, Netherlands

Corresponding Author:
Willianne LDM Nelen, MD, PhD
Radboud University Nijmegen Medical Center
Department of Obstetrics and Gynecology
Radboud University Nijmegen
internal postal code 791 P.O. Box 9101
Nijmegen, 5600 HB
Netherlands
Phone: 31 24 3668615
Fax: 31 24 3668597
Email: w.nelen@obgyn.umcn.nl

Abstract

Background: Patient participation is essential in developing high-quality guidelines but faces practical challenges. Evidence on timing, methods, evaluations, and outcomes of methodologies for patient participation in guideline development is lacking.

Objective: To assess the feasibility of a wiki as a participatory tool for patients in the development of a guideline on infertility determined by (1) use of the wiki (number of page views and visitors), (2) benefits of the wiki (ie, number, content, and eligibility of the recommendations to be integrated into the guideline), and (3) patients' facilitators of and barriers to adoption, and the potential challenges to be overcome in improving this wiki.

Methods: To obtain initial content for the wiki, we conducted in-depth interviews (n = 12) with infertile patients. Transcripts from the interviews were translated into 90 draft recommendations. These were presented on a wiki. Over 7 months, infertile patients were invited through advertisements or mailings to formulate new or modify existing recommendations. After modifying the recommendations, we asked patients to select their top 5 or top 3 recommendations for each of 5 sections on fertility care. Finally, the guideline development group assessed the eligibility of the final set of recommendations within the scope of the guideline. We used a multimethod evaluation strategy to assess the feasibility of the wiki as a participatory tool for patients in guideline development.

Results: The wiki attracted 298 unique visitors, yielding 289 recommendations. We assessed the 21 recommendations ranked as the top 5 or top 3 for their eligibility for being integrated into the clinical practice guideline. The evaluation identified some challenges needed to be met to improve the wiki tool, concerning its ease of use, website content and layout, and characteristics of the wiki tool.

Conclusions: The wiki is a promising and feasible participatory tool for patients in guideline development. A modified version of this tool including new modalities (eg, automatically limiting the number and length of recommendations, using a fixed format for recommendations, including a motivation page, and adding a continuous prioritization system) should be developed and evaluated in a patient-centered design.

(J Med Internet Res 2012;14(5):e138) doi:10.2196/jmir.2080

KEYWORDS

Wiki; patient participation; infertility; Web 2.0; guideline development

Introduction

Having patients participate in clinical practice guideline (CPG) development is essential but challenging [1,2]. Their participation is particularly assumed to result in higher-quality guidelines in terms of applicability, acceptability, usefulness, and enhancement of implementation [1-7]. For instance, patient participation is one of the key criteria of the Appraisal of Guidelines Research and Evaluation (AGREE) instrument [8], which is used to assess the methodological quality of guidelines. However, only 25%-50% of CPG developers regularly involve patients [9].

Several practical limitations could explain why patient participation is not common practice in CPG development. First, various methods for patient participation in CPG development can be used, such as conducting in-depth interviews or focus group meetings to explore patients' preferences, asking patients' representatives to comment on drafts of the CPG, or including patients' representatives or patients in the CPG development group [3,6,10-14]. However, practical guidance on how and when to apply these methods is lacking [15]. Second, all methods are restricted to including a selected number of patients or patients' representatives and do not involve a large population of patients. Third, transparently integrating patients' preferences into CPG recommendations is difficult and often unclear [16]. Fourth, organizational (eg, recruitment of participants), financial (eg, costs of patients' education or for conducting focus groups), and sociopolitical barriers (eg, CPG developers' resistance to including patients in the CPG group) may impede patient participation in CPG development [13]. Finally, studies on the effectiveness and impact of patient participation are limited [15]. A new methodology for patient participation in CPG development that enables overcoming most of these drawbacks is thus necessary.

Web 2.0 tools offer opportunities to let nonorganized groups participate in a complex process such as CPG development [17-20]. In particular a wiki, such as Wikipedia, seems to be an easily accessed tool, which enables patients to collaborate in formulating guideline recommendations directly. Ideally, to test the feasibility of such a new method for patient participation in CPG development, an Internet-using young target group such as infertile patients [21-23] is preferred. Infertility is commonly defined as “any form of reduced fertility with prolonged time of unwanted non-conception” [24] and affects approximately 80 million couples worldwide [25,26]. In this study, we applied wiki technology as a participatory tool for patients in the development of a multidisciplinary CPG on infertility and aimed to assess its feasibility.

Methods

Setting

Fertility Care

In the Netherlands, fertility care is mostly publically arranged and provided by various professionals. First, fertility care is provided by the general practitioner and may be part of an initial fertility assessment after a prolonged time of unwanted nonconception. Second, the general practitioner can refer couples to a gynecologist in a general (secondary care) or a university (tertiary care) hospital to complete the fertility assessment, determine a cause of infertility, and define a suitable treatment policy. Third, if a severe male factor is diagnosed, a urologist may be consulted. Furthermore, since infertility has a high emotional and psychological impact, which also interferes with work, a psychologist and occupational physician are regularly engaged in the care pathway. In vitro fertilization and intracytoplasmic sperm injection are provided by 13 licensed hospitals (8 university hospitals, 4 general hospitals, and 1 private clinic). Ovulation induction and intrauterine insemination are performed in all Dutch hospitals. Ovulation induction, intrauterine insemination cycles, and the first three in vitro fertilization or intracytoplasmic sperm injection treatment cycles are reimbursed as part of the basic health care package according to the Health Insurance Act.

Guideline Development

In February 2008, a collaboration of stakeholders (a general practitioner, 2 gynecologists, a urologist/sexologist, a clinical embryologist, a clinical chemical specialist, a medical psychologist, an occupational physician, 2 patients' representatives, and a researcher) was set up to develop a national multidisciplinary paper-based CPG on infertility. CPGs are defined as sets of evidence- or consensus-based recommendations describing optimal patient care to assist health care professionals and patients in clinical decision making [2]. The aim of the CPG was to focus on organizational and patient-centered aspects of fertility care. Two representatives of the Dutch patients' association for infertility, Freya, participated in the CPG development group. However, for direct patient participation in this guideline, we applied a wiki concurrently with this guideline development phase.

Study Objectives

The objective of this study was to assess the feasibility of the wiki as a participatory tool for patient participation in CPG development. The feasibility of the wiki was determined by three end points: (1) use of the wiki and users' characteristics (number of page views and visitors), (2) wiki content quality, particularly the assessment of various aspects of the final set of unique recommendations (ie, number, content, and their eligibility for integration into the CPG) for high-quality fertility care, and (3) wiki system quality (ie, patients' facilitators of and barriers to adoption of this wiki as a participatory tool for direct patient involvement in CPG development, as well as potential suggestions for improvement).

Wiki Tool Development

We developed a conventional wiki website using MediaWiki software and made accessible through the Freya website, called FreyaWIKI [27]. During the preparation phase, we first conducted in-depth interviews to obtain initial content for the wiki. Next, we structured the wiki tool according to the topics of the recommendations derived from the interviews.

Obtaining Initial Content of the Wiki Tool From In-Depth Interviews

To obtain the initial content for this wiki, we first conducted 12 semistructured in-depth interviews with infertile couples during different phases of care, from the first visit to the general practitioner, to (non)pregnant status after completing medically assisted reproduction techniques [28]. Patients visiting outpatient clinics in Nijmegen and Amsterdam were consecutively invited to participate through an information letter. Subsequently, 1 researcher (EB) obtained the final consent by telephone. Participants were asked to specify perceived bottlenecks in their fertility care pathway, using an interview guide including the main treatment stages of their fertility care pathway (eg, treatment by a general practitioner, gynecologist, or urologist). Interviews were audiotaped and transcribed verbatim. Next, a researcher (EB) and the chief executive of Freya (JK) independently translated these bottlenecks into draft patient recommendations. These draft recommendations were formulated as “I want my doctor to....” Consensus on the formulation of patient draft recommendations was reached through discussion.

Structuring the Wiki Tool

Division of the draft recommendations into sections and subsections determined the structure of the wiki. Draft recommendations were divided into 4 sections (EB,JK), consisting of 3 sections referring to the care delivered by the 3 most involved professionals and a general section for recommendations important to all professionals: general care, care delivered by a general practitioner, gynecological care, and urological care. To provide more structure in the wiki, the draft recommendations in each of these 4 sections were subdivided into 8 subsections (EB, JK) based on aspects of care that are known to be important to infertile patients: 3 medical-technical aspects (ie, examination, therapy, and referral), 4 patient-centered aspects (ie, organization of care, information provision, communication, and staff attitudes), and 1 general aspect (ie, general) [29]. These subsections were presented on the wiki in the following order: general (recommendations in general and those that don’t apply to other care aspects), information provision (recommendations on oral and paper-based information provision), organization (recommendations on the organization of fertility care, for example, adjustment of care between different health care professionals, accessibility of care), staff attitudes (recommendations on the attitude of health care professionals toward the patient, for example, having empathy), communication (recommendations on communication between the health care professional and the patient), examination (recommendations on examinations during fertility care), therapy (recommendations on therapy, namely infertility treatment by, for example, in vitro fertilization), and referral (recommendations on referral from one health care professional to another, for example, from a general practitioner to a gynecologist). Discrepancies in division and subdivision of the draft recommendations were resolved through discussion.

Patient Participation in CPG Development

Recruiting Participants

We recruited participants for the wiki evaluation through mailings to members of Freya, the Dutch patients’ association for infertility; advertisements in Freya’s quarterly journal; links on websites of Freya and the professional societies (eg, general practitioners, gynecologists, urologists, and clinical embryologists); and links in social media (eg, Hyves, Twitter, and Facebook). In addition, we sent advertising posters to all 103 clinics offering fertility treatments in the Netherlands for their waiting rooms.

Obtaining Recommendations From Wiki Participants

Formulating Recommendations

From May to December 2008, we presented the draft recommendations for fertility care on the wiki. Patients were invited to modify or refine these recommendations and to add new recommendations. During this process, we asked patients to subscribe voluntarily through an email address and to provide background characteristics for study purposes. After 2 months, when the number of recommendations started to increase, patients and patients’ representatives requested us to add 2 sections to the existing structure of the wiki: 1 regarding the care delivered by the laboratory (eg, recommendations regarding semen analysis), and 1 regarding the care delivered by the remaining professionals who were not represented in a separate section (eg, recommendations regarding the medical psychologist). Hence, we added 2 sections to the wiki: laboratory and remaining. Next, we recategorized recommendations from the general section regarding care delivered by the laboratory or care delivered by professionals other than the general practitioner, gynecologist or urologist. After this restructuring of the wiki, the general section contained only recommendations on fertility care in general, thus not referring to the care delivered by 1 of the professionals involved.

Modifying Recommendations

After 7 months, we modified the recommendations in several steps. First, we removed duplicate recommendations. Then, if necessary, we moved recommendations into the appropriate sections (EB, JK). Since all recommendations in the remaining section turned out to be more suited to other sections, we eliminated this section. Next, 2 researchers (EB, WN) and the chief executive of Freya (JK) independently assessed the implementability of all recommendations using the Guideline Implementability Appraisal (GLIA) instrument [30]. Discrepancies were discussed and resolved through consensus. Based on the results of this assessment, the recommendations were independently textually refined or modified by a researcher (EB) and the chief executive of Freya (JK). Finally, after consensus was reached on the final formulation, we reentered the recommendations into the wiki.

Prioritizing Recommendations

All patients visiting the wiki website were invited to prioritize their top 5 (modified) recommendations in each section (for the laboratory section, we asked them to identify their top 3 due to the small number of recommendations). This prioritization was privately conducted by assigning 5, 4, 3, 2, and 1 points for the most important recommendations for determining high-quality fertility care for each of the 5 sections and independently from the subsections.

Assessing Eligibility of the Selected Recommendations

Initially, the CPG development group had intended to integrate this final top selection of patients' recommendations directly into the CPG. However, before integrating these recommendations, the entire CPG development group (n = 11) assessed the eligibility of the recommendations for inclusion in terms of the scope of the guideline.

Evaluation of the Wiki

To evaluate the feasibility of the wiki, we performed a multimethod evaluation study including three components [31]. First, to assess the ability to involve large and diverse patient populations compared with other methods such as interviews, we evaluated wiki use and users' characteristics. Second, we evaluated wiki content quality (ie, recommendations) and, third, wiki system quality (eg, ease of use, layout), identifying factors that could potentially influence adoption of the wiki (barriers and facilitators) as well as potential factors for improvement.

Evaluation of Wiki Use and Users' Characteristics

Data on actual use of the wiki (eg, number of unique visitors, page views) were generated through log files on the website of the patient association (Freya). Unique visitors were determined by IP address logged and stored on the website.

Evaluation of Wiki Content Quality

To evaluate the content quality of the wiki, we assessed various aspects of the final set of unique recommendations, particularly the number of recommendations, their content, and their eligibility for integration into the CPG for high-quality fertility care.

Evaluation of Wiki System Quality

To evaluate the quality of the wiki system and to identify facilitators, barriers, and potential areas of improvement, we conducted an online questionnaire. To gain insight into the thoughts underlying the resulting factors that formed potential facilitators of or barriers to adoption of the wiki and aspects of improvement, we conducted in-depth interviews with wiki users who completed the evaluation questionnaire.

Online Evaluation Questionnaire

During the prioritization phase, patients visiting the wiki website were invited to complete an online evaluation questionnaire. This questionnaire included items regarding users' background characteristics (eg, age, type of infertility), use of the wiki (eg, number of visits), and factors that could potentially influence adoption of the wiki (quality of the wiki website, satisfaction, and net benefits) based on the relevant evaluation factors derived from the Human, Organization, and Technology-fit framework [32]. Questions on the potential influencing factors were grouped into 5 sections: ease of use of the wiki website, layout of the wiki website, value of the wiki methodology as a participatory tool for CPG development, content of the wiki website, and experienced privacy on the wiki website. Patients were asked to rate 22 accompanying positively formulated statements on these factors on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree) (Multimedia Appendix 1). After each section, patients were invited to comment. Next, patients were asked to describe their three advantages and disadvantages of the wiki website and potential areas of improvement. Finally, patients were asked for their willingness to participate again in a similar project and for their intention to recommend this wiki.

In-Depth Interviews With Wiki Users

We first summarized the identified influencing factors on adoption of the wiki website and suggested potential areas of improvement. Next, we translated these into a topic list to guide the in-depth interviews. To get both confirmation of and saturation in the thoughts underlying the facilitators of and barriers to adoption and potential areas of improvement of the wiki, 1 researcher (EB) conducted semistructured in-depth interviews with wiki users by telephone. Participants in the questionnaire survey who left their email address were randomly recruited by email. The first part of the interview consisted of open-ended questions, related to thoughts underlying the identified influencing factors on adoption and potentials for improvement of the wiki. Next, patients were asked for additional influencing factors and suggestions for improving the wiki. Recruitment continued until saturation of data was achieved. Regarding the starting and stopping criteria according to Francis and colleagues [33], we started with 2 interviews and aimed to repeat cycles of 2 interviews until we obtained no new data. If data saturation was achieved, an additional interview was conducted to attain data saturation.

Data Analysis

We used SPSS 16.0 for Windows, Data Entry 4.0 (IBM Corporation, Somers, NY, USA) to perform descriptive statistical tests on the background characteristics of the wiki participants and to analyze patients' top rankings of the recommendations. The final top selection of recommendations in each section was determined by identifying those with the highest sumscores derived. For analyzing the results of the online evaluation questionnaire, we grouped the responses on the 5-point Likert scale into the categories agree (scores 1and 2), neutral (score 3), and disagree (scores 4 and 5). Items were a priori identified as facilitators of adoption if >50% chose agree (scores 1 and 2) and as barriers to adoption of the wiki website if >50% of the evaluators chose disagree (scores 4 and 5). We used the reported top three advantages and disadvantages and the potential areas of improvement of the wiki to determine the frequency of occurrence of each aspect. We conducted an initial content analysis of all free-text responses to the questionnaire, to determine additional points to be improved (EB, WN).

Qualitative Analysis of the Interviews

All interviews were audiotaped and transcribed verbatim. Data were analyzed iteratively and thematically across accounts (EB, JK) [34], according to the relevant factors of the evaluation framework, as used in the questionnaire to identify barriers to and facilitators of adoption and potential areas of improvement of the wiki [32]. Another researcher (WN) independently checked the coding framework and analysis.

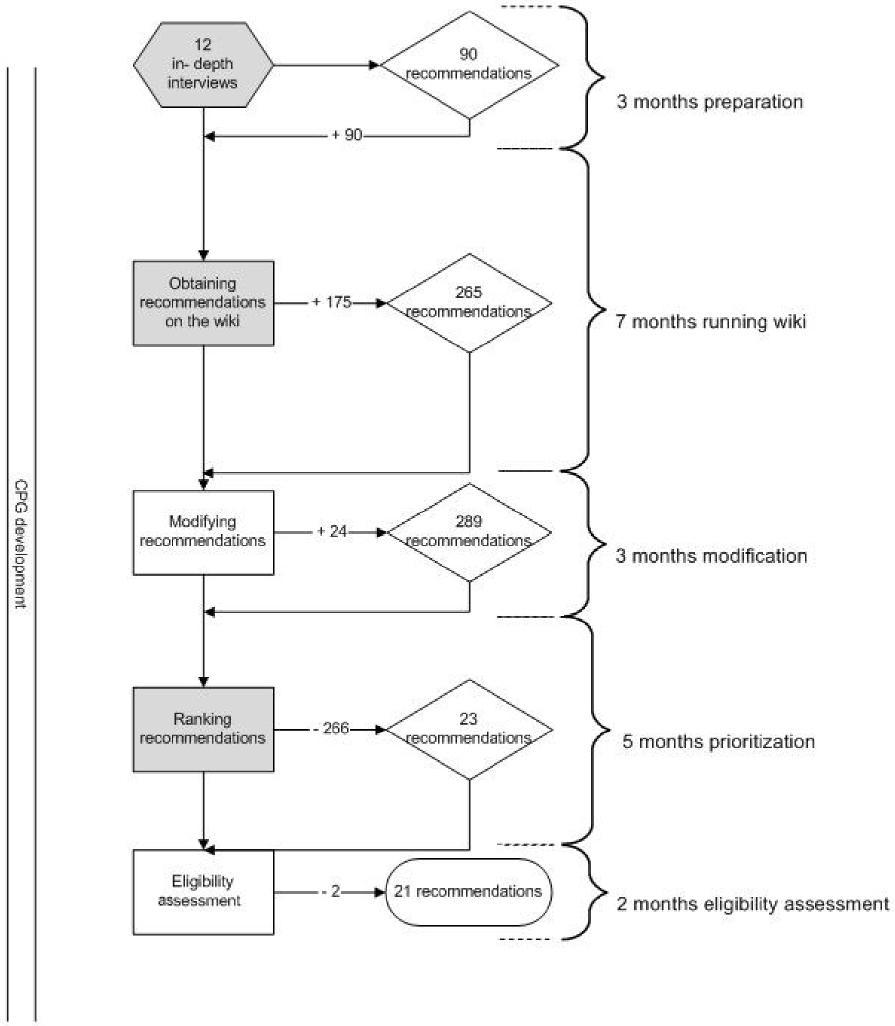
Results

Wiki Tool Development

In-Depth Interviews

From the transcripts of 12 in-depth interviews with infertile patients, we translated the perceived bottlenecks into a set of 90 draft patient recommendations and entered them into the wiki (Figure 1). For example, patients perceived a bottleneck in that appointments were possible only during working hours instead of also during the evening, which resulted in difficulties with work. The resulting draft recommendation was formulated as “I want the hospital to provide possibilities to make appointments during evening hours.” Other examples of the bottlenecks mentioned were the variation between hospitals’ laboratories in performing a semen analysis, unavailability of separate waiting rooms for pregnancy and infertility consultations, and gynecologists’ lack of empathy.

Figure 1. The process of obtaining recommendations for clinical practice guideline (CPG) development.



Structure of the Wiki

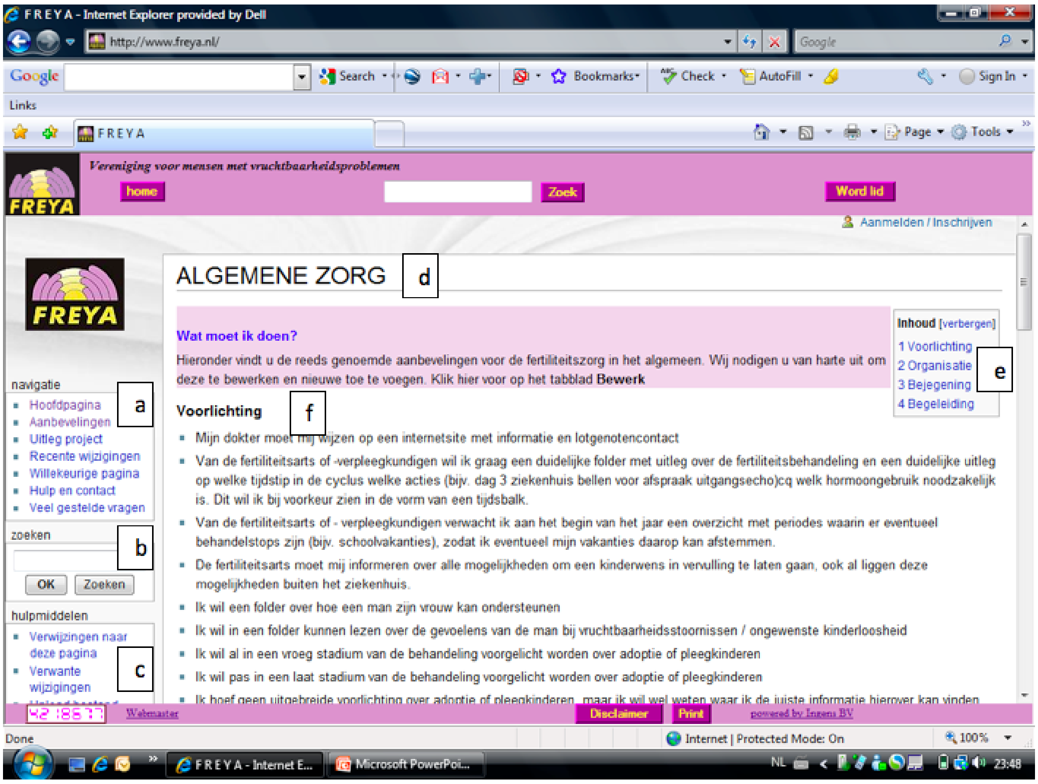
FreyaWIKI was structured through the division of recommendations into 6 sections. Each of these sections was subdivided into 8 subsections (Figures 2 and 3, see Multimedia Appendices 2 and 3 for translations).

Patient Participation in CPG Development

Wiki Use and Users’ Characteristics

During 7 months of access, 36,473 wiki pages were viewed. We identified 298 unique users, including 81 registered users who provided background characteristics (Figure 4). The majority of them were female 78/81 (96%), highly educated 54/81(67%), and middle aged (mean 33 years). Median duration of infertility was 30 months (range 0–71 months). More than half 43/81(53%) underwent medically assisted reproduction techniques during the period of their visit. Another 14% (n = 11) stayed childless despite treatment.

Figure 2. Screenshot of recommendations on FreyaWIKI.



ALGEMENE ZORG

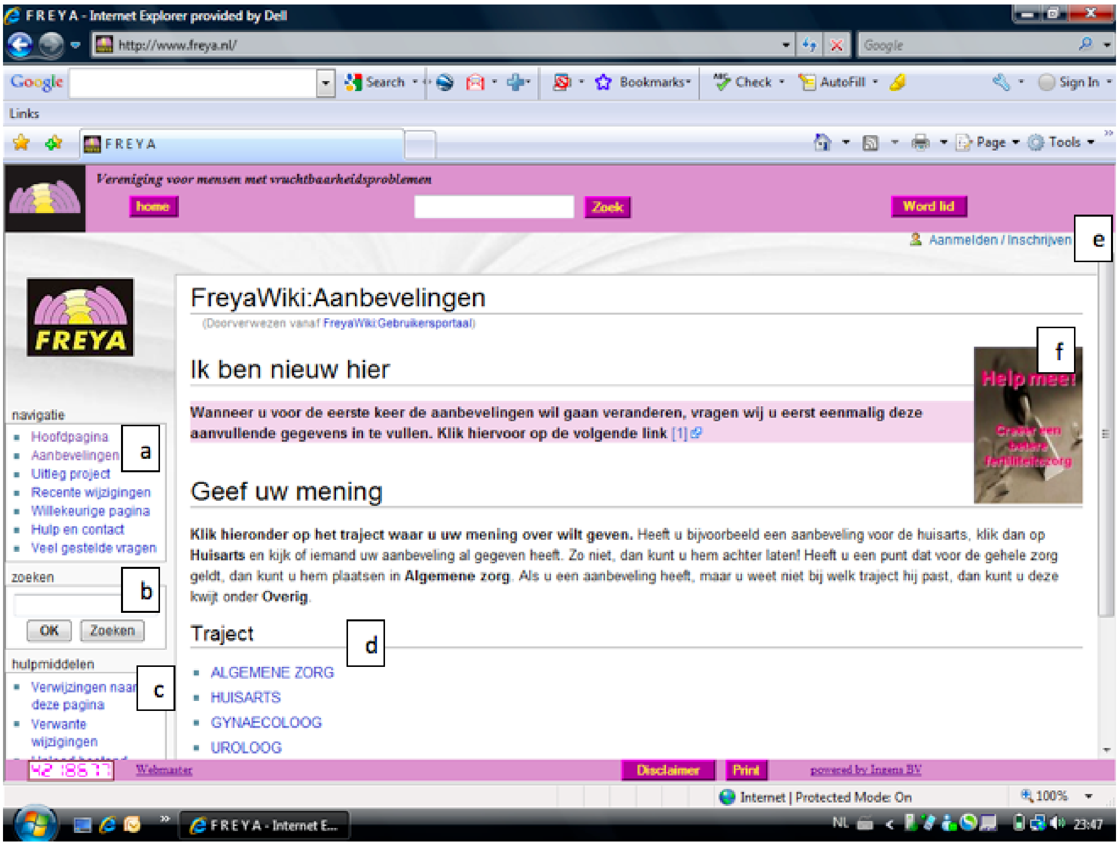
Wat moet ik doen?

Hieronder vindt u de reeds genoemde aanbevelingen voor de fertiliteitszorg in het algemeen. Wij nodigen u van harte uit om deze te bewerken en nieuwe toe te voegen. Klik hier voor op het tabblad **Bewerk**

Voorlichting

- Mijn dokter moet mij wijzen op een internetsite met informatie en lotgenotencontact
- Van de fertiliteitsarts of -verpleegkundigen wil ik graag een duidelijke folder met uitleg over de fertiliteitsbehandeling en een duidelijke uitleg op welke tijdstip in de cyclus welke acties (bijv. dag 3 ziekenhuis bellen voor afspraak uitgangsecho)cq welk hormoongebruik noodzakelijk is. Dit wil ik bij voorkeur zien in de vorm van een tijdsbalk.
- Van de fertiliteitsarts of - verpleegkundigen verwacht ik aan het begin van het jaar een overzicht met periodes waarin er eventueel behandelstops zijn (bijv. schoolvakanties), zodat ik eventueel mijn vakanties daarop kan afstemmen.
- De fertiliteitsarts moet mij informeren over alle mogelijkheden om een kinderwens in vervulling te laten gaan, ook al liggen deze mogelijkheden buiten het ziekenhuis.
- Ik wil een folder over hoe een man zijn vrouw kan ondersteunen
- Ik wil in een folder kunnen lezen over de gevoelens van de man bij vruchtbaarheidsstoornissen / ongewenste kinderloosheid
- Ik wil al in een vroeg stadium van de behandeling voorgelicht worden over adoptie of pleegkinderen
- Ik wil pas in een laat stadium van de behandeling voorgelicht worden over adoptie of pleegkinderen
- Ik hoef geen uitgebreide voorlichting over adoptie of pleegkinderen, maar ik wil wel weten waar ik de juiste informatie hierover kan vinden

Figure 3. Screenshot of the FreyaWIKI homepage.



FreyaWiki:Aanbevelingen

(Doorverwezen vanaf FreyaWiki:Gebruikersportaal)

Ik ben nieuw hier

Wanneer u voor de eerste keer de aanbevelingen wil gaan veranderen, vragen wij u eerst eenmalig deze aanvullende gegevens in te vullen. Klik hiervoor op de volgende link [1]

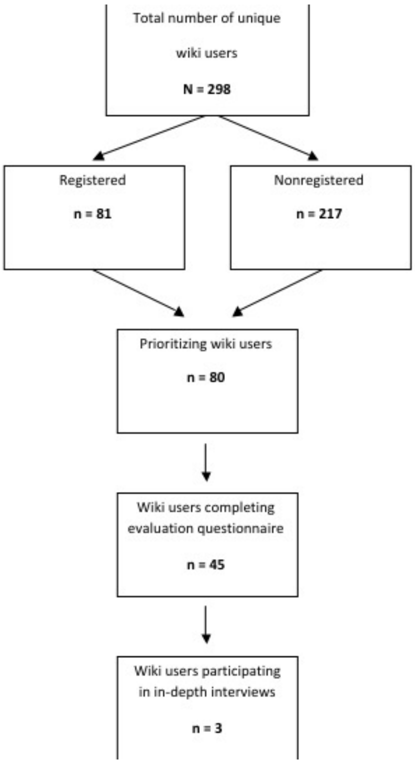
Geef uw mening

Klik hieronder op het traject waar u uw mening over wilt geven. Heeft u bijvoorbeeld een aanbeveling voor de huisarts, klik dan op **Huisarts** en kijk of iemand uw aanbeveling al gegeven heeft. Zo niet, dan kunt u hem achter laten! Heeft u een punt dat voor de gehele zorg geldt, dan kunt u hem plaatsen in **Algemene zorg**. Als u een aanbeveling heeft, maar u weet niet bij welk traject hij past, dan kunt u deze kwijt onder **Overig**.

Traject

- ALGEMENE ZORG
- HUISARTS
- GYNAECOLOOG
- UROLOOG

Figure 4. Flow of wiki participants through the study.



Wiki Content Quality

Overall, we collected 265 recommendations and modified them into 289 unique recommendations (Figure 1). After patients (n = 80) prioritized the recommendations by ranking the top 5 or top 3 in each section (we had eliminated the remaining section) according to their importance to high-quality fertility care, we selected 23 recommendations (4 sets of top 5 and 1 set of top 3) for eligibility assessment by the CPG development group (Table 1). We excluded 2 insurance-related recommendations, since they did not meet the scope of the CPG. The CPG development group accepted all of the remaining 21 recommendations, which were directly integrated into the CPG. More than half (n = 11) of the final set of recommendations concerned the organization of care. Similar to the quality assessment scale used in evidence-based recommendations (levels of evidence A-D) [35], a level of P (patients) was provided for the patients' recommendations and formulated as " Patients would like to…." Participants were informed by email, on the wiki website, and through the websites of Freya and the professional associations of the final CPG that included the untouched eligible recommendations of the patients.

Evaluation of Wiki System Quality

Online Evaluation Questionnaire

Of the 80 patients who participated in the prioritization, 45 completed the questionnaire. Of these, 53% (n = 24) visited FreyaWIKI for the first time while completing the questionnaire, and 93% (n = 42) had never worked with a wiki, other than this one, before. Other background characteristics of the respondents are presented in Table 2. Facilitators of adoption of the wiki, defined as >50% of respondents agreeing (scores 1 and 2) to the relevant statements, were not identified. Barriers, defined as >50% disagreeing (scores 4 and 5) to the relevant statements, were identified in 3 of the 5 sections: ease of use, content of the website, and value of the wiki methodology (Table 3). In decreasing order of the proportion of evaluators who disagreed with the relevant statement, the main identified barriers concerned the findability (82%) and accessibility (78%) of the website, and the suitability of this wiki for obtaining recommendations for CPG development (71%).

Reported advantages of the wiki were the privacy they experienced on the website, the structure of the website linking recommendations to sections on care delivered by fertility professionals, ease of navigation through the website, and the additional value of the wiki website as a source of information and as an opportunity to provide feedback to the care services.

Reported disadvantages of the wiki concerned the content of the wiki website, in terms of the unstructured recommendations not being formulated in a similar way, too much content being visible on one screen, and the nonattractive layout of the wiki website.

The main potential areas of improvement were providing information on treatment options and causal factors of infertility on the wiki website, broadening the marketing of the wiki by placing advertisements in commercial magazines, and communicating information on related activities (Table 4). Overall, 98% of the patients said they would recommend the website and 84% would participate again in a similar project.

In-Depth Interviews

Overall, 11 of the 30 patients who gave their email address in the evaluation questionnaire agreed to participate in the interviews. We conducted 3 interviews. All 3 interviews confirmed barriers to adoption as well as suggestions to improve the wiki, and saturation of the related underlying thoughts was reached (Table 4). All patients reported problems with formulating a recommendation and expressed their wish to add a personal touch to the recommendation (eg, to explain why something should be done). The introduction of a motivation page, where patients could describe why they formulated a recommendation, might meet this request. Patients also embraced the missing community feeling as mentioned in the evaluation questionnaire. Introducing a monthly newsletter and automatically sending an email to the person who made the recommendation were suggested. All 3 interviewees regarded the website as a valuable source of information, rather than as a tool for modifying recommendations for CPG development. They mentioned that the content of the wiki had been helpful to them in searching for information on experiences regarding infertility treatment and in searching for recognition of their own experiences. Challenges faced by users in understanding the purpose of the website would be addressed by clearer instructions.

Table 1. Final set of the patients' top-5 and top-3 recommendations (n = 23) for the 5 sections, ranked by importance to the quality of fertility care as formulated on the wiki website.

Section, rank, and recommendation		Subsection (aspect of care)[a]
General care		
1	I want insurance companies to reimburse six attempts at in vitro fertilization[b]	General
2	I want insurance companies to start counting in vitro fertilization attempts only after oocyte retrieval or even after embryo transfer has been performed[b]	General
3	I want my doctor to practice empathy, instead of only working on the technical or financial part	Staff attitudes
4	I want the hospital to have separate waiting rooms for pregnant women and couples being treated for infertility	Organization
5	I want to be able to arrange appointments during the daytime as well as in the evenings	Organization
General practice care		
1	I want my gynecologist and GP[c] to have good communication, so my GP knows what is going on with us	Referral
2	I want my GP to make a referral immediately after we have been trying to conceive for a year	Referral
3	I want to have my first medical consultation with my gynecologist within 1 month after referral.	Organization
4	I want my GP to be informed of possible causes of infertility, in both women and men	General
5	I want my GP to pay attention to nonmedical issues, such as stress, anxiety, relational problems, and sexuality	Communication
Gynecologic care		
1	I want also to be able to receive treatments on weekends (Saturdays and Sundays)	Organization
2	I want all members of the fertility treatment team to apply one policy regarding my infertility treatment	Organization
3	I want my gynecologist to inform me of all possible fertility treatment options, even if these are outside the hospital	Information provision
4	I want my gynecologist to inform me about the different phases of treatment and their expected time span	Information provision
5	I want assisted hatching to be possible or available in the Netherlands	Therapy
Urologic care		
1	I want my urologist and gynecologist to have good communication	Organization
2	I want to be informed of the investigations that are to be performed by the urologist	Examination
3	I want to have a permanent urologist who is specialized in infertility	Organization
4	I want to have a consultation with a urologist within 1 month after referral	Organization
5	I want my urologist to involve my partner in the conversation	Communication
Laboratory		
1	I want to be informed as soon as possible when our embryos do not divide correctly	Organization
2	I want Dutch laboratories to share protocols and learn from each other's experiences	Organization
3	I want to be informed of the causes of nonviability of our frozen embryos, if appropriate	Organization

[a] Subsections were derived from the website's structure and defined by the user.

[b] Recommendation was excluded, since it fell out of the scope of the clinical practice guideline.

[c] General practitioner.

Table 2. Background characteristics of respondents (n = 45) to the evaluation questionnaire.

Characteristic	Data
Gender, n (%)	
Male	0 (0%)
Female	45 (100%)
Age (years), mean (SD)	35 (5.24)
Type of infertility, n (%)	
Primary	15 (33%)
Secondary	30 (67%)
Duration of infertility (months), median (range)	36 (0–71)
Current phase in fertility care, n (%)	
Gynecologic	19 (42%)
No pregnancy after fertility treatment	8 (18%)
Pregnant achieved by fertility treatment	4 (9%)
Unknown	14 (31%)
Level of education, n (%)	
Low	0 (0%)
Intermediate	14 (31%)
High	31 (69%)
Membership in Freya, n (%)	24 (53%)

Table 3. Patients’ barriers to adoption of the wiki (n = 45).

Factor influencing adoption of the wiki	Proportion disagreeing with the factor[a]	
	n	%
Ease of use of the website		
Findability of the website	37	82%
Accessibility of the website	35	78%
Clarity of log-in location on the website	27	60%
Clarity on the goal of the website	28	62%
Clarity on instructions for using the website	24	53%
Efficiency of the website (ie, speed at which the website enabled users to accurately and successfully add and modify recommendations)	24	53%
Content of the website		
Comprehensiveness of the clarifying text on the website	30	66%
Satisfaction with the content of the formulated recommendations	25	56%
Usefulness of clustering recommendations into sections in searching for existing recommendations	23	51%
Similarity between formulated recommendations and participants’ actual opinions on fertility care	23	51%
Value of using the wiki		
Suitability of the wiki for obtaining recommendations for clinical practice guideline development	32	71%
Ease of using the wiki	24	53%
Accessibility of the wiki	27	60%

[a] Number (%) of participants who rated the positively formulated statements on the evaluation factors as disagree (scores 4 or 5).

Table 4. Participants' (n = 45[a]) suggestions for improving the wiki website.

Aspect of improvement	Respondents suggesting the aspect		Sample quotes (translated from Dutch) from in-depth interviews (I) and online questionnaires (Q)
	n	%	
Usability of the website			
Findability of the website	10	22%	Q:*Hard to find* Q: *I think it is awkward that the website is only findable through the Freya website* I: *I wouldn't know how to find the website, unless through the Freya website*
Accessibility of the website	2	4%	I: *I was unable to find the log-in location or request a new password*
Content of the website			
Comprehensiveness of clarifying text	1	2%	Q: *Unclear*
Clearness of description of the goal of the wiki	4	8%	I: *The description is a bit unclear; therefore, I previously thought to check it more precisely, but I still haven't done this* I: *I had not concluded that the recommendations were directly integrated in a professional guideline*
Clearness of instructions for use	1	2%	
Satisfaction with formulated recommendations	8	16%	I:*...but there are recommendations I am not satisfied with, I would suggest that participants can prioritize recommendations that they are satisfied with in an earlier stage, then you only have to list the most important recommendations in one screen*
Similarity between actual preferences and recommendations	4	8%	Q: *I would like to see why a specific recommendation was formulated, separately from the recommendation* I: *There are too many recommendations on the website, but there are recommendations I am not satisfied with. I would suggest that participants can prioritize recommendations that they are satisfied with*
Clarity of the structure in which recommendations are placed on the website	30	66%	I: *Structure is good but the provided sections are incomplete, for example the care provided by a psychologist or other forms of mental counseling. Psychosocial concerns are always underestimated in fertility care* Q: *The used structure is good, but for searching an existing recommendation it would be valuable to add a search function to the website*
Relationship between length and number of recommendations and their presentation on one screen	32	71%	Q: *There are too many recommendations on the website* I: *Recommendations are too long, sometimes it's more like a story, which is very interesting, but I wonder if the doctors are taking this as serious input to a guideline* Q: *The prioritization is hard due to the large number of recommendations*
Education provision on the website	19	42%	Q: *It might be valuable if the website provides usable links to high-quality websites* Q: *Information on treatment options might enrich the website* Q: *I would like to find information on causal factors of infertility* Q: *Practical information about compensations for treatment per insurance company, regional psychological services, plural miscarriages, infertility, and referral*
Characteristics of the wiki			
Usability of wiki methodology	6	13%	Q: *The website is not user friendly...the number of visible recommendations makes it unclear* *Q: Recommendations given contain too many words* I: *I really don't have a clue about what constitutes a high-quality recommendation* I: *It would be valuable to apply an automatic program, through which patients are able to formulate recommendations*
Accessibility of wiki methodology			
Efficiency of wiki methodology	5	11%	Q: *Prioritizing is hard and not efficient in this stage, the list of recommendations is too long* I: *The efficiency might be improved if you ask patients immediately after formulating a recommendation to prioritize the most important recommendations*

Aspect of improvement	Respondents suggesting the aspect		Sample quotes (translated from Dutch) from in-depth interviews (I) and online questionnaires (Q)
	n	%	
Layout of the website			
Impression of the layout	33	73%	Q: *Nonattractive/not a modern/not a fashionable website* Q: *The layout is not from today* Q: *Looks unprofessional*
Communication with wiki users			
Marketing	6	13%	Q: *This good initiative requires a better marketing approach to reach more participants*
Community feeling of the wiki	3	6%	I: *More communication on related activities and results will increase the number of patients that will come back* Q: *Effect of the recommendations on the guideline is unclear*

[a] 45 participants completed the online evaluation questionnaire, of whom 3 participated in the in-depth interviews.

Discussion

Principal Results

In this study, we showed that the wiki is a feasible tool to ensure active patient participation in the development of a Dutch multidisciplinary CPG on infertility. The high numbers of page views (36,473), unique visitors (298), and recommendations formulated (289) implies patients' willingness and ability to contribute to CPGs through a wiki-based method. We also showed that such a wiki is a useful information source for patients.

Second, we gained a final set of 21 selected recommendations, which were assessed as being eligible to be integrated directly and transparently into the CPG. Third, patients had positive views on the experienced privacy, ease of navigation, divisional structure of the wiki, and its potential befits. A total of 98% of the patients would recommend the website and 84% would participate again in a similar project. This study also provided some important suggestions to improve this participatory tool for patients in the development of CPGs, concerning ease of use, content and layout of the website, and characteristics of the wiki tool.

Comparison With Existing Techniques

Several studies on specialized medical wikis (eg, wikis that fall outside the scope of a general encyclopedia) have been published, but most particularly focus on education of medical students [36] or collaboration between health care professionals [37,38], rather than on patients, and did not include a process evaluation. Only Gupta and colleagues [39] and Archambault [40] involved a group of preselected patients as well as professionals in the development of an asthma action plan through a wiki. However, results are premature, since this study was conducted over a very short time period (weeks), and a wiki needs more time to build content (approximately 7-8 months) [41]. Furthermore, Gupta and colleagues' and Archambault's evaluation of the wiki tool was not focused on patients' experiences and was less extensive than our multifaceted approach to gaining insight into patients' barriers to adoption of our wiki. In this study, we involved a large number of patients (298), which cannot be realized using traditional methods, such as focus groups, in which participation is generally restricted to a maximum of 8 participants [42]. We even assessed the final selection of top recommendations for their eligibility for direct integration into the CPG. Thus, the patients' contribution to the CPG was clearly illustrated by integrating their recommendations in their entirety, indicated by the new P level (Patients). We also addressed other practical limitations of the methods used to enhance patient participation in CPG development, such as organizational (eg, recruitment of participants), financial (eg, travel costs), and sociopolitical (eg, professional resistance to including patients in CPG development group) constraints.

Professionals and patients' representatives could also use the wiki and had the opportunity to informed themselves about patients' views and to bring up content for discussion in the CPG development group. According to the results of the evaluation questionnaire and the interviews, this content was also helpful to patients as an information source, which may also explain the relatively large number of page views. Although providing information was not the initial goal of this wiki, its relevance is in agreement with published literature on conventional wikis [19] and with European patients' perception of the importance of the Internet as a source of information [43]. Hence, this unintended consequence concurrently yields challenges for improvement and might be aided by providing clearer instructions for use and description of the goal of the wiki, but also addresses important implications for future studies in this field. Next to the informational value of formulated recommendations for high-quality care, attention should be paid to useful links to relevant websites that may potentially attract more patients to the wiki website and increase the chances for adoption of an improved version of the wiki.

Although drawbacks to active patient participation methods were reduced, this study drew attention to some other potential implications derived from patients' suggestions that might improve the use of a future medical dedicated wiki exclusively for actively involving patients in CPG development. First, structuring recommendations and limiting the number and length of recommendations to presentation on one screen may improve usability [44]. Second, using a fixed format in the formulation of recommendations, based on relevant items of the GLIA instrument, may not only improve usability and accessibility of the wiki [30,44], but may also improve the efficiency of the wiki and the usefulness of recommendations in being integrated directly into the CPG. Introducing a motivation page might give patients the opportunity to add a personal touch to the recommendation. Third, a prioritization system, continuously refining the similarity between patients' perspectives and the top5 recommendations (eg, by rating recommendations after every contribution), could improve the tools' efficiency by avoiding separate prioritization of recommendations and could improve patients' satisfaction with the highest-rated recommendations. This modality would also allow more flexible use by CPG developers at the time of their choosing. In addition to the suggested modalities, some known refinements in overall usability (eg, findability, prominent log-in location), content (comprehensiveness of text), and layout of the website might improve use of the wiki and would be reduced by repeated cycles of design, evaluation, and redesign [45,46]. Furthermore, a user-centered design, in which patients codevelop such new modalities, may improve future implementability and provide chances for local adaptation of a redesigned wiki website [47,48]. Both the feasibility of a wiki as a participatory tool for patients in the development of CPGs and the recommendations for future wiki-based initiatives illustrate the value of eHealth. With this in mind, numerous participatory applications based on wikis are conceivable and may be valuable in various fields of research. In the field of guideline development, guideline-derived initiatives actively involving patients in the development of patient information leaflets or treatment action plans, in addition to fully online-based CPGs, may also benefit from our results. Finally, our results add to the knowledge base about wikis in health care [49].

Limitations

This wiki has been tested in the field of infertility care, representing a relatively young target group [50]. More than 98% of this group use the Internet [21]. This participant characteristic is associated with more frequent health-related Internet use [51-53]. Therefore, the participants in our study were an ideal subgroup for testing and evaluating a wiki-based method, which argues against the generalizability of our findings to other patient groups. Nevertheless, health-related Internet use in Europe is increasing over time [54]. Hence, it seems to be a question of time until older people or their caregivers, or both, will be using such tools [55]. Furthermore, this feasibility study provided an important exploratory evaluation component, which resulted in valuable information for future studies in this field but also had certain limitations. First, based on the results of a recent systematic review from Gagnon and colleagues [56], we acknowledge that the items used in our evaluation questionnaire might be incomplete. However, the results of our study add to those from the limited number of previously published studies on patients' facilitators of and barriers to adoption of eHealth applications [57,58]. Second, the heuristics used were not based on a validated questionnaire and were too limited for drawing conclusions on the usability that patients perceived. Therefore, a next step in future development of a wiki-based participatory tool for patients in CPG development should be to include a broader evaluation of the potentially influencing factors on adoption, including more organizational factors and a heuristic evaluation.

Third, the participation rate in the evaluation of the wiki might have subjected our study to a participation bias of potentially the most motivated wiki users. However, this is a known limitation in the active use of wikis in general: the most motivated users provide most of the content [19]. Finally, this feasibility study did not assess the representativeness of either the participants or the final set of recommendations in the wiki.

Conclusions

The wiki is a promising and feasible tool to actively involve patients in CPG development. To improve the tool's ease of use and practical aspects to enhance direct integration of recommendations into the CPG, a more specialized and refined wiki should be developed. This should include new modalities, such as automatically shortening the number and length of recommendations, using a fixed format for formulation of recommendations, using a continuous prioritization system for selection of the most important recommendations, and including a separate motivation page. Furthermore, in the development, attention should be paid to the informational character of such a wiki. To improve future implementability, a modified tool should preferably be codeveloped and evaluated by patients in a user-centered design study. Furthermore, representativeness of patients and recommendations should be integrated into this next phase.

Acknowledgments

The authors thank all participants of FreyaWIKI.

Funding: This project was funded by the Netherlands Organization for Health Research and Development (ZonMW), grant no. 150020015, The Hague. This open access publication was funded by the Dutch Organization for Scientific Research (NWO), grant no. 036.001.927. Both funders did not have any role in the design and conduct of the study, the collection, management, analysis, and interpretation of data, or in the preparation, review, or approval of the manuscript.

All researchers are independent from the funders.

Previous presentation/report: The content of this paper has not been copyrighted, published, or presented previously and is not under consideration for publication elsewhere.

Ethics approval: In this study, ethics approval was not required. However, all participants gave informed consent before taking part. All observational data and transcripts of interviews were anonymized.

Authors' Contributions

EB was guarantor, designed the study, modified the recommendations, conducted and (qualitatively) analyzed the online evaluation questionnaire and in-depth interviews, interpreted the analysis of the use of the wiki and its evaluation, and drafted the paper. WN designed the study, checked the analysis, interpreted the analysis, and drafted the paper. JK designed the study, modified the recommendations, interpreted the analysis, and drafted the paper. JB designed the study, interpreted the analysis, and drafted the paper. RH designed the study, interpreted the analysis, and drafted the paper. JAMK led the research team, designed the study, was principal investigator, interpreted the analysis, and drafted the paper. All authors reviewed consecutive drafts of the paper and gave their final approval of the version to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Constructs of the online evaluation questionnaire.

[PDF File (Adobe PDF File), 50KB - jmir_v14i5e138_app1.pdf]

Multimedia Appendix 2

Screenshot of recommendations on FreyaWIKI.

[PDF File (Adobe PDF File), 221KB - jmir_v14i5e138_app2.pdf]

Multimedia Appendix 3

Screenshot of the FreyaWIKI homepage.

[PDF File (Adobe PDF File), 233KB - jmir_v14i5e138_app3.pdf]

References

1. Nilsen ES, Myrhaug HT, Johansen M, Oliver S, Oxman AD. Methods of consumer involvement in developing healthcare policy and research, clinical practice guidelines and patient information material. Cochrane Database Syst Rev 2006(3):CD004563. [doi: 10.1002/14651858.CD004563.pub2] [Medline: 16856050]
2. Committee on Standards for Developing Trustworthy Clinical Practice Guidelines. In: Graham RL, Mancher M, Wolman DM, Greenfield S, Steinberg E, editors. Clinical Practice Guidelines We Can Trust. Washington, DC: National Academies Press; 2011.
3. van Wersch A, Eccles M. Involvement of consumers in the development of evidence based clinical guidelines: practical experiences from the North of England evidence based guideline development programme. Qual Health Care 2001 Mar;10(1):10-16 [FREE Full text] [Medline: 11239138]
4. Crawford MJ, Rutter D, Manley C, Weaver T, Bhui K, Fulop N, et al. Systematic review of involving patients in the planning and development of health care. BMJ 2002 Nov 30;325(7375):1263. [Medline: 12458240]
5. Burgers JS, Fervers B, Haugh M, Brouwers M, Browman G, Philip T, et al. International assessment of the quality of clinical practice guidelines in oncology using the Appraisal of Guidelines and Research and Evaluation Instrument. J Clin Oncol 2004 May 15;22(10):2000-2007. [doi: 10.1200/JCO.2004.06.157] [Medline: 15143093]
6. Bastian H. Raising the standard: practice guidelines and consumer participation. Int J Qual Health Care 1996 Oct;8(5):485-490 [FREE Full text] [Medline: 9117202]
7. Gagliardi AR, Brouwers MC, Palda VA, Lemieux-Charles L, Grimshaw JM. How can we improve guideline use? A conceptual framework of implementability. Implement Sci 2011;6:26 [FREE Full text] [doi: 10.1186/1748-5908-6-26] [Medline: 21426574]
8. AGREE Enterprise. Appraisal of Guidelines for Research and Evaluation (AGREE Instrument) URL: http://www.agreetrust.org/mod_sys/errors/error.html?aspxerrorpath=/index.aspx [accessed 2012-02-16] [WebCite Cache ID 65UXoyUUk]
9. Schünemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: 10. Integrating values and consumer involvement. Health Res Policy Syst 2006;4:22 [FREE Full text] [doi: 10.1186/1478-4505-4-22] [Medline: 17147811]
10. Rogers WA. Are guidelines ethical? Some considerations for general practice. Br J Gen Pract 2002 Aug;52(481):663-668. [Medline: 12171228]

11. National Institute for Health and Clinical Excellence. 2009 Apr 03. The Guidelines Manual 2009 URL: http://www.nice.org.uk/aboutnice/howwework/developingniceclinicalguidelines/clinicalguidelinedevelopmentmethods/GuidelinesManual2009.jsp [accessed 2012-10-01] [WebCite Cache ID 6B5JzkRSR]
12. Healthcare Improvement Scotland. Scottish Intercollegiate Guidelines Network (SIGN). 2011 Nov. SIGN 50: A Guideline Developer's Handbook, Revised Edition URL: http://www.sign.ac.uk/pdf/sign50.pdf [accessed 2012-02-16] [WebCite Cache ID 65UY1tSVw]
13. Díaz Del Campo P, Gracia J, Blasco JA, Andradas E. A strategy for patient involvement in clinical practice guidelines: methodological approaches. BMJ Qual Saf 2011 Sep;20(9):779-784. [doi: 10.1136/bmjqs.2010.049031] [Medline: 21460393]
14. Légaré F, Boivin A, van der Weijden T, Pakenham C, Burgers J, Légaré J, et al. Patient and public involvement in clinical practice guidelines: a knowledge synthesis of existing programs. Med Decis Making 2011;31(6):E45-E74. [doi: 10.1177/0272989X11424401] [Medline: 21959267]
15. Boivin A, Currie K, Fervers B, Gracia J, James M, Marshall C, G-I-N PUBLIC. Patient and public involvement in clinical guidelines: international experiences and future perspectives. Qual Saf Health Care 2010 Oct;19(5):e22. [doi: 10.1136/qshc.2009.034835] [Medline: 20427302]
16. Krahn M, Naglie G. The next step in guideline development: incorporating patient preferences. JAMA 2008 Jul 23;300(4):436-438. [doi: 10.1001/jama.300.4.436] [Medline: 18647988]
17. Cross M. How the internet is changing health care. BMJ 2008;337:a883. [Medline: 18647776]
18. Giustini D. How Web 2.0 is changing medicine. BMJ 2006 Dec 23;333(7582):1283-1284. [doi: 10.1136/bmj.39062.555405.80] [Medline: 17185707]
19. Heilman JM, Kemmann E, Bonert M, Chatterjee A, Ragar B, Beards GM, et al. Wikipedia: a key tool for global public health promotion. J Med Internet Res 2011;13(1):e14 [FREE Full text] [doi: 10.2196/jmir.1589] [Medline: 21282098]
20. Trevena L. WikiProject medicine. BMJ 2011;342:d3387. [Medline: 21653617]
21. Haagen EC, Tuil W, Hendriks J, de Bruijn RP, Braat DD, Kremer JA. Current Internet use and preferences of IVF and ICSI patients. Hum Reprod 2003 Oct;18(10):2073-2078 [FREE Full text] [Medline: 14507823]
22. Weissman A, Gotlieb L, Ward S, Greenblatt E, Casper RF. Use of the internet by infertile couples. Fertil Steril 2000 Jun;73(6):1179-1182. [Medline: 10856479]
23. Wischmann T. Implications of psychosocial support in infertility--a critical appraisal. J Psychosom Obstet Gynaecol 2008 Jun;29(2):83-90. [doi: 10.1080/01674820701817870] [Medline: 18484439]
24. Gnoth C, Godehardt E, Frank-Herrmann P, Friol K, Tigges J, Freundl G. Definition and prevalence of subfertility and infertility. Hum Reprod 2005 May;20(5):1144-1147 [FREE Full text] [doi: 10.1093/humrep/deh870] [Medline: 15802321]
25. Taylor A. ABC of subfertility: extent of the problem. BMJ 2003 Aug 23;327(7412):434-436. [doi: 10.1136/bmj.327.7412.434] [Medline: 12933733]
26. Nachtigall RD. International disparities in access to infertility services. Fertil Steril 2006 Apr;85(4):871-875. [doi: 10.1016/j.fertnstert.2005.08.066] [Medline: 16580367]
27. UMC St Radboud. 2008. FreyaWIKI URL: https://www.freyawiki.nl/ [accessed 2012-02-16] [WebCite Cache ID 65UYXww7d]
28. Zegers-Hochschild F, Adamson GD, de Mouzon J, Ishihara O, Mansour R, Nygren K, International Committee for Monitoring Assisted Reproductive Technology, World Health Organization. International Committee for Monitoring Assisted Reproductive Technology (ICMART) and the World Health Organization (WHO) revised glossary of ART terminology, 2009. Fertil Steril 2009 Nov;92(5):1520-1524. [doi: 10.1016/j.fertnstert.2009.09.009] [Medline: 19828144]
29. Dancet EA, Nelen WL, Sermeus W, De Leeuw L, Kremer JA, D'Hooghe TM. The patients' perspective on fertility care: a systematic review. Hum Reprod Update 2010;16(5):467-487 [FREE Full text] [doi: 10.1093/humupd/dmq004] [Medline: 20223789]
30. Shiffman RN, Dixon J, Brandt C, Essaihi A, Hsiao A, Michel G, et al. The GuideLine Implementability Appraisal (GLIA): development of an instrument to identify obstacles to guideline implementation. BMC Med Inform Decis Mak 2005;5:23 [FREE Full text] [doi: 10.1186/1472-6947-5-23] [Medline: 16048653]
31. Arain M, Campbell MJ, Cooper CL, Lancaster GA. What is a pilot or feasibility study? A review of current practice and editorial policy. BMC Med Res Methodol 2010;10:67 [FREE Full text] [doi: 10.1186/1471-2288-10-67] [Medline: 20637084]
32. Yusof MM, Kuljis J, Papazafeiropoulou A, Stergioulas LK. An evaluation framework for Health Information Systems: human, organization and technology-fit factors (HOT-fit). Int J Med Inform 2008 Jun;77(6):386-398. [doi: 10.1016/j.ijmedinf.2007.08.011] [Medline: 17964851]
33. Francis JJ, Johnston M, Robertson C, Glidewell L, Entwistle V, Eccles MP, et al. What is an adequate sample size? Operationalising data saturation for theory-based interview studies. Psychol Health 2010 Dec;25(10):1229-1245. [doi: 10.1080/08870440903194015] [Medline: 20204937]
34. Silverman D. Doing Qualitative Research. 3rd edition. London: Sage Publications Ltd; 2009.
35. Phillips B, Ball C, Sackett D, Badenoch D, Straus S, Haynes B, et al. Oxford Centre for Evidence Based Medicine. 2009 Mar. Levels of Evidence URL: http://www.cebm.net/index.aspx?o=4590 [accessed 2012-06-21] [WebCite Cache ID 68aRCTclX]

36. Jalali A, Mioduszewski M, Gauthier M, Varpio L. Wiki use and challenges in undergraduate medical education. Med Educ 2009 Nov;43(11):1117. [doi: 10.1111/j.1365-2923.2009.03480.x] [Medline: 19874526]
37. Naik AD, Singh H. Electronic health records to coordinate decision making for complex patients: what can we learn from wiki? Med Decis Making 2010;30(6):722-731. [doi: 10.1177/0272989X10385846] [Medline: 21183759]
38. Archambault PM, Légaré F, Lavoie A, Gagnon MP, Lapointe J, St-Jacques S, et al. Healthcare professionals' intentions to use wiki-based reminders to promote best practices in trauma care: a survey protocol. Implement Sci 2010;5:45 [FREE Full text] [doi: 10.1186/1748-5908-5-45] [Medline: 20540775]
39. Gupta S, Wan FT, Newton D, Bhattacharyya OK, Chignell MH, Straus SE. WikiBuild: a new online collaboration process for multistakeholder tool development and consensus building. J Med Internet Res 2011;13(4):e108 [FREE Full text] [doi: 10.2196/jmir.1833] [Medline: 22155694]
40. Archambault PM. WikiBuild: a new application to support patient and health care professional involvement in the development of patient support tools. J Med Internet Res 2011;13(4):e114 [FREE Full text] [doi: 10.2196/jmir.1961] [Medline: 22155746]
41. Hansen DL. Knowledge Sharing, Maintenance, and Use in Online Support Communities [doctoral dissertation]. Ann Arbor, MI: University of Michigan; 2007. URL: http://deepblue.lib.umich.edu/bitstream/2027.42/57608/2/shakmatt_1.pdf [accessed 2012-06-21] [WebCite Cache ID 68aQHFxWs]
42. Pope C, van Royen P, Baker R. Qualitative methods in research on healthcare quality. Qual Saf Health Care 2002 Jun;11(2):148-152 [FREE Full text] [Medline: 12448807]
43. Kummervold PE, Chronaki CE, Lausen B, Prokosch HU, Rasmussen J, Santana S, et al. eHealth trends in Europe 2005-2007: a population-based survey. J Med Internet Res 2008;10(4):e42 [FREE Full text] [doi: 10.2196/jmir.1023] [Medline: 19017584]
44. Spyridakis JH. Guidelines for authoring comprehensible Web pages and evaluating their success. Tech Commun 2000 Aug 01;47(3):359-382 [FREE Full text]
45. Gustafson DH, Wyatt JC. Evaluation of ehealth systems and services. BMJ 2004 May 15;328(7449):1150. [doi: 10.1136/bmj.328.7449.1150] [Medline: 15142895]
46. Kushniruk A. Evaluation in the design of health information systems: application of approaches emerging from usability engineering. Comput Biol Med 2002 May;32(3):141-149. [Medline: 11922931]
47. Taylor HA, Sullivan D, Mullen C, Johnson CM. Implementation of a user-centered framework in the development of a web-based health information database and call center. J Biomed Inform 2011 Oct;44(5):897-908. [doi: 10.1016/j.jbi.2011.03.001] [Medline: 21396486]
48. Roehrer E, Cummings E, Ellis L, Turner P. The role of user-centred design within online community development. Stud Health Technol Inform 2011;164:256-260. [Medline: 21335720]
49. Archambault PM, van de Belt TH, Grajales FJ, Eysenbach G, Aubin K, Gold I, et al. Wikis and collaborative writing applications in health care: a scoping review protocol. JMIR Res Protoc 2012;1(1):e1. [doi: 10.2196/resprot.1993]
50. Brandes M, Hamilton CJ, Bergevoet KA, de Bruin JP, Nelen WL, Kremer JA. Origin of multiple pregnancies in a subfertile population. Acta Obstet Gynecol Scand 2010 Sep;89(9):1149-1154. [doi: 10.3109/00016349.2010.498495] [Medline: 20804340]
51. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]
52. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the Internet and e-mail for health care information: results from a national survey. JAMA 2003 May 14;289(18):2400-2406. [doi: 10.1001/jama.289.18.2400] [Medline: 12746364]
53. Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. Soc Sci Med 2004 Nov;59(9):1795-1806. [doi: 10.1016/j.socscimed.2004.02.020] [Medline: 15312915]
54. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]
55. Karunanithi M. Monitoring technology for the elderly patient. Expert Rev Med Devices 2007 Mar;4(2):267-277. [doi: 10.1586/17434440.4.2.267] [Medline: 17359231]
56. Gagnon MP, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012 Feb;36(1):241-277. [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
57. Chhanabhai P, Holt A. Consumers are ready to accept the transition to online and electronic records if they can be assured of the security measures. MedGenMed 2007;9(1):8 [FREE Full text] [Medline: 17435617]
58. Gagnon MP, Shaw N, Sicotte C, Mathieu L, Leduc Y, Duplantie J, et al. Users' perspectives of barriers and facilitators to implementing EHR in Canada: a study protocol. Implement Sci 2009;4:20 [FREE Full text] [doi: 10.1186/1748-5908-4-20] [Medline: 19358739]

Abbreviations

AGREE: Appraisal of Guidelines Research and Evaluation
CPG: clinical practice guideline
GLIA: Guideline Implementability Appraisal

Edited by G Eysenbach; submitted 16.02.12; peer-reviewed by N Nijland, P Archambault; comments to author 14.04.12; revised version received 22.06.12; accepted 26.07.12; published 26.10.12.

Please cite as:
den Breejen EME, Nelen WLDM, Knijnenburg JM, Burgers JS, Hermens RPMG, Kremer JAM
Feasibility of a Wiki as a Participatory Tool for Patients in Clinical Guideline Development
J Med Internet Res 2012;14(5):e138
URL: http://www.jmir.org/2012/5/e138/
doi:10.2196/jmir.2080
PMID:23103790

Original Paper

Conducting Research on the Internet: Medical Record Data Integration with Patient-Reported Outcomes

Elisa Cascade[1*], MBA; Paige Marr[2*], MSc; Matthew Winslow[1*]; Andrew Burgess[3*]; Mark Nixon[3*], PhD

[1]Digital Patient Unit, Quintiles, Durham, NC, United States
[2]Outcomes Health Information Solutions, Alpharetta, GA, United States
[3]Quintiles, Reading, United Kingdom
[*]all authors contributed equally

Corresponding Author:
Elisa Cascade, MBA
Digital Patient Unit
Quintiles
4820 Emperor Blvd
7th Floor #773C
Durham, NC, 27707
United States
Phone: 1 703 597 6642
Fax: 1 919 518 9081
Email: elisa.cascade@quintiles.com

Abstract

Background: The growth in the number of patients seeking health information online has given rise to new direct-to-patient research methods, including direct patient recruitment and study conduct without use of physician sites. While such patient-centric designs offer time and cost efficiencies, the absence of physician-reported data is a key concern, with potential impact on both scientific rigor and operational feasibility.

Objective: To (1) gain insight into the viability of collecting patient-reported outcomes and medical record information in a sample of gout patients through a direct-to-patient approach (ie, without the involvement of physician sites), and (2) evaluate the validity of patient-reported diagnoses collected during a patient-reported outcomes plus medical record (PRO+MR) direct-to-patient study.

Methods: We invited a random sample of MediGuard.org members aged 18 to 80 years to participate via email based on a gout treatment or diagnosis in their online profiles. Interested members clicked on an email link to access study information, consent to participate electronically, and be screened for eligibility. The first 50 consenting participants completed an online survey and provided electronic and wet signatures on medical record release forms for us to obtain medical charts from their managing physicians.

Results: A total of 108 of 1250 MediGuard.org members (8.64%) accessed study information before we closed the study at 50 completed surveys. Of these 108 members who took the screener, 50 (46.3%) completed the study, 19 (17.6%) did not pass the screening, 5 (4.6%) explicitly declined to participate due to the medical record requirement, and 34 (31.5%) closed the browser without completing the survey screener. Ultimately, we obtained 38 of 50 charts (76%): 28 collected using electronic signature and 10 collected based on wet signature on a paper form. Of the 38 charts, 37 cited a gout diagnosis (35 charts) or use of a gout medication (2 charts). Only 1 chart lacked any mention of gout.

Conclusions: Patients can be recruited directly for observational study designs that include patient-reported outcomes and medical record data with over 75% data completeness. Although the validity of self-reported diagnosis is often a concern in Internet-based studies, in this PRO+MR study pilot, nearly all (37 of 38) charts confirmed patient-reported data.

(J Med Internet Res 2012;14(5):e137) doi:10.2196/jmir.2202

KEYWORDS

Direct-to-patient study; patient-reported outcomes; observational research; medical record review; Internet recruitment; online patient communities

Introduction

The emergence of the e-patient—an individual seeking health information online—offers important opportunities to advance research. The ability to access targeted patient populations, coupled with technology to capture patient-reported outcomes data via the Internet, can reduce study timelines and cost, thus increasing the operational feasibility of real-world drug evaluation. These benefits are becoming increasingly important for observational research in light of growing demands for postapproval noninterventional studies, registries, and Risk Evaluation and Mitigation Strategy programs to evaluate real-world treatment effects.

In these direct-to-patient observational research designs, participant recruitment and data collection are conducted directly with patients rather than through clinic-based physician investigators. Although direct-to-patient studies initially focused on collection of patient-reported outcomes data only [1,2], these research designs have been evolving to incorporate clinical data from medical records as well as collection of genomic samples and laboratory data. In this paper, we report results from what we believe to be the first direct-to-patient research study involving collection of patient-reported outcomes data and clinical information extracted from patient medical records.

Potential Limitations of Direct-to-Patient Studies

One of the greatest concerns associated with direct-to-patient observational studies is the quality of the information provided by patients outside of the investigator’s office. To investigate this issue, we conducted a pilot study in US gout patients using a design combining patient-reported outcomes and medical record data to evaluate the feasibility of recruiting a representative population of patients via the Internet, the willingness of physicians to provide medical record data, and the validity of self-reported diagnosis. We believe that this pilot study is the first to evaluate the ability to conduct observational direct-to-patient studies using a patient-reported outcomes plus medical record (PRO+MR) approach. As a result, we had no a priori hypothesis related to the proportion of charts that we could obtain or the validity of self-reported diagnosis in Internet research.

Methods

Study Design

This was a real-world, observational pilot study that combined patient-reported outcomes and medical records from gout patients recruited from an Internet database.

Recruitment

We recruited participants from MediGuard.org, a free online service that monitors the safety of prescription medicines, over-the-counter medicines, and health care supplements for more than 2.5 million patients in the United States, United Kingdom, France, Germany, Spain, and Australia. MediGuard.org attracts members through online search engines and social media as well as outreach efforts with physicians, pharmacies, and health-related websites. Patients who enroll in the MediGuard.org service consent to be contacted about research opportunities as part of the registration process, and they double consent to participate in any individual study following receipt of an email invitation.

In July 2011, we invited US MediGuard.org members, aged 18 to 80 years, to participate in the study via email. Members whose MediGuard.org profiles cited a diagnosis of gout or use of a gout treatment (allopurinol or colchicine) were randomly selected for invitation to the study. Interested members clicked on an email link to access study information, provided their electronic signature for consent, and screened for eligibility based on a self-reported diagnosis of gout and willingness to release their medical records as part of the study design. Participants who completed the online survey and returned the paper medical record release form within 2 weeks were offered a US $50 MasterCard gift card as compensation for their time, to be mailed to the participant.

Data Collection: Online Survey

The first 50 consenting participants were directed to complete an online survey of 20 questions that asked for information regarding disease and treatment experience, including time since first diagnosis of gout, specialty of treating physician, number of gout attacks in the previous year, number of physician visits for gout in the past year, satisfaction with treatment, and adherence to treatment regimens. Patients also reported demographic data such as age, gender, employment status, insurance status, household income, and highest level of education obtained. Prior to launching the online survey, we performed a quality control test to validate the functionality of the survey and data reporting tool.

Data Collection: Medical Chart Authorization

After completing the online survey, participants also completed an electronic release form consenting to a medical record review. The purpose of accessing the patient’s medical record was to confirm the diagnosis of gout reported by the patient.

The medical chart release form asked participants to provide their physician’s name, address, specialty, and telephone and fax numbers, as well as the participant’s signature, name, telephone number, and date of birth. The participant’s personal information was electronically transferred to Outcomes Health Information Solutions (Alpharetta, GA, USA) in a secured manner to protect the participant’s personal information. After completing the electronic consent, participants were sent a paper copy of the medical record release form by regular mail. Participants were provided with a postage-paid, return-addressed envelope and requested to return the completed form within 1 week of receipt. Completed forms were mailed to Outcomes Health Information Solutions, a third-party contractor that specializes in health care abstraction and quality compliance programs.

Medical Record Extraction

On receipt of the medical record release forms, Outcomes Health Information Solutions contacted patients’ managing physicians to schedule an appointment for chart extraction. Our approach to obtaining access to the chart was to first offer the physician the electronic medical record release. If the physician requested it, we subsequently provided a copy of the signed paper medical record release form when one was available.

According to the information provided in the medical record release form, Outcomes Health Information Solutions extracted data for the 2-year period from June 2009 to June 2011. Using jointly approved abstraction guidelines, nurses with clinical coding certifications reviewed the collected charts for evidence of gout. The abstraction guidelines defined acceptable physician notes and *International Classification of Diseases*, 9th revision (ICD-9) codes that could be used to support a diagnosis and excluded differential diagnoses as evidence of disease. If a diagnosis was not present in the 2-year chart data, the nurses examined the chart for mention of a prescription for colchicine or allopurinol, which are medications commonly used to treat gout. The date on which the paper medical record release form was received, the type of release form (paper or electronic) used for chart retrieval, the date when the chart was received, and the presence of a gout diagnosis or medication was entered into an Excel database by Outcomes Health Information Solutions, along with the de-identified patient study identification number.

Validation of Patient-Reported Diagnosis

Diagnosis data obtained from the nurses were coded as confirmed, suspected, or missing based on the following formula: the patient's self-reported diagnosis was considered confirmed if the medical chart contained a diagnosis; the patient's self-reported diagnosis was considered suspected if the chart contained one of the medications commonly used for the treatment of gout; and charts containing neither a diagnosis nor medication were coded as missing.

Statistical Analysis

Medical chart data elements were merged with information from the patient questionnaire based on the de-identified study identification number. We conducted all analyses on the aggregated data set.

Metrics related to the demographics and gout experience of the sample were analyzed. To evaluate the representativeness of the MediGuard.org population completing the study, we compared age and gender demographics with data obtained from Encuity Research's Physician Drug and Diagnosis Audit (PDDA). According to Encuity's website, "PDDA surveys more than 3,100 office-based physicians representing 29 specialties across the US. Physicians report on all patient activity during one typical workday each month. Once collected, the drug and diagnosis information is projected by region and specialty to estimate national activity for a universe of more than 460,000 physicians" [3]. We compared age and gender from our sample against Encuity's PDDA data using Pearson chi-square tests. We were not able to access a data source to examine the representativeness of other gout-related study variables.

To describe the feasibility of chart extraction and diagnosis validation, we calculated two metrics related to the medical record process: the proportion of medical record release forms returned and charts received. In addition, we examined the type of release required to obtain the chart (electronic vs wet paper signature) and the status of the chart diagnosis (confirmed, suspected, or missing).

With respect to factors related to chart access, we compared the proportion of charts collected based on physician specialty, time since diagnosis, number of visits in the past year (3 items), and number of visits in the past year (binary). Binary variables (specialty and visits in the past year) were compared using the Fisher exact test. Time since diagnosis and the nonbinary physician visit variable were compared using the Cochran-Mantel-Haenszel statistic.

Results

Enrollment

MediGuard.org contains more than 8250 US members whose profile contains a diagnosis of gout or a medication likely to be used in the treatment of gout (eg, allopurinol or colchicine), or both. A total of 1250 emails were sent to MediGuard.org members, prompting 108 members (8.64%) to click on the email link to access more information about the study. Of the members screening for the study, 5 (4.6%) explicitly declined to participate due to the medical record requirement, and 34 (31.5%) closed the browser without completing the survey screener. If all individuals who closed the browser were to be included, the number of members who declined could be as high as 39 (36.1%).

Sample Characteristics

A total of 50 members completed the study. It should be noted that the completion rate may have been higher, but the survey closed after the first 50 members had consented to participate. Table 1 displays demographic characteristics self-reported by the 50 participants completing the study. As seen in Table 1, participants represented a broad range of age, gender, employment status, insurance access, income, and education characteristics.

Table 1. Participants' demographic characteristics (n = 50).

Characteristic	n	%
Age (years)		
<40	3	6%
40–59	19	38%
60+	28	56%
Gender		
Male	34	68%
Female	16	32%
Employment status		
Employed	15	30%
Not employed	35	70%
Insurance status		
Insured—employer	17	34%
Insured—public (Medicare, Medicaid)	27	54%
Insured—Military/Department of Defense	3	6%
Uninsured/self-pay	3	6%
Household annual income (US $)		
<$25,000	14	28%
$25,000–$34,999	7	14%
$35,000–$49,999	7	14%
$50,000–$74,999	9	18%
$75,000–$99,999	1	2%
$100,000+	9	18%
Declined to answer	3	6%
Highest level of education		
High school graduate	11	22%
Some college	18	36%
College graduate	13	26%
Postgraduate studies	5	10%
Declined to answer	3	6%

Comparison with demographic data from Encuity's PDDA suggests that our sample had slightly more women than the general population of patients with gout (68% male in our sample vs 82% male in Encuity's PDDA). We performed a Pearson chi-square test compare the gender in the sample against the Encuity's PDDA data, and the difference was significant at the 5% level (*P* = .01). With respect to age, the sample was extremely similar to Encuity's PDDA data: <40 years (6% in our sample vs 9% in Encuity's PDDA), 40–59 years (38% sample vs 45% in Encuity's PDDA), and 60+ years (56% sample vs 46% in Encuity's PDDA). We performed a Pearson chi-square test comparing the sample against Encuity's PDDA data with age pooled into two categories (<59, 60+), the difference was not significant at the 5% level (*P* = .17).

Table 2 summarizes self-reported characteristics of the participants' gout condition. The majority of study participants (80%) were managed by a primary care physician. Two-thirds of patients in the gout sample had experienced a flare in the past year; however, only 50% of patients had visited their physician in the past year due to gout.

Table 2. Participants' gout experience (n = 50).

Characteristic	n	%
Time since first diagnosis of gout (years ago)		
<5	18	36%
5–10	11	22%
>10	21	42%
Number of gout attacks in past year		
0	17	34%
1	9	18%
2	5	10%
3	6	12%
4+	9	18%
I do not know/not sure	4	8%
Number of physician visits for gout in past year		
0	25	50%
1	10	20%
2	4	8%
3	2	4%
4+	5	10%
I do not know/not sure	4	8%
Specialty of physician managing gout		
Primary care physician	40	80%
Rheumatologist	7	14%
Other	3	6%

Medical Chart Data: Feasibility Evaluation

Table 3 displays metrics related to the medical record extraction process. As Table 3 shows, 42 of 50 participants (84%) who completed the online survey and electronic medical record release form also completed and returned the paper form. In total, we obtained 38 of 50 charts (76%): 28 (74%) were provided in response to the electronic medical record release; and 10 (26%) were provided in response to receipt of a wet signature on a paper form. The remaining 12 charts were not retrieved due to physician refusal (2 charts), the participant's wet signature form was requested by the physician but not returned by the patient (2 charts), or the physician's office did not complete the request by the time the study closed (8 charts).

In Table 4, we provide details on the number of charts accessed by characteristics such as physician specialty, time since diagnosis, and number of physician visits in the past year. While there were no significant differences based on these characteristics, a larger sample size might have yielded differences, particularly for the variable number of physician visits within the past year.

Table 3. Medical record process metrics.

Outcome	n	%
Paper medical record release form returned with wet signature (n = 50)		
Yes	42	84%
No	8	16%
Chart received (n = 50)		
Yes	38	76%
No	12	24%
Type of release required (n = 38)		
Electronic	28	74%
Paper	10	26%

Table 4. Factors related to number of charts access.

Type of access (n = 38)	n	%	*P* value
By physician specialty[a]			1.00
Primary care (n = 40)	30	75%	
Rheumatologist (n = 7)	6	86%	
By years since diagnosis[b]			.31
<5 (n = 18)	13	72%	
5–10 (n = 11)	7	64%	
>10 (n = 21)	18	86%	
By number of visits in past year [b]			.67
0 (n = 25)	17	68%	
1 (n = 10)	9	90%	
2+ (n = 6)	4	67%	
By number of visits in past year (binary) [a]			.48
0 (n = 25)	17	68%	
1+ (n = 16)	13	81%	

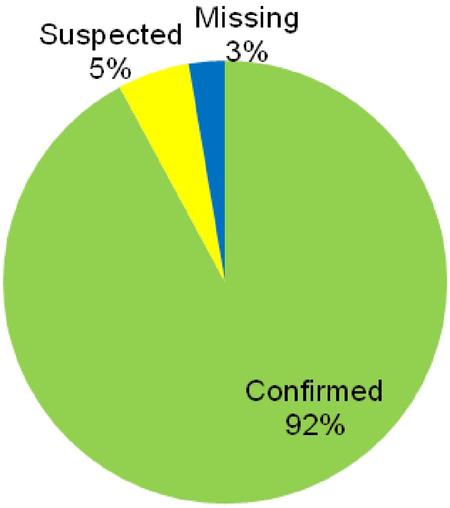
[a] *P* values calculated by using Fisher exact test.

[b] *P* values calculated using Cochran-Mantel-Haenszel statistic.

Medical Chart Data: Diagnosis Validation

As Figure 1 shows, 35 of the 38 charts obtained (92%) contained a physician's diagnosis of gout to confirm the patient's self-reported diagnosis. An additional 2 charts (5%) contained notes regarding treatment with the gout medications allopurinol or colchicine; in these 2 cases, we considered gout to be likely and thus coded them as suspected. Only 1 chart (3%) was missing any diagnosis or notation of prescribed medication associated with gout.

Figure 1. Medical record validation of self-reported diagnosis.



Discussion

Results of this pilot study are a positive step forward in confirming the viability of the PRO+MR direct-to-patient study approach and the validity of patient-reported outcomes when collected in this manner—a foundational step toward broader use of this design in collecting real-world, observational data. In this study, the findings demonstrated that patients could be recruited, screened, and enrolled directly from online patient communities for observational studies that collect patient-reported outcomes and medical record data, with more than 75% data completeness. Although there is concern that patients cannot be relied on to accurately report data in the absence of a physician investigator, in this PRO+MR pilot study, nearly all medical charts (37 of 38) confirmed the accuracy of a patient-reported gout diagnosis. Further, regarding the 1 chart that did not confirm diagnosis, it is unclear whether the respondent falsely claimed a diagnosis of gout or whether the chart was simply missing a note regarding patient diagnosis.

Study Limitations

Although these findings have important implications for advancing direct-to-patient study design, the study has some limitations. First, we conducted our pilot PRO+MR study among a small sample of members with a single diagnosis (gout). Other studies are now being conducted in other conditions (eg, chronic obstructive pulmonary disease and rheumatoid arthritis) and with much larger sample sizes to determine whether the findings described in this paper are reproducible across patient types and physician practice specialties and settings.

Another limitation is that the medical record review in this study was restricted to a 2-year look-back in a single provider's office for two primary fields: diagnosis and medications. Requesting more chart data or requiring a longer look-back period could have a negative impact on provider compliance. Additionally, while the current study extracted information from a single provider's office, future study designs could require data that are dispersed across multiple care settings (eg, primary care and hospital records), also having a potential negative impact on data completeness. Developing a better understanding of the number of fields, type of information, and length of review period that is feasible in a medical record review will be a key objective of future projects.

It is also possible that unique characteristics of the MediGuard.org population influenced members' interest and ability to participate in online clinical studies. For example, enrollment in MediGuard.org may self-select individuals with special interest in medical information that distinguishes them from the general population. Another distinguishing characteristic may be members' facility in using the Internet. A higher level of Internet skills among members would also limit the generalization of findings to the broader population. However, according to a 2011 Pew report, 74% of all US adults use the Internet and 80% of these Internet users actively seek health information online, including information about a specific disease or treatment [4]. As the population ages and Internet-enabling technology continues to evolve with devices including smartphones and tablet devices, remaining concerns should diminish regarding the generalization of findings from direct-to-patient studies recruited from online patient communities.

Finally, there is the possibility that the methodology of combining patient-reported outcomes and medical record data resulted in a higher validation of self-reported diagnosis than would be achievable with patient-reported outcomes data alone. In other words, requiring a patient to consent to medical record extraction could have biased the sample toward patients who actually had the condition.

Potential Benefits of Direct-to-Patient Studies

In addition to eliminating site-based costs, direct-to-patient studies offer major time and cost efficiencies. In traditional trials, estimates of recruitment costs range from US $10 to $1300 per completed participant [5,6], and recruitment delays can account for up to 45% of study delays [7]. The rise of online patient communities—such as PatientsLikeMe.com, CysticFibrosis.com, Inspire.com, and MediGuard.org—offers great recruitment advantages. For example, a recent study found that recruitment through MediGuard.org returned the highest percentage of candidates interested in a rheumatoid arthritis study at the lowest cost (US $4.82 per interested patient), compared with direct mail (US $86.28) and email generated by a third-party email list (US $195.65) [8].

Direct-to-patient studies can also overcome traditional recruitment barriers, such as patients' lack of study awareness and limitations due to travel or site location. Patients learn about study opportunities through online communities and by searching social network sites; rather than traveling to an investigator's office, patients participate in their own homes. Enhanced by greater awareness and wide geographical capture of study participants, and enabled by the high penetration of Internet access such as in the United States and Western Europe, direct-to-patient studies also offer a broader range of participant demographics and physician specialties and practice settings. This makes study results potentially more representative of the overall community of patients with a condition than do physician-centric studies, which are limited to the population of a few sites. Study retention is also likely to be higher in direct-to-patient studies because of the explicit alignment of patient incentives: the patient learns about the study directly, understands what is required to be compensated for participating in the study, self-consents to participate, and then self-reports study information.

Finally, direct-to-patient studies allow researchers in the United States to meet the Health Insurance Portability and Accountability Act requirements for maintaining patient information because the researchers obtain patients' authorization to capture and store personal health information. Specifically, in direct-to-patient studies, researchers can engage patients directly via the Internet (no physicians involved); patients can choose to self-consent to participate in the study; patients voluntarily share identifiers to access their medical record; and patients provide an electronic signature and a wet signature on a paper form to have their medical record accessed. To maximize protection of personal health information, we use identifiers only to link survey and medical record data; once the linkage was completed, the study database was maintained in a de-identified format to minimize future risk of breach.

Implications for Future Research

Confirmation of the validity of patient-reported data in observational, direct-to-patient remote studies is an essential step forward toward scaling up this method in real-world research. As direct-to-patient observational studies grow in number and size, experience and insights from these designs can be considered for integration into interventional clinical trials. While efforts to implement direct-to-patient approaches in interventional clinical trials, most notably the REMOTE study initiated by Pfizer in 2011 [9], are as yet inconclusive, knowledge gained through observational designs such as the PRO+MR model described here enables researchers to better identify study challenges and implement potential corrective solutions. Whether direct-to-patient studies are integrated in totality for interventional studies, or individual processes are adopted for recruitment, reporting, retention, and long-term extensions, one trend remains clear: the pool of e-patients will continue to grow globally, and harnessing the power of these patients offers the potential to drive a paradigm shift in clinical research.

Acknowledgments

This study was funded internally by Quintiles and Outcomes Health Information Solutions. No sponsor was involved.

Conflicts of Interest

Individuals involved in this paper are employees of Quintiles and Outcomes Health Information Solutions, both of which are named in this paper.

References

1. Wade AG, Crawford GM, Pumford N, Koscielny V, Maycock S, McConnachie A. Baseline characteristics and patient reported outcome data of patients prescribed etanercept: web-based and telephone evaluation. BMC Med Res Methodol 2011;11:91 [FREE Full text] [doi: 10.1186/1471-2288-11-91] [Medline: 21672203]
2. Frost J, Okun S, Vaughan T, Heywood J, Wicks P. Patient-reported outcomes as a source of evidence in off-label prescribing: analysis of data from PatientsLikeMe. J Med Internet Res 2011;13(1):e6 [FREE Full text] [doi: 10.2196/jmir.1643] [Medline: 21252034]
3. Encuity Research, LLC. 2012. Specialties Surveyed URL: http://www.encuity.com/services/syndicated/pdda/specialties/ [accessed 2012-09-30] [WebCite Cache ID 6B3xrvrpS]
4. Fox S. The social life of health information, 2011. Washington, DC: Pew Internet & American Life Project; 2011 May 12. URL: http://www.pewinternet.org/~/media//Files/Reports/2011/PIP_Social_Life_of_Health_Info.pdf [accessed 2012-09-26] [WebCite Cache ID 6AyFrAV8X]
5. Gren L, Broski K, Childs J, Cordes J, Engelhard D, Gahagan B, et al. Recruitment methods employed in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. Clin Trials 2009 Feb;6(1):52-59. [doi: 10.1177/1740774508100974] [Medline: 19254935]
6. Roy S, Patel S, Sheehan KH, Janavs J, Sheehan D. Efficacy of print advertising for a bipolar disorder study. Psychopharmacol Bull 2008;41(1):136-141. [Medline: 18362876]
7. Anderson DL. A Guide to Patient Recruitment: Today's Best Practices & Proven Strategies. Boston, MA: CenterWatch; 2001.
8. Stephenson H, Bhamal M. Applied Clinical Trials Online. 2010 Sep 1. Direct-to-Patient Enrollment Strategies: A Comparison of the Yields and Costs of Online Outreach Methods to Other Recruitment Techniques URL: http://www.appliedclinicaltrialsonline.com/appliedclinicaltrials/CRO%2FSponsor/Direct-to-Patient-Enrollment-Strategies/ArticleStandard/Article/detail/686202 [accessed 2012-05-09] [WebCite Cache ID 67X80Y8PM]
9. Taylor N. OutsourcingPharma.com. 2012 Mar 6. Pfizer Director Defends Virtual Trial After Recruitment Struggle URL: http://www.outsourcing-pharma.com/Clinical-Development/Pfizer-director-defends-virtual-trial-after-recruitment-struggle [accessed 2012-05-10] [WebCite Cache ID 67Yj6QtHP]

Abbreviations

ICD-9: International Classification of Diseases, 9th revision
PDDA: Physician Drug and Diagnosis Audit
PRO+MR: patient-reported outcomes plus medical record

Edited by G Eysenbach; submitted 04.06.12; peer-reviewed by J Banks, P Wicks; comments to author 20.06.12; revised version received 19.07.12; accepted 18.09.12; published 11.10.12.

Please cite as:
Cascade E, Marr P, Winslow M, Burgess A, Nixon M
Conducting Research on the Internet: Medical Record Data Integration with Patient-Reported Outcomes
J Med Internet Res 2012;14(5):e137
URL: http://www.jmir.org/2012/5/e137/
doi:10.2196/jmir.2202
PMID:23060427

Original Paper

Validation of Web-Based Physical Activity Measurement Systems Using Doubly Labeled Water

Hideyuki Namba[1], MEng, PhD; Yukio Yamaguchi[1], MEd; Yosuke Yamada[2,3], PhD; Satoru Tokushima[1], BSc; Yoichi Hatamoto[1], MSc; Hiroyuki Sagayama[1], MSc; Misaka Kimura[2], PhD; Yasuki Higaki[1], PhD; Hiroaki Tanaka[1], PhD

[1]Faculty of Sports and Health Science, Fukuoka University, Fukuoka, Japan

[2]Kyoto Prefectural University of Medicine, Kyoto, Japan

[3]Japan Society for the Promotion of Science, Tokyo, Japan

Corresponding Author:
Hideyuki Namba, MEng, PhD
Faculty of Sports and Health Science
Fukuoka University
8-19-1 Nanakuma, Jonan-ku
Fukuoka, 814-0180
Japan
Phone: 81 92 871 6631 ext 6759
Fax: 81 92 865 6029
Email: hnanba@fukuoka-u.ac.jp

Abstract

Background: Online or Web-based measurement systems have been proposed as convenient methods for collecting physical activity data. We developed two Web-based physical activity systems—the 24-hour Physical Activity Record Web (24hPAR WEB) and 7 days Recall Web (7daysRecall WEB).

Objective: To examine the validity of two Web-based physical activity measurement systems using the doubly labeled water (DLW) method.

Methods: We assessed the validity of the 24hPAR WEB and 7daysRecall WEB in 20 individuals, aged 25 to 61 years. The order of email distribution and subsequent completion of the two Web-based measurements systems was randomized. Each measurement tool was used for a week. The participants' activity energy expenditure (AEE) and total energy expenditure (TEE) were assessed over each week using the DLW method and compared with the respective energy expenditures estimated using the Web-based systems.

Results: The mean AEE was 3.90 (SD 1.43) MJ estimated using the 24hPAR WEB and 3.67 (SD 1.48) MJ measured by the DLW method. The Pearson correlation for AEE between the two methods was $r = .679$ ($P < .001$). The Bland-Altman 95% limits of agreement ranged from –2.10 to 2.57 MJ between the two methods. The Pearson correlation for TEE between the two methods was $r = .874$ ($P < .001$). The mean AEE was 4.29 (SD 1.94) MJ using the 7daysRecall WEB and 3.80 (SD 1.36) MJ by the DLW method. The Pearson correlation for AEE between the two methods was $r = .144$ ($P = .54$). The Bland-Altman 95% limits of agreement ranged from –3.83 to 4.81 MJ between the two methods. The Pearson correlation for TEE between the two methods was $r = .590$ ($P = .006$). The average input times using terminal devices were 8 minutes and 10 seconds for the 24hPAR WEB and 6 minutes and 38 seconds for the 7daysRecall WEB.

Conclusions: Both Web-based systems were found to be effective methods for collecting physical activity data and are appropriate for use in epidemiological studies. Because the measurement accuracy of the 24hPAR WEB was moderate to high, it could be suitable for evaluating the effect of interventions on individuals as well as for examining physical activity behavior.

(J Med Internet Res 2012;14(5):e123) doi:10.2196/jmir.2253

KEYWORDS

Physical activity; energy expenditure; doubly labeled water; Japan

Introduction

Despite the known hazards of physical inactivity, it continues to be a major risk in the development of chronic diseases [1]. Physical inactivity is associated with increased incidence rates of obesity, diabetes, cardiovascular diseases, osteoporosis, and cancer [2-4]. Therefore, it is important that instruments be developed that allow accurate measurement of the level of physical activity in the population.

Physical activity questionnaires are the most frequently used instruments in epidemiological studies for estimating physical activity or energy expenditure [5]. However, the validity of traditional self-reported questionnaires is low when compared with the ideal doubly labeled water (DLW) method [6]. In addition, the tasks associated with conventional questionnaires, including collection and analysis, require time and effort. The validity of the triaxial accelerometer is higher than that of questionnaires when compared with the DLW method [7]. However, it is difficult to measure cycling, swimming, and activities involving only the upper limb using an accelerometer [5]. Accelerometers are easily available and include memory for long-term data collection. However, because they are expensive, they cannot be used for epidemiological studies that require physical activity or energy expenditure measurements in large populations. The DLW method is one of the most accurate and valid systems used for evaluating total energy expenditure (TEE) under free-living conditions [8]. The method can also be used to assess physical activity energy expenditure (AEE) and physical activity level (PAL) in association with measurement of basal metabolic rate (BMR) [9,10]. However, because it too is expensive, the DLW method is not suitable for epidemiological studies.

Therefore, it is necessary to develop new, inexpensive methods of evaluating the levels of physical activity in large populations. Web-based measurement systems have been suggested as being convenient for collecting self-reported physical activity data compared with traditional questionnaires. In addition, Web-based systems may improve measurement accuracy because of the interactive communication associated with responses to questions. Responses can be obtained from individuals anywhere and at any time because Web-based systems can be completed using terminal devices, such as cell phones and smartphones. Two recent studies validated Web-based systems for measuring physical activity against the DLW method and found their validity was equivalent to that of traditional questionnaires [11,12]. More accurate Web-based systems for measuring physical activity are required. The successful and systematic collection of demographic and lifestyle data is central to any epidemiological study. Therefore, such technologies as the Internet and mobile phones have great potential for use in this kind of study [13]. The purpose of the present study was to develop two Web-based physical activity systems and examine their validity against the DLW method.

Methods

Participant Recruitment and Exclusion Criteria

We recruited 20 healthy people (25–61 years, 10 men and 10 women) to participate in the study. Participants were recruited by email through advertisements in a local email newsletters of 5000 registrants and a campus of the university in Fukuoka, Japan.

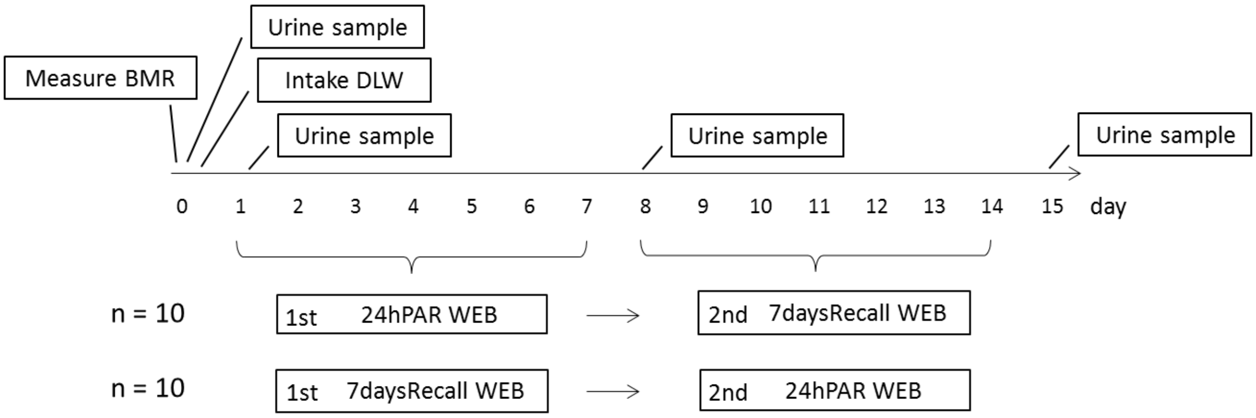
The participants received a ¥10,000 gift certificate for participating in the experiments. Inclusion criteria were that each participant should have one of the following: smartphone, tablet device, personal computer, or mobile phone; they also had to have Internet and email access. Exclusion criteria were metabolic disease, pregnancy, currently lactating, or being a competitive athlete; this was because these conditions alter normal energy expenditure. The study was conducted with the approval of the Ethics Committee of Fukuoka University, Fukuoka, Japan. All participants received a full explanation of the study’s purpose and content, and we confirmed that they understood this before they gave their written consent.

Study Overview

We developed two Web systems that functioned via a website and allowed for the automatic delivery of email. Diary [14] and 7-day recall [15] methods, developed previously, were adjusted so that they could be completed online. We named these the 24-hour Physical Activity Record Web version (24hPAR WEB) and the 7 days Recall Web version (7daysRecall WEB), respectively. The study took place over 2 weeks, and participants were requested to respond to an email sent at 8:00 every evening before bedtime using the terminal device of their choice.

Figure 1 shows the time schedule for the DLW and Web-based system measurements. For the first week, 10 participants were randomly selected to complete the 24hPAR WEB, with the remaining 10 participants answering the 7daysRecall WEB. In the second week, the participants used the Web-based system that they had not completed during the first week. To determine the validity of the systems, we calculated TEE, AEE, and PAL each week. TEE was determined using the DLW method, while AEE and PAL were based on the calculation of BMR from expired gas analysis.

Figure 1. Schedule for measurement of doubly labeled water (DLW) and delivery of the Web-based systems over a 2-week period. 7daysRecall WEB = 7 days Recall Web, 24hPAR WEB = 24-hour Physical Activity Record Web, BMR = basal metabolic rate.



Web-Based Physical Activity Measurements

The diary method [14] required that participants record activities performed in each of four different categories (work-related activities, way to work, leisure-time activities, and sports activities) every 15 minutes for a cross-tabulation. In total, there was a choice of 66 different activity behaviors among the four categories. Multimedia Appendix 1 shows a sample screen of the 24hPAR WEB activities and a list of their metabolic equivalents (METs). For every 15 minutes over each 24-hour period for 1 week, participants chose the behavior on the left of the screen and then completed the table on the right. For the 24hPAR WEB, we added an extra 33 types of behavior compared with the original 33 [14] to ensure that we included a greater variety of lifestyle behaviors. We included behaviors suggested by the NHK's (Japan Broadcasting Corporation) National Lifetime Study 2010 [16] and the Sasakawa Sports Foundation's 2010 National Sports-Life Survey [17] as being common to Japanese adults. We determined the intensity (in METs) of the new activities based on previous research [18]. Response results, expressed as the intensity of each activity, were recorded on a Web server every 15 minutes.

We explained to the participants how to use the system during a briefing session. Since this system operates in JAVA script, launching the software requires access to the URL of the email address that was registered.

The TEE (in MJ) for each 24-hour period was calculated using the following equation: BMR × 24-hour average METs × 1.1 ÷ 0.9 × 4.184 × 10^{-3} [19].

The AEE (in MJ) for each 24-hour period was calculated using the following equation: TEE × 0.9 – BMR.

BMR was the value obtained from expired gas analysis. The constant 1.1 reflected the resting metabolic rate for sitting and was equivalent to an increase of approximately 10% of the BMR. The resting metabolic rate in the sitting position (1 MET) is different from the supine position BMR with fasting. The constant 0.9 reflected dietary thermogenesis of approximately 10%, and the conversion factor of 4.184 × 10^{-3} was used to transform the values from kilocalories to megajoules.

The 7-day recall method [15] requires an individual to recall the previous 7 days' physical activity behavior during an interview. In contrast, the participants in the present study had to input data into the 7daysRecall WEB system every day for 1 week. The day was divided into three periods—morning, afternoon, and evening—and the participants had to recall their physical activity during those periods as well as the intensity of that activity. Multimedia Appendix 2 describes the 7daysRecall WEB, the input screen, illustrations of the activities, and their intensity. The activity intensity was divided as follows: light (1.5-2.9 METs), moderate (3.0-3.9 METs), moderate to high (4.0-5.9 METs), high (6.0-7.9 METs), and very high (≥8.0 METs). The participants were able to select the intensity of the physical activity based on illustrations that demonstrated the differences between the various intensity classifications. In addition, to reduce errors in the choice of activities that represented a specific intensity, the first and second days of the 7daysRecall WEB assessment included an interactive quiz. The quiz required that the participants match activities with the appropriate intensity, and it demonstrated how the participants were required to complete the assessment. The quiz continued until the participant answered two questions correctly in a row. For scoring, the following METs levels were assigned to each class of activities: sleep = 0.9 METs; light level of activity = 2.2 METs; moderate = 3.5 METs; moderate to high = 4.5 METs; high = 7.0 METs, and very high = 10 METs.

Times of very light-intensity activity, such as reading and television viewing, were not included in the selection screen. Therefore, we subtracted sleep and other activity times from 24 hours and classed activity intensity for this time as very light (= 1.3 METs).

Activity-intensity data for both systems were submitted over the Internet and converted to energy expenditure, which was then transferred to a server. Feedback regarding energy expenditure during the experimental period was not provided to the participants to avoid influencing individual behavior.

DLW Method

TEE was measured using the previously described DLW method over a 14-day period [7]. On arrival at the testing facility on day 0, participants gave a urine sample for the measurement of baseline ^2H and ^{18}O enrichment. Between 8:00 and 9:00 AM, a premixed dose containing approximately 0.12 g/kg predicted total body water of ^2H_2O (99.8 atom%; Taiyo Nippon Sanso, Tokyo, Japan) and 2.5 g/kg predicted total body water of $H_2{}^{18}O$ (10.0 atom%; Taiyo Nippon Sanso) was given to each participant to drink. Urine samples were collected at the following time points: twice on the next morning (day 1) and twice on the mornings of days 8 and 15. Aliquots of the urine samples were stored frozen at –15°C for later analysis by isotope ratio mass spectrometry (Hydra 20-20 Stable Isotope Mass Spectrometers; SerCon Ltd, Crewe, UK). The gas for equilibration of ^{18}O was CO_2, and H_2 was used for the ^2H. We used platinum catalyst for equilibration of ^2H. Each sample and the corresponding reference were analyzed in duplicate. The average standard deviations for the analyses were 0.7‰ for ^2H and 0.05‰ for ^{18}O [7].

The ^{18}O and ^2H dilution spaces (N_O and N_d) were determined by dividing the dose of the administered tracer (as moles of ^2H- or ^{18}O-labeled water) by the intercept method (^2H and ^{18}O enrichments at time 0) [20,21]. N_d/N_O in the present study was mean 1.027 (SD 0.007), range 1.011-1.043, which is similar to values reported in previous studies [7,22]. Therefore, total body water (mol) was calculated as the mean of N_d (mol) divided by 1.041 for the dilution space (estimated by ^2H and N_O [mol]) divided by 1.007 for dilution space (estimated by ^{18}O) [22].

CO_2 production rates were determined using the following equation: 0.4554 × total body water × (1.007 × ^{18}O elimination rate – 1.041 × ^2H elimination rate). For this calculation, we used assumptions of isotope fractionation applied to breath water using equation A6 of Schoeller et al [8] with the revised dilution space constant of Racette et al [22]. The TEE calculation was performed using the modified Weir formula [23] based on CO_2 production rates, and the assumed respiratory quotient was 0.85. PAL was calculated as TEE/BMR [9].

Measurement of BMR

We measured BMR between 6:00 and 6:30 AM on day 0 of the study using indirect calorimetry (ARCO 2000; ARCO System, Chiba, Japan). CO_2 production and O_2 consumption were measured after an overnight fast and 30 minutes of rest. Participants were required to be transported to the laboratory by car to keep physical activity to a minimum. CO_2 production and O_2 consumption were converted to BMR through the Weir equation [23]. We used a variation in O_2 consumption of less than 25 mL/min to determine whether the collection was acceptable [24]. Each participant was monitored periodically to ensure that he or she remained awake. Data were collected in a thermally regulated environment with minimal light and noise. The calorimeter system was calibrated before each measurement.

Statistical Analysis

All statistical analyses were performed using SPSS for Windows 19.0 (IBM Corporation, Somers, NY, USA). All data are shown as mean (SD); for participant characteristics, a range is also given. We compared men versus women for general characteristics by the independent *t* test.

We calculated zero-order and partial Pearson correlation coefficients as measures of association between AEE measured by DLW and by 24hPAR WEB; TEE measured by DLW and by 24hPAR WEB; PAL measured by DLW and by 24hPAR WEB; AEE measured by DLW and by 7daysRecall WEB; TEE measured by DLW and by7daysRecall WEB; and PAL measured by DLW and by 7daysRecall WEB. Using Bland-Altman plots, we related the difference in AEE between 24hPAR WEB and DLW (y-axis) to the arithmetic mean of AEE for 24hPAR WEB and DLW (x-axis) [25]. In addition, we related the difference in AEE between 7days WEB and DLW (y-axis) to the arithmetic mean of AEE for 7days WEB and DLW (x-axis) [25]. The limits of agreement are given as ±1.96 SD of the difference. To determine the source of error for TEE between 24hPAR WEB and DLW, and for TEE between 7daysRecall WEB and DLW, we calculated the Pearson correlation coefficient to determine the association of the difference between 24hPAR WEB and DLW, and the difference between 7daysRecall WEB and DLW, for TEE, plotted by gender. We defined a significant difference as $P <. 01$.

Results

Table 1 displays the general characteristics of the study population. We intended to recruit a heterogeneous group (both genders and a wide age and body mass index range), and this is demonstrated in the table. Of the 20 participants, 7 (35%) were overweight (body mass index >25 kg/m^2). A total of 10 (50%) were employed full-time, 2 (10%) were employed part-time, 4 (20%) were self-employed, and 4 (20%) were full-time homemakers. Men had significantly greater BMRs ($P < .001$), DLW-derived AEE ($P < .001$), and DLW-derived TEE ($P < .001$) than women. There was no significant difference between men and women in PAL, as determined by the DLW method ($P = .09$).

Table 1. General characteristics of the study population.

	All participants		Male (n = 10)			Female (n = 10)			*P* value
	Mean	SD	Mean	SD	Range	Mean	SD	Range	
Age (years)	42.6	10.9	45.2	12.3	26–61	40.0	9.1	25–51	.30
BMI[a] (kg/m^2)	24.0	3.8	26.4	3.4	19.2–33.9	21.5	2.4	19.2–27.4	.001
Weight (kg)	66.6	16.3	79.2	13.9	57.5–108.5	54.0	3.6	48.7–58.0	<.001
Height (cm)	165.8	10.4	172.6	7.3	161.9–185.4	159.0	8.5	141.8–171.0	.001
Measured BMR[b] (MJ)	5.68		6.50	0.89	4.77–8.00	4.87	0.57	4.40–6.24	<.001
TEE_{DLW}[c] (MJ/day)[d]	10.46	2.58	12.39	2.09	8.46–15.50	8.53	1.18	7.41–11.26	<.001
AEE_{DLW}[e] (MJ/day)[d]	3.73	1.35	4.65	1.07	2.85–6.16	2.82	0.89	1.48–4.17	<.001
PAL_{DLW}[f] (TEE/measured BMR)[d]	1.83	0.18	1.90	0.12	1.68–2.09	1.76	0.22	1.42–2.14	.09

[a] Body mass index.

[b] Basal metabolic rate.

[c] Total energy expenditure measured by doubly labeled water.

[d] Average value of the total over 2 weeks.

[e] Activity energy expenditure measured by doubly labeled water.

[f] Physical activity level measured by doubly labeled water.

Average input times using the terminal devices were 8 minutes and 10 seconds for the 24hPAR WEB and 6 minutes and 38 seconds for the 7daysRecall WEB system. The average number of input data in the 24hPAR WEB was 8.6 (SD 2.6) types; this refers to the average number of types of behavior, such as reading a book, watching television, and walking. The average number of inputted hours in 7daysRecall WEB was 6.0 (SD 1.5) times; this refers to the average number of inputs for each intensity of activity in the morning, afternoon, and evening. Compliance relating to data input using the terminal devices was 122 of 140 days (87.1%) with the 24hPAR WEB and 133 of 140 days (95.0%) with the 7daysRecall WEB. When the system administrator identified a participant who was not compliant, this was recorded, and the participant received a phone call the next day to remind him or her to complete the input. Using this method, 100% data were collected in this study. Figure 2 shows the response time over the 7 days for each measurement system based on the weblog data. The figure shows that the response time became gradually faster.

The mean AEE was 3.90 (SD 1.43) MJ measured by the 24hPAR WEB method and 3.67 (SD 1.48) MJ by the DLW method. The mean TEE was 10.65 (SD 2.57) MJ by the 24hPAR WEB method and 10.39 (SD 2.68) MJ by the DLW method. Mean PAL was 1.87 (SD 0.20) by 24hPAR WEB and 1.82 (SD 0.22) by DLW. Mean AEE was 4.29 (SD 1.94) MJ by the 7daysRecall WEB method and 3.80 MJ (SD 1.36) by DLW. The mean TEE was 11.08 (SD 2.82) MJ by the 7daysRecall WEB method and 10.53 (SD 2.58) MJ by DLW, and mean PAL was 1.96 (SD 0.32) for the 7daysRecall WEB method and 1.84 (SD 0.19) by DLW. There were no significant differences between the Web-based measurements and the DLW method for the above energy expenditure variables.

Figure 3 shows the Pearson correlation coefficient between the 24hPAR WEB and DLW methods for daily AEE. AEE measured by 24hPAR WEB and by DLW was correlated (r = .679, P < .001). Correlation coefficients for TEE and PAL by the 24hPAR WEB and the DLW methods were r = .874 (P < .001) and r = .404 (P = .08), respectively. Figure 4 shows the Bland-Altman plot for AEE as measured by 24hPAR WEB compared with DLW. The mean difference for the 24hPAR WEB and DLW methods was small (0.23 MJ), and the limits of agreement were 2.33 MJ (±1.96 SD). The test for trend was not statistically significant. The regression equation was y = –0.03x + 0.36 (r = –.382, P = .87). In the Bland-Altman plot for TEE measured by 24hPAR WEB compared with DLW, the mean difference between the two methods was small (0.26 MJ), and the limits of agreement were 2.59 MJ (±1.96 SD). The test for trend was not statistically significant. The regression equation was y = –0.04x + 0.72 (r = –.085, P = .72).

The Pearson correlation for AEE between the 7daysRecall WEB and DLW methods was r = .144 (P = .54). TEE as measured by 7daysRecall WEB and DLW was correlated (r = .590, P = .006). The correlation coefficient for PAL as measured by the 7daysRecall WEB and DLW methods was r = –.085, with no significant correlation.

Figure 5 shows the Bland-Altman plot for AEE as measured by 7daysRecall WEB compared with DLW. The mean difference for the 7daysRecall WEB and the DLW method was small (0.49 MJ), and the limits of agreement were large at 4.32 MJ (±1.96 SD). The test for trend was not statistically significant. The regression equation was y = 0.59x – 1.93 (r = .343, P = .14). In the Bland-Altman plot for TEE measured by 7daysRecall WEB compared with DLW, the mean difference between the two methods was small (0.55 MJ), and the limits of agreement were large at 4.80 MJ (±1.96 SD). The test for trend was not statistically significant. The regression equation was y = 0.11x – 0.66 (*r* = .110, *P* = .65).

Figure 6 shows the correlation of the difference between the DLW and 24hPAR WEB measures of TEE and the difference between the DLW and 7daysRecall WEB measures of TEE (*r* = .673, *P* = .001). The figure demonstrates the profile of the methods regarding overestimation or underestimation of TEE. Most of the women and 1 man overestimated TEE in both the 24hPAR WEB and the 7daysRecall WEB methods (Figure 6).

Figure 2. Input time trends (in minutes and seconds for days 1 to 7) for each Web-delivered system. 7daysRecall WEB = 7 days Recall Web, 24hPAR WEB = 24-hour Physical Activity Record Web.

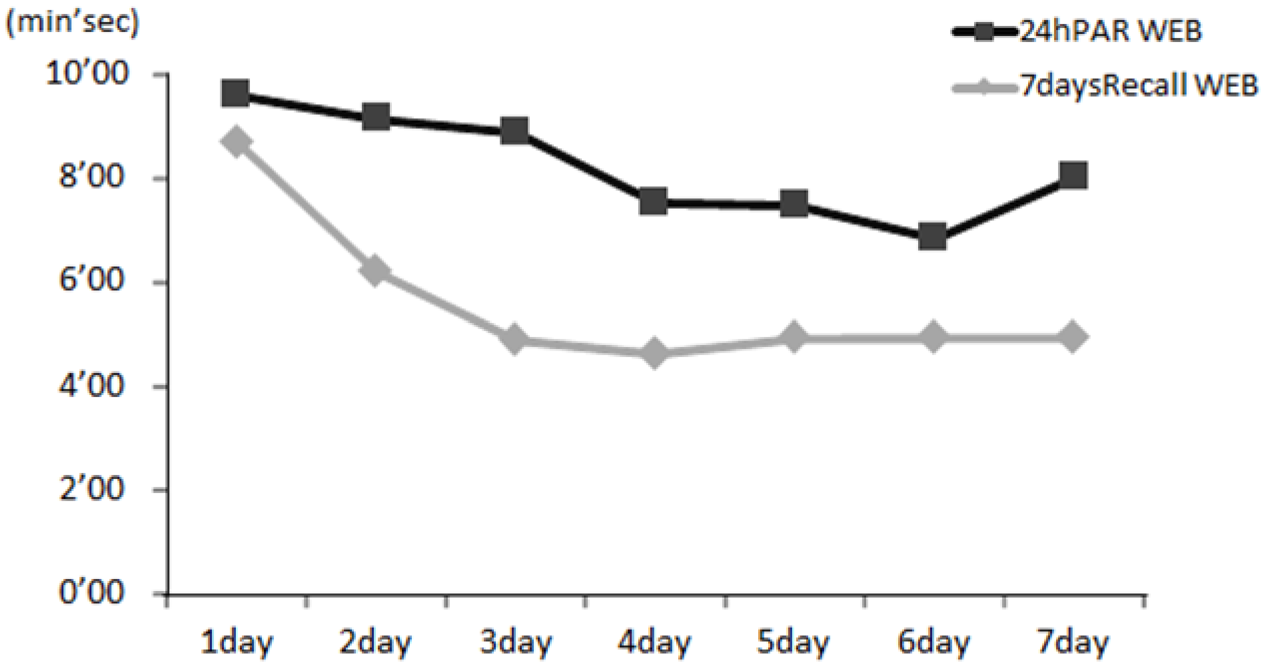


Figure 3. Pearson correlation coefficient for daily activity energy expenditure (AEE) measured by the 24-hour Physical Activity Record WEB (24hPAR WEB) and doubly labeled water (DLW) methods. CI = confidence interval.

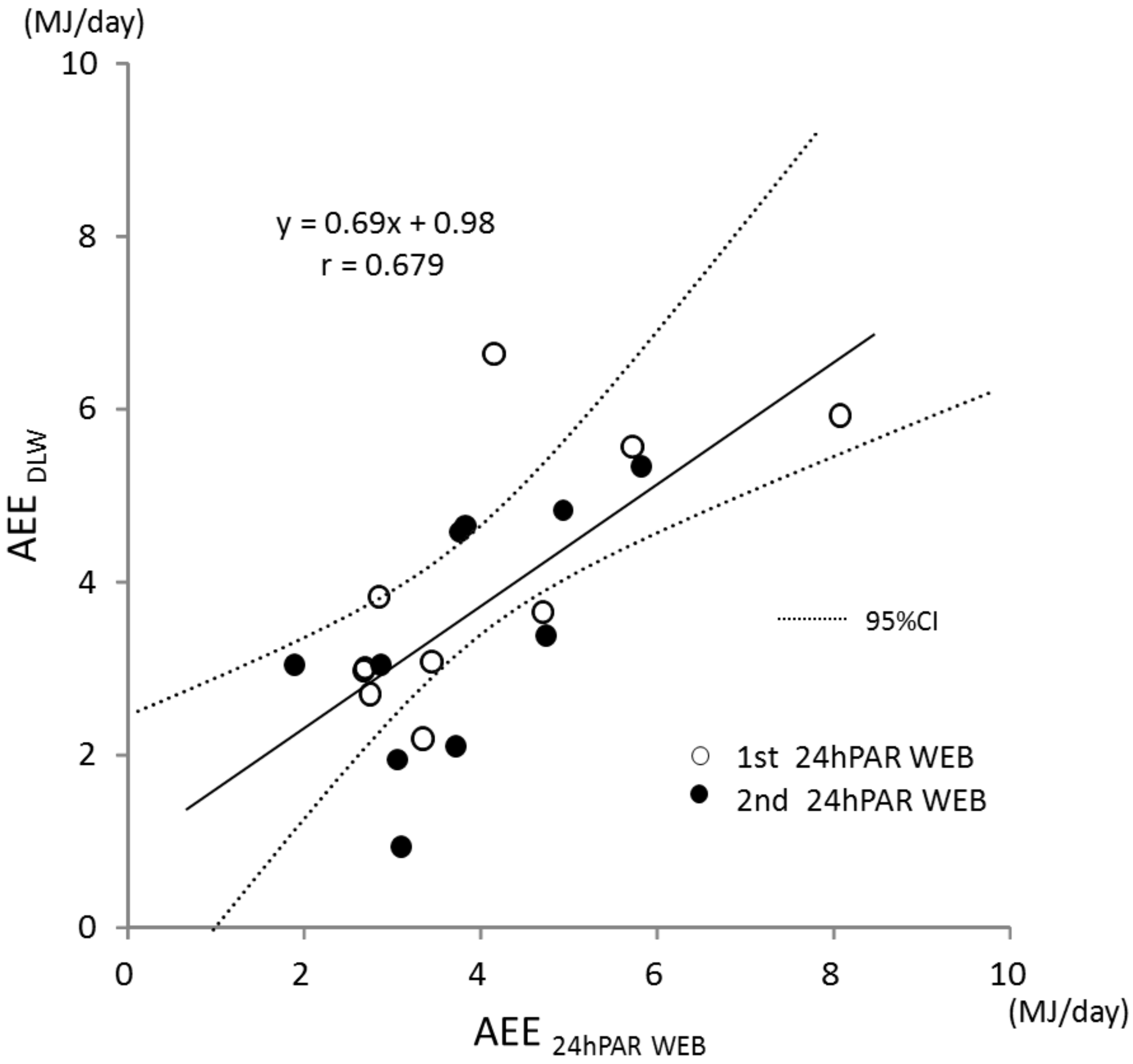


Figure 4. Bland–Altman plot illustrating the difference in activity energy expenditure (AEE) between the 24-hour Physical Activity Record Web (24hPAR WEB) and the doubly labeled water (DLW) methods.

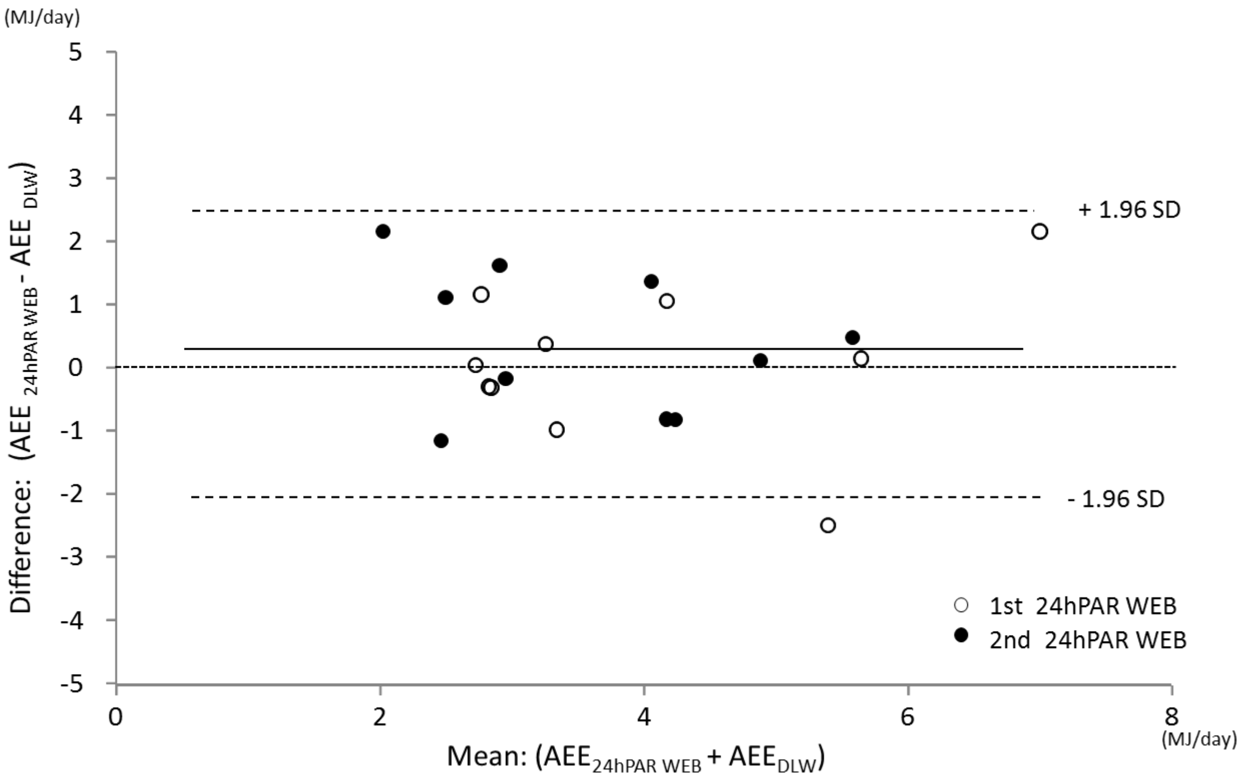


Figure 5. Bland–Altman plot illustrating the difference in activity energy expenditure (AEE) between the 7 days Recall Web (7daysRecall WEB) and the doubly labeled water (DLW) methods.

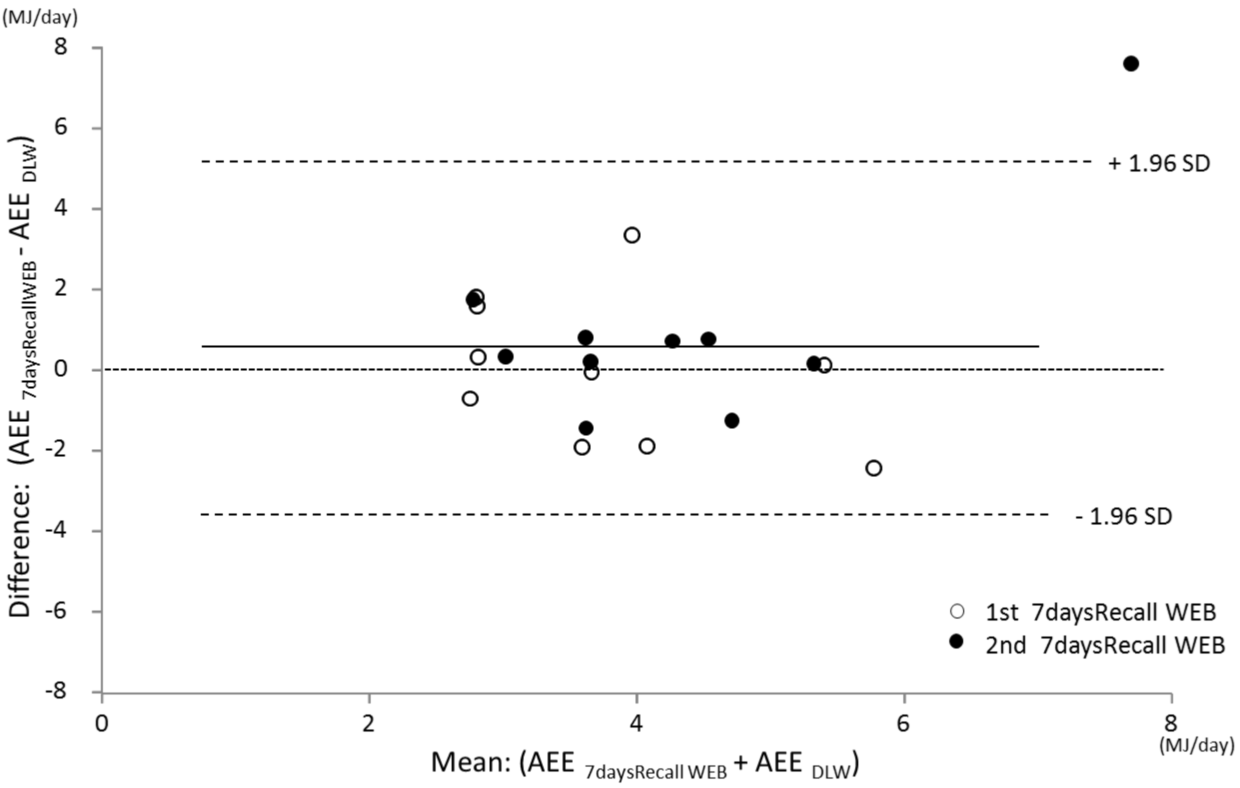
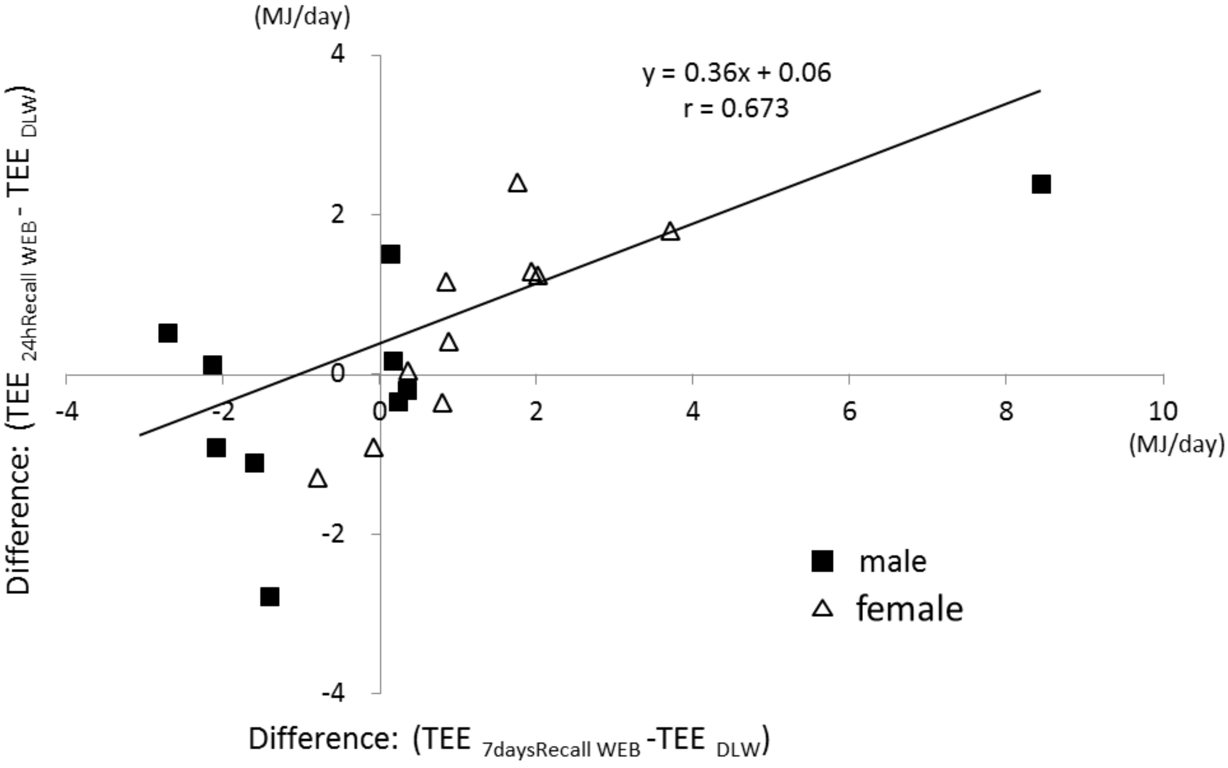


Figure 6. Correlation of the difference between the doubly labeled water (DLW) and 24-hour Physical Activity Record Web (24hPAR WEB) measures of total energy expenditure (TEE) and the difference between the DLW and 7 days Recall Web (7daysRecall WEB) measures of TEE.



Discussion

In this study, we developed a new physical activity assessment method using Web-based measurement systems and validated them against the DLW method. These measurement systems are low in cost, practical, and user friendly. They can be completed quickly, with an input time of less than 10 minutes. With both systems, the daily response time tended to become progressively faster; this was particularly the case with the 7daysRecall WEB method. It is possible that increased operational familiarity aided quicker recall of behavior. The systems can measure daily physical activity in many people simultaneously because they are compatible with many terminal devices, such as cell phones, smartphones, and personal computers, allowing complete assessments to be made via the Internet. These two systems were an improvement compared with the diary methods [14] and interview methods [15] developed previously.

Diary methods for the assessment of physical activity have been used since the 1960s [26]. The method requires that individuals record their physical activity behavior every 15 minutes. We used this method in a previous study [27], though the activities selected for the present study were modified from a previous study [14]. An advantage of the Web-based measurement systems used in the present study is that they are compatible with a variety of terminal devices. They also allowed for a greater variety of physical activity behaviors to be assessed, and they simplified the recording of data.

In the present study, we found a significant association between 24hPAR WEB and DLW measurements of AEE. This result is important because BMR accounts for 60%-75% of TEE [6]. In addition, the relationship between 24hPAR WEB and DLW measurements of TEE was similar to that between 3-axis accelerometry and DLW measurements of TEE [7]. This suggests that TEE can be assessed accurately in under 10 minutes using the 24hPAR WEB. The 24hPAR WEB allows for the assessment of swimming, cycling, and climbing activities, which cannot be measured using an accelerometer. The 24hPAR WEB can also classify such activities as traffic behavior, sleeping, working, and leisure time.

The correlation between 24hPAR WEB and DLW measurements of TEE in the present study was similar to that found in previous research [14]. The present study also demonstrated that physical activity can be reported on a single digital screen without the need for cross-tabulation from several pieces of paper. A significant finding was that the accuracy of 24hPAR WEB measurement of AEE was superior to that in conventional physical activity questionnaires when compared with the DLW method. Bonn et al reported a Spearman correlation between Active-Q, a Web-based questionnaire, and DLW measurements of TEE of $r = .52$ ($P < .001$, $n = 37$) [11]; similarly, Ishikawa-Takata et al reported a Spearman correlation between Japan Arteriosclerosis Longitudinal Study Physical Activity

Questionnaire and DLW measurements of TEE of r = .742 (P <.001, n = 226) [28]. A weakness of the present study was that we had fewer participants than in previous studies; however, we saw a moderately high correlation even with AEE (Figure 3). Our study findings confirm that the diary method is the most accurate for self-reporting.

Figure 4 indicates that the error range of the estimate did not change regardless of the highest and lowest average values for AEE. Because there was a meaningful correlation with the DLW methods, and the error range was small (±2.33 MJ, 95% confidence interval) regardless of the size of the estimate, we found that the 24hPAR WEB method provided better results than methods used in previous studies [11,14]. When using assessment tools that require greater effort for recalling behavior, such as the Active-Q [11] and the 7daysRecall WEB in the present study, study participants have been found to over- or underestimate their physical activity. With regard to the time of each behavior, they may be able to recall impressive behavior but unable to recall nonimpressive behavior. However, for the 24hPAR WEB, we observed no significant over- or underestimation. One reason for this finding may be that selecting the order of behavior after awaking results in greater accuracy. Another reason may be that behaviors were selected every 15 minutes.

The 7-day physical activity recall method was originally developed in 1979 for use in the Stanford Five-City Project [15]. An advantage of this method is that the previous 7 days of physical activity could be estimated from a 15-minute interview [15]. For the present study, we were unable to develop an interactive communication system that could simulate an interview. Therefore, we devised a method of effective communication in a manner that obtained similar information to that obtained in a face-to-face interview. This method included the following: (1) input once per day recalling that day's activities, compared with the original 7-day physical activity recall frequency of once per week, (2) illustrations of physical activity intensity, and (3) quizzes over the first 2 days that required participants to select the intensity among activities (Multimedia Appendix 2). The quiz questions continued until the participants were able to answer two questions correctly in a row. Those who did not look at the illustrations were unable to answer correctly; consequently not allowing them to continue until they answered correctly reduced potential mistakes in the self-reported exercise intensity of their activities.

Compared with previous studies that examined the 7-day recall [29-31] method, the 7daysRecall WEB predicted TEE accurately. Figure 5 indicates that the error range of the estimate would not change regardless of the high and low average value of AEE. The 95% confidence interval was ± 4.32 MJ, demonstrating a large variation compared with 24hPAR WEB. However, the error was small compared with the 7-day recall diary method [29-31], and it was similar to that of the Active-Q [11].

In using the 7daysRecall WEB, 1 man greatly overestimated (3SD above the average) energy expenditure. Therefore, we excluded his data and reanalyzed the data (n = 19). The recalculated Pearson correlation coefficient for TEE between the 7daysRecall WEB and DLW methods was r = .788 (P <.001). AEE measured by 7daysRecall WEB and DLW was not significantly correlated (r = .346); nor was and PAL measured by 7daysRecall WEB and DLW (r = -.001). After completion of the study, we found that the reason for energy expenditure being overestimated by the male participant was that, regardless of his physical activity, the intensity level he inputted corresponded to light-intensity child care time. To prevent this type of mistake, we programmed the system to request that the participant confirm each input. However, accurate recording of physical activity behavior appears to be a limitation of self-reported physical activity assessment methods [32].

The present study employed a crossover design, whereby participants alternately used two Web-based methods. With respect to response time, there was no difference (P = .32) between the group that used 24hPAR WEB during the first week (450, SD 163 seconds) and the group that used 24hPAR WEB during the second week (596, SD 329 seconds). There was likewise no difference (P = .68) between the group that used 7daysRecall WEB during the first week (353, SD 225 seconds) and the group that used 7daysRecall WEB during the second week (318, SD 140 seconds). In Figure 3, Figure 4, and Figure 5, correlation charts present the validity of the Web-based methods compared with the DLW method. No difference in the validity was evident as a result of using the crossover design. However, future investigations should examine this issue of the crossover design, since we had just 10 participants in each group.

With both systems, women tended to overestimate energy expenditure (Figure 6). Of the 7 women who made such overestimations, 6 were homemakers or were employed part-time. A possible reason for this estimation relates to the difficulty in determining the intensity of child care and housework activity. Housework and child care require complex physical activity. It is possible that, rather than selecting the time spent moving, the participants selected the time engaged in housework and child care. However, further study is required to investigate this issue.

Another limitation of this study was that conventional cell phones cannot support JavaScript. Therefore, conventional cell phones could not operate the 24hPAR WEB. In Japan, 95% of people aged 30-49 years have access to the Internet, whereas just 32.9% of those aged 70-79 years have such access [33]. Therefore, sampling is an important issue when using the 24hPAR WEB for epidemiological studies in the elderly. Compliance during the 7-day input is also an important consideration for obtaining accurate measurements.

In recent years, public health research has focused on disease prevention and community intervention to promote physical activity [34]. Over the last decade, many programs have also developed Web-based interventions to help increase physical activity [35,36]. For research targeting large groups of individuals, evaluation of physical activity is important in determining the effects of interventions. The Web-based measurements developed in the present study will be useful for accurately assessing physical activity at low cost.

In conclusion, the 24hPAR WEB appears to be valid for estimating AEE and TEE, and the 7daysRecall WEB appears to be valid for estimating TEE. Both methods are effective for collecting physical activity data in epidemiological studies. The 24hPAR WEB is more accurate than the 7daysRecall WEB, and it is useful for evaluating physical activity behavior and the effect of interventions. The input time for the 7daysRecall WEB is short. The system is easy to operate and suitable for evaluating physical activity in large communities.

Acknowledgments

The authors thank the participants for their cooperation in this study. This work was supported by grants from Grant-in-Aid for Research Activity Start-up (No. 22800075; to the first author), Grant-in-Aid for Young Scientists (B) (No. 24700740; to the first author), The Twenty-Eight Research-Aid (2011) Meiji Yasuda Life Foundation of Health and Welfare (to the first author), Grant-in-Aid for JSPS Fellows (No.11J00333; to YY), and the Sports Psychology Laboratory and Central Research Institute for Physical Activity, Fukuoka University.

Conflicts of Interest

The authors have a patent application for an improved version of the 24h Physical Activity Record WEB (Pat. Pend. 2011-227032 Japan Patent Office).

Multimedia Appendix 1

24h Physical Activity Record WEB.

[PDF File (Adobe PDF File), 76KB - jmir_v14i5e123_app1.pdf]

Multimedia Appendix 2

7daysRecall WEB.

[PDF File (Adobe PDF File), 250KB - jmir_v14i5e123_app2.pdf]

References

1. World Health Organization. Global Health Risks: Mortality and Burden of Disease Attributable to Selected Major Risks. Geneva: World Health Organization; 2009.
2. Lakka HM, Laaksonen DE, Lakka TA, Niskanen LK, Kumpusalo E, Tuomilehto J, et al. The metabolic syndrome and total and cardiovascular disease mortality in middle-aged men. JAMA 2002 Dec 4;288(21):2709-2716. [Medline: 12460094]
3. Lakka TA, Laaksonen DE. Physical activity in prevention and treatment of the metabolic syndrome. Appl Physiol Nutr Metab 2007 Feb;32(1):76-88. [doi: 10.1139/h06-113] [Medline: 17332786]
4. Pate RR, Pratt M, Blair SN, Haskell WL, Macera CA, Bouchard C, et al. Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. JAMA 1995 Feb 1;273(5):402-407. [Medline: 7823386]
5. Montoye HJ, Kemper HC, Saris WH, Washburn RA. Measuring Physical Activity and Energy Expenditure. Champaign, IL: Human Kinetics; 1996.
6. Neilson HK, Robson PJ, Friedenreich CM, Csizmadi I. Estimating activity energy expenditure: how valid are physical activity questionnaires? Am J Clin Nutr 2008 Feb;87(2):279-291 [FREE Full text] [Medline: 18258615]
7. Yamada Y, Yokoyama K, Noriyasu R, Osaki T, Adachi T, Itoi A, et al. Light-intensity activities are important for estimating physical activity energy expenditure using uniaxial and triaxial accelerometers. Eur J Appl Physiol 2009 Jan;105(1):141-152. [doi: 10.1007/s00421-008-0883-7] [Medline: 18853176]
8. Schoeller DA, Ravussin E, Schutz Y, Acheson KJ, Baertschi P, Jéquier E. Energy expenditure by doubly labeled water: validation in humans and proposed calculation. Am J Physiol 1986 May;250(5 Pt 2):R823-R830. [Medline: 3085521]
9. Manini TM, Everhart JE, Patel KV, Schoeller DA, Colbert LH, Visser M, et al. Daily activity energy expenditure and mortality among older adults. JAMA 2006 Jul 12;296(2):171-179. [doi: 10.1001/jama.296.2.171] [Medline: 16835422]
10. Ishikawa-Takata K, Tabata I, Sasaki S, Rafamantanantsoa HH, Okazaki H, Okubo H, et al. Physical activity level in healthy free-living Japanese estimated by doubly labelled water method and International Physical Activity Questionnaire. Eur J Clin Nutr 2008 Jul;62(7):885-891. [doi: 10.1038/sj.ejcn.1602805] [Medline: 17522602]
11. Bonn SE, Trolle Lagerros Y, Christensen SE, Möller E, Wright A, Sjölander A, et al. Active-Q: validation of the web-based physical activity questionnaire using doubly labeled water. J Med Internet Res 2012 Feb;14(1):e29 [FREE Full text] [doi: 10.2196/jmir.1974] [Medline: 22356755]

12. Bexelius C, Löf M, Sandin S, Trolle Lagerros Y, Forsum E, Litton JE. Measures of physical activity using cell phones: validation using criterion methods. J Med Internet Res 2010 Jan;12(1):e2 [FREE Full text] [doi: 10.2196/jmir.1298] [Medline: 20118036]
13. Ekman A, Litton JE. New times, new needs; e-epidemiology. Eur J Epidemiol 2007;22(5):285-292. [doi: 10.1007/s10654-007-9119-0] [Medline: 17505896]
14. Koebnick C, Wagner K, Thielecke F, Moeseneder J, Hoehne A, Franke A, et al. Validation of a simplified physical activity record by doubly labeled water technique. Int J Obes (Lond) 2005 Mar;29(3):302-309. [doi: 10.1038/sj.ijo.0802882] [Medline: 15672111]
15. Sallis JF, Haskell WL, Wood PD, Fortmann SP, Rogers T, Blair SN, et al. Physical activity assessment methodology in the Five-City Project. Am J Epidemiol 1985 Jan;121(1):91-106. [Medline: 3964995]
16. NHK (Japan Broadcasting Corporation). 2010. National Lifetime Study 2010 URL: http://www.nhk.or.jp/bunken/summary/yoron/lifetime/pdf/110223.pdf [accessed 2012-08-09] [WebCite Cache ID 69mniRHTY]
17. Sasakawa Sports Foundation. The 2010 SSF National Sports-Life Survey. Tokyo: Sasakawa Sports Foundation; 2010.
18. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, et al. Compendium of physical activities: an update of activity codes and MET intensities. Med Sci Sports Exerc 2000 Sep;32(9 Suppl):S498-S504. [Medline: 10993420]
19. Tanaka S. Metabolic equivalent and Basal metabolic rate. J Health Phys Educ Recreat 2009;59(10):657-663 [FREE Full text]
20. Coward WA. Calculation of pool sizes and flux rates. In: Prentice AM, editor. The Doubly-Labelled Water Method for the Measurement of Energy Expenditure. Vienna: International Atomic Energy Agency; 1990:48-65.
21. Yamada Y, Masuo Y, Yokoyama K, Hashii Y, Ando S, Okayama Y, et al. Proximal electrode placement improves the estimation of body composition in obese and lean elderly during segmental bioelectrical impedance analysis. Eur J Appl Physiol 2009 Sep;107(2):135-144. [doi: 10.1007/s00421-009-1106-6] [Medline: 19533166]
22. Racette SB, Schoeller DA, Luke AH, Shay K, Hnilicka J, Kushner RF. Relative dilution spaces of 2H- and 18O-labeled water in humans. Am J Physiol 1994 Oct;267(4 Pt 1):E585-E590. [Medline: 7943308]
23. Weir JB. New methods for calculating metabolic rate with special reference to protein metabolism. J Physiol 1949 Aug;109(1-2):1-9 [FREE Full text] [Medline: 15394301]
24. Turley KR, McBride PJ, Wilmore JH. Resting metabolic rate measured after subjects spent the night at home vs at a clinic. Am J Clin Nutr 1993 Aug;58(2):141-144 [FREE Full text] [Medline: 8338039]
25. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307-310. [Medline: 2868172]
26. Edholm OG. The Assessment of Habitual Activity in Health and Disease. Oslo: Williams and Wilkins; 1966:187-197.
27. Bouchard C, Tremblay A, Leblanc C, Lortie G, Savard R, Thériault G. A method to assess energy expenditure in children and adults. Am J Clin Nutr 1983 Mar;37(3):461-467 [FREE Full text] [Medline: 6829488]
28. Ishikawa-Takata K, Naito Y, Tanaka S, Ebine N, Tabata I. Use of doubly labeled water to validate a physical activity questionnaire developed for the Japanese population. J Epidemiol 2011 Mar 5;21(2):114-121 [FREE Full text] [Medline: 21258166]
29. Bonnefoy M, Normand S, Pachiaudi C, Lacour JR, Laville M, Kostka T. Simultaneous validation of ten physical activity questionnaires in older men: a doubly labeled water study. J Am Geriatr Soc 2001 Jan;49(1):28-35. [Medline: 11207839]
30. Conway JM, Seale JL, Jacobs DR, Irwin ML, Ainsworth BE. Comparison of energy expenditure estimates from doubly labeled water, a physical activity questionnaire, and physical activity records. Am J Clin Nutr 2002 Mar;75(3):519-525 [FREE Full text] [Medline: 11864858]
31. Mahabir S, Baer DJ, Giffen C, Clevidence BA, Campbell WS, Taylor PR, et al. Comparison of energy expenditure estimates from 4 physical activity questionnaires with doubly labeled water estimates in postmenopausal women. Am J Clin Nutr 2006 Jul;84(1):230-236 [FREE Full text] [Medline: 16825700]
32. Squire LR, Zola-Morgan S. Memory: brain systems and behavior. Trends Neurosci 1988 Apr;11(4):170-175. [Medline: 2469186]
33. Japan Ministry of Internal Affairs and Communications. 2010. Results of the Telecom use Trend Survey URL: http://www.soumu.go.jp/main_content/000064217.pdf [accessed 2012-08-09] [WebCite Cache ID 69mntzKXm]
34. Mummery WK, Brown WJ. Whole of community physical activity interventions: easier said than done. Br J Sports Med 2009 Jan;43(1):39-43. [doi: 10.1136/bjsm.2008.053629] [Medline: 18981043]
35. Laplante C, Peng W. A systematic review of e-health interventions for physical activity: an analysis of study design, intervention characteristics, and outcomes. Telemed J E Health 2011 Sep;17(7):509-523. [doi: 10.1089/tmj.2011.0013] [Medline: 21718092]
36. van den Berg MH, Schoones JW, Vliet Vlieland TP. Internet-based physical activity interventions: a systematic review of the literature. J Med Internet Res 2007;9(3):e26 [FREE Full text] [doi: 10.2196/jmir.9.3.e26] [Medline: 17942388]

Abbreviations

7daysRecall WEB: 7 days Recall Web

24hPAR WEB: 24-hour Physical Activity Record Web
AEE: activity energy expenditure
BMR: basal metabolic rate
DLW: doubly labeled water
METs: metabolic equivalents
Nd: 2H dilution space
NO: 18O dilution space
PAL: physical activity level
TEE: total energy expenditure

Edited by G Eysenbach; submitted 03.07.12; peer-reviewed by K Westerterp, S Bonn; comments to author 23.07.12; revised version received 17.08.12; accepted 25.08.12; published 25.09.12.

Please cite as:
Namba H, Yamaguchi Y, Yamada Y, Tokushima S, Hatamoto Y, Sagayama H, Kimura M, Higaki Y, Tanaka H
Validation of Web-Based Physical Activity Measurement Systems Using Doubly Labeled Water
J Med Internet Res 2012;14(5):e123
URL: http://www.jmir.org/2012/5/e123/
doi:10.2196/jmir.2253
PMID:23010345

Original Paper

The Umeå University Database of Facial Expressions: A Validation Study

Hanna Samuelsson[1*], MSc; Karl Jarnvik[1*], MSc; Hanna Henningsson[1*], MSc; Josefin Andersson[1*], MSc; Per Carlbring[1*], MSc, PhD

Department of Psychology, University of Umeå, Umeå, Sweden
[*]all authors contributed equally

Corresponding Author:
Per Carlbring, MSc, PhD
Department of Psychology
University of Umeå
IP
Umeå, SE90187
Sweden
Phone: 46 907867833
Fax: 46 700400399
Email: per@carlbring.se

Abstract

Background: A set of face stimuli, called the Umeå University Database of Facial Expressions, is described. The set consists of 30 female and 30 male models aged 17–67 years (M = 30.19, SD = 10.66). Each model shows seven different facial expressions (angry, surprised, happy, sad, neutral, afraid, and disgusted). Most models are ethnic Swedes but models of Central European, Arabic, and Asian origin are also included.

Objective: Creating and validating a new database of facial expressions that can be used for scientific experiments.

Methods: The images, presented in random order one at a time, were validated by 526 volunteers rating on average 125 images on seven 10-point Likert-type scales ranging from “completely disagree” to “completely agree” for each emotion.

Results: The proportion of the aggregated results that were correctly classified was considered to be high (M = 88%).

Conclusions: The results lend empirical support for the validity of this set of facial expressions. The set can be used freely by the scientific community.

(J Med Internet Res 2012;14(5):e136) doi:10.2196/jmir.2196

KEYWORDS

Face; Expression; Stimulus set; Emotion; Multiracial; Validity; Reliability

Introduction

There is a wealth of published research into face perception, face processing, and facial expressions using images of facial expressions of emotions [1]. Images of facial expressions of emotions are frequently used in research on emotions and are increasingly being used in neuroscience [2]. Recently, their use has been extended to the treatment of anxiety and depression, using a modified dot-probe task [3]. This paper presents the Umeå University Database of Facial Expressions, which is freely available to researchers. We present information about the database, along with the results of an Internet-based validation study.

The human face is an integral part of daily life. Facial muscles allow a wide range of expressions and functions [4]. These expressions communicate emotions in interaction with others and are an important part of the emotional experience [5]. The link between emotion and facial expression is the main driver of research interest in facial expressions and their interpretation [1].

Ekman and Friesen published their pioneering Pictures of Facial Affect (PFA) in 1976, which became the most frequently used set in research [2], greatly improving our understanding of the universality of the facial expression of emotion [6]. Today, however, this set of facial expressions has limitations, such as the small number of images, and is considered rather dated in regard to quality and the range of models, which adversely affects its ecological validity [2,7]. This and some other currently available sets do not contain models representing different ethnicities (KDEF, see [2,8]; FACES [9]).

The understanding and perception of emotions has been shown to be more accurate if those that are evaluating emotional expressions have the same ethnicity and national and regional background as the expressers. This may be because people of different ethnicities develop different nuances in their expressions. However, when different cultural groups spend more time together, the in-group advantage decreases [10]. Evaluation of facial expressions of emotions is also influenced by stereotypical ideas about people of other ethnicities [11].

The NimStim Set of Facial Expressions (NimStim) [1], the Japanese and Caucasian Facial Expressions of Emotion (JACFEE) [12], and the Montreal Set of Facial Displays of Emotion (MSFDE) [13] include models of different ethnicities but contain fewer than 145 images in total (JACFEE, MSFDE) [2]. They lack representation of different age groups, and it is not possible to use them in Internet-based experiments, even with password protection (NimStim) [1].

This project has attempted to address problems identified in previous sets of facial expressions and validation studies. The aim of producing the Umeå University Database of Facial Expressions was to create a database for Internet-based research, containing a large number of images across a spectrum of age, ethnicity, and gender.

This database has several advantages. First, it contains a large number of color images—a total of 424, posed by 60 models (2720x4080 pixels). The models express the most consistently recognized facial expressions of emotions, which are anger, surprise, happiness, sadness, fear, and disgust [14] as well as a neutral expression. Databases containing a large number of images are preferable because research often requires a large body of stimulus material in order to avoid the effects of habituation [2]. Second, the models represent different age groups, ethnicities, and genders, which gives the database good ecological validity with regard to these variables. Third, the database is available with password protection for scientific experiments on the Internet.

The aim of this validation study was to examine the extent to which facial expressions as depicted in the images were correctly interpreted as the intended emotion. It was done over the Internet in order to recruit participants with as great a range as possible of age, gender, and ethnicity. Swedish law does not, however, permit the registration of individual ethnicity. However, researchers based in a country without this restriction on reporting of individual ethnicities are free to do so after inspection of the photographs.

The genders of both model and rater may influence evaluation of facial expressions [15-19]. The age of the rater also plays an important role in the correct recognition of facial expression [20,21]. Validation studies commonly include only university or college students [2,7,8,12]. In order to reach a more heterogeneous group of people, we recruited participants both within and outside higher education.

Each image was evaluated by participants. Nuanced answer options were used in the validation study in order to reduce the risk of influencing responses to a specific expression. Response scales with fixed response options can be problematic, as different response scale formats may influence the results obtained [22]. Predetermined emotion labels can be regarded as a contextual variable that influences the participant's response to a specific expression [22,23]. In this validation study, participants were therefore given the option of rating expressions for several different emotions.

We hypothesized that the Internet-based validation study would provide sufficient data to support the validity of the Umeå University Database of Facial Expressions.

Methods

Participants

Data were collected from 526 participants. The mean age was 37.7 years (18–73, SD =13.0). 70% (369/526) were female and 30% (157/526) were male. Participants were recruited by disseminating information about the study via the local Swedish newspaper. All those who volunteered were allowed to take part in the study, and no financial compensation or remuneration was given.

Stimuli

The stimuli were 424 facial images from the Umeå University Database of Facial Expressions. A total of 60 subjects participated as amateur models (30 female, 30 male; 17-67 years old; M=30.19, SD=10.66). Most of these models were ethnic Swedes, but models of Central European, Arabic, and Asian origin were also included. During the photography session, models were instructed to display seven different facial expressions (angry, surprised, happy, sad, neutral, afraid, and disgusted). Instructions on how to make the facial expressions were based on the work of Ekman [24] and Ekman and Friesen [6] and presented to models before and during the sessions. Before the shoots, models were encouraged to practice making the facial expressions, and during the shoots models were instructed to make the expressions as they saw fit, to look at pictures of facial expressions (POFA) [25] and to move certain muscle groups. Models were instructed not to wear make-up. The shoots took place at Umeå University. Models were compensated for their participation by receiving uncompressed high-quality personal photographs. They also signed a legal agreement allowing the images to be used in research and education.

Selection of Images

The photo shoots produced over 8,000 images. The best image of each expression from each model was chosen to be validated empirically. However, a clear decision could not be made in four instances, and these images were therefore added to the validation phase, making a total of 424.

Validation Procedure

The validation procedure took place on the Internet. Before obtaining access to the images, the potential participant had to register his/her age, gender, and email address. A confirmatory email, including a unique login link, was sent to the registered email address, ensuring that all participants had registered a valid email address. Instructions to participants were to sit alone in a quiet, private setting and base the evaluations on their own opinion. Participants evaluated the images at their own pace and were free to evaluate as many images as they wished. They were allowed to discontinue the evaluation at any time and were free to return and continue at another time during a two-week period in October 2011. Images were randomly presented to each participant. However, each of the 424 images was presented only once. 526 participants started the validation process, rating an average of 125.5 out of 424 faces (SD=137.4).

Each of the 424 images (320x480 pixels, color) was presented on its own with the text "This person seems to be…" above each image. As shown in Figure 1, the options "angry", "surprised", "happy", "sad", "neutral", "afraid", and "disgusted" were presented below each image together with a 10-point Likert-type scale ranging from "completely disagree" to "completely agree". Participants could specify to what extent they agreed or disagreed with one or more of the listed emotions.

Data Analytic Procedure

We used a binary logistic model (specified through generalized linear equations), and variance–covariance for all models was assumed to be block diagonal but independent within a block defined by individual, which implies that we assumed that the scoring of one image did not affect the score given by that individual to the next randomized image.

The seven outcome variables were defined as 1/0 for each "true" emotion. The independent factors were gender and age of the rater and model and the rating score for the seven emotions. We studied the adjusted association between each outcome and the 11 independent factors. We present the estimated odds ratios and their 95% Wald confidence intervals (CIs) and their significance (see Supplemental Tables 1–7 in Appendix 1). All tests were two-sided. The results were considered significant if $P<.05$. All analyses were performed using SPSS, version 20 (SPSS, Inc., Chicago, IL).

We considered an image to be correctly classified if the highest score was given to the emotion corresponding to the true emotion. For example, if the emotion "sad" was scored 7 and the other emotions between 0 and 6 points, then sad would be counted as the response. That response would then be compared to the intended emotion when calculating the hit rate. In addition, in order to obtain a measure of the reliability of the interpretation, we also calculated the sum of the scores given to emotions that did not correspond to the true emotion, and the number of emotions rated.

Table 1 describes the scores given by raters for all images where models attempted to display a given emotion in the following columns: 1, "Percentage correctly perceived", how often the emotion that models displayed was rated higher than all other emotions; 2, "Number of unintended emotions scored (0−6)", how many unintended emotions were given a score of greater than 0; and 3, "Total score (0−9) given to unintended emotions", the sum of scores given to emotions that did not correspond to the true emotion.

Results

Validity

The validity measure (proportion interpreted correctly) was performed for every image. The data for these 424 individual images are presented separately on the Internet database. The proportion correctly interpreted for each portrayed emotion is, however, shown in Table 1. The overall value for the aggregated sum of results was considered high (mean=88%). Five out of the seven expressions had a mean percentage of correct interpretation of over 90%, while the emotions of fear and sadness had a mean percentage of 73% and 78% respectively. Ratings for emotions other than that intended are considered to be low (mean 0.13–0.65). There was a difference between the emotions in ratings given to emotions other than those intended (mean range from 0.38 points for happiness to 3.15 points for fear).

As shown in Table 2, levels of incorrect perception of expressed emotions were, with a few exceptions (eg, the intended facial expression of fear being perceived as surprise), consistently low.

Figure 1. A screen shot of the web-based validation.

The person seems to be

0: Disagree completely, 9: Completely agree

Table 1. Summary of the proportion of images correctly perceived, number of unintended emotions scored and the total scores given to unintended emotions.

Emotion expressed	Number of images	Proportion correctly perceived (%)			Number of unintended emotions scored (0–6)			Total score given to unintended emotions (0–9)		
		Mean[a]	Min[b]	Max[c]	Mean[a]	Min[b]	Max[c]	Mean[a]	Min[b]	Max[c]
Anger (n=9581)	61	94	72	100	0.25	0.07	0.73	0.87	0.12	2.57
Surprise (n=9357)	60	94	76	99	0.33	0.14	0.66	1.25	0.42	3.26
Happiness (n=9721)	62	98	85	100	0.13	0.05	0.44	0.38	0.08	1.65
Sadness (n=9393)	61	78	25	98	0.54	0.14	1.23	2.41	0.45	6.55
Neutral (n=9406)	60	91	56	99	0.38	0.14	0.94	1.21	0.36	4.21
Fear (n=9211)	60	73	39	95	0.65	0.33	1.08	3.15	1.43	6.32
Disgust (n=9325)	60	90	60	100	0.36	0.10	0.86	1.42	0.22	4.23
Total (n=65994)	424	88	25	100	0.38	0.05	1.23	1.52	0.08	6.55

Note: An image was considered to be correctly classified if the highest score was given to the emotion corresponding to the true emotion.

[a] Mean proportion of correct perception (n=9211–9721).

[b] The value for the image with the lowest proportion of correct perception.

[c] The value for the image with the highest proportion of correct perception.

Table 2. Confusion matrix for images of expressed emotion and rater response (only scores 7–9 shown).

Expressed emotion	Rater response (7–9) (%)						
	Anger	Surprise	Happiness	Sadness	Neutral	Fear	Disgust
Anger	74.6[a]	0.4	0.2	0.9	0.5	0.8	0.9
Surprise	0.2	81.7[a]	1.0	0.2	0.5	3.7	0.3
Happiness	0.2	0.3	92.5[a]	0.3	0.5	0.2	0.2
Sadness	1.1	1.0	0.4	55.6[a]	5.9	3.9	2.3
Neutral	1.0	0.8	0.4	1.7	81.6[a]	0.6	0.1
Fear	2.6	14.2	0.6	0.9	0.5	55.5[a]	1.7
Disgust	2.2	0.9	0.3	2.5	0.2	0.8	71.8[a]

[a] Intended emotion.

Odds Ratio

The odds ratios (presented in the supplemental tables in Appendix 1), ie, the relation between ratings 9 and 0, are high. The highest odds ratio was found in the expression of happiness (OR=1945.6, *P*<.0001), and the lowest was in fear (OR=72.0, *P*=<.0001).

The most noteworthy results relating to the four background variables (model age, model gender, rater age, and rater gender), presented in the supplemental tables in Appendix 1, were in the expressions of surprise and anger. Female facial models aged ≥46 years (OR=0.4, *P*<.05) and 26–45 years (OR=0.8, *P*<.05) were significantly less strongly associated with the expression anger in comparison with the reference group, but at the same time, significantly more strongly associated with the expression of surprise (OR=1.8, and OR=1.1, *P*<.05).

Female facial models were more frequently significantly associated with three of the intended expressions in comparison with male facial models. Those were the expressions anger (OR=1.2, *P*<.05), surprise (OR=1.2, *P*<.05) and neutral (OR=1.9, *P*<.05). There were no statistically significant differences in the expression of happiness, fear, and disgust. The expression of sadness was more frequently associated with male models than female models (OR=0.6, *P*<.05).

Discussion

Principal Results

The purpose of this study was to present a database of facial expressions and the results of an Internet-based validation study. The database contains 424 color images of models across a spectrum of age, ethnicity, and gender expressing a variety of different emotions. The database is freely available for scientific experiments both online and offline.

The validity of the database was based on how accurate the raters were in identifying the expressions in the presented images. Scores were generally high. The overall mean proportion of this database that was correctly interpreted was 88%. The corresponding values are 79% for NimStim [1] and 88% for Pictures of Facial Affect [25], with the Karolinska Directed Emotional Faces achieving a mean biased hit rate of 72% [2].

The results did not show any consistent advantage related to age or gender in either the models or in the validating participants. There were significant differences when the seven expressions were studied individually, but the stronger and weaker association varies across the four background variables. Hall and Matsumoto found that women made more correct interpretations than men when multiple scales were used [26]. These results were not replicated in our study, with the exception that women were better than men at identifying the neutral emotion and worse than men at recognizing disgust.

The results of previous studies [27] of the effect of age and gender have suggested that even higher identification scores are obtained if images are validated exclusively by women aged 20-30 years. However, our data did not show any consistently significant differences between gender groups or age groups, and previous results were not replicated in this study population. The only significant differences were that raters aged ≥46 years were better at identifying disgust and worse at identifying anger and that raters aged 26-45 were worse at identifying happiness than the reference group. That lower capacity of older raters to correctly identify anger is consistent with the results of Ebner et al [9].

The results of our study show that facial expressions of people ≥46 years showing anger, fear, and sadness were less reliably identified than those posed by younger faces aged ≤25 years. Faces of participants aged 26-45 years portraying anger, neutral, and disgust were also less reliably identified than the same expressions in younger people aged ≤25 years. This is consistent with the findings of Ebner et al [9] who showed that angry, disgusted, happy, neutral, and sad expressions were less accurately identified in older faces than in the faces of the young or middle-aged. Disgusted, neutral, and sad middle-aged faces were less accurately identified than young faces portraying the same emotions.

The validation study was Internet-based. A large number of participants from different age groups evaluated the images, which provides this study with a more heterogeneous population of raters than previous studies [2,7,8,12]. The number of ratings for each image was higher than in previous studies [1,2,8]. While this type of validation has merits, we had no control over how the raters were complying with the task or if they instead carried out the validation in a detrimental way.

As response scales with fixed response options can be problematic, Russell [22] recommended more studies with quantitative ratings on multiple scales. These are considered more neutral as they are not biased towards a single expression. Therefore, we chose a ten-degree Likert-type scale ranging from “completely disagree” to “completely agree”. Participants could choose to agree on one or more of the response scales. Giving the participants the opportunity to rate each image on a continuum and to choose to rate several emotions for each image provided important information about each image, enabling an assessment of how reliably it was depicting the intended emotion as compared to other emotions.

Some facial expressions of emotion are easier to identify correctly than others. In validation studies, happy facial expressions are usually recognized more reliably than negative facial expressions [1,2,25,28,29]. In this validation study, happy was the facial expression that had the highest proportion of correct identification and the lowest association with other emotions. These results are consistent with previous findings that happiness is the facial expression that is the most reliably identified and the least likely to be confused with other facial expressions [12].

Sadness and fear had the lowest proportion of correct identification, also consistent with previous research [2,12]. These expressions were also confused with other expressions to a greater extent. In particular, raters often thought that fear had an element of surprise. This confusion may be due to the similarities in these two facial expressions, with the eyes being wide open in both. It may be difficult for untrained models to make the facial movements necessary to distinguish these expressions. There may also be a measure of confusion on the basis of interpretation, because when fear is experienced, it is often preceded by surprise [24].

The method of creating facial expressions can affect their interpretation. Currently existing databases have been produced by instructing the photo shoot models in two different ways. One is to instruct the models to move particular muscle groups while making the facial expressions [8,25], and the other is to instruct them to make the emotional expressions as they see fit [1].

One advantage of asking models to move particular muscle groups is that it creates uniform expressions. The disadvantage is that the ecological validity may be affected [22]. Naturally produced facial expressions can be perceived as more authentic, but the variation within the same expression may be greater [22], and this could be regarded as a background variable in scientific experiments [1]. When models follow instructions about which muscles to move or imitate a picture, a larger proportion of expressions are correctly identified compared to studies in which the models made the expressions as they saw fit or spontaneous expressions were induced [10].

As we wanted models to make authentic expressions and still maintain uniformity within the same emotional expressions, the instructions given to models were a combination of the instructions used in previous studies. The models in the Umeå University Database of Facial Expressions were instructed to make the expressions as they saw fit, to look at pictures of facial expressions, and to move certain muscle groups.

Limitations

The database has, however, a number of shortcomings. First, as the validation study was Internet-based, it was difficult to control for the authenticity of participant responses and other contextual variables, eg, how closely participants followed the instructions. However, the requirement for personal information such as name, age, gender, and email address should have decreased the risk of non-valid answers. In addition, the relatively large number of participants (n=526) would have reduced the impact of deliberately false responses. The lack of remuneration also meant there was no financial reward in providing false responses.

Second, models may have validated their own images, which may have inflated the proportion of correct identification in the database. However, the number of models who may have validated their own images was small in relation to the large number of ratings made for each image.

Third, there may have been a subjective interpretation of the meaning of the response scales. The scale steps between 0 and 9 could have been interpreted as a measure of intensity, authenticity, or purity. However, giving the participants the opportunity to rate every image on a continuum and to rate for several expressions, provided important information about each image. Valuable information about the extent to which each image was rated for expressions other than the one intended is available online, as well as the proportion correctly identified for each image.

A fourth limitation is that a forced-choice scale was used to calculate the proportion of correct identification. The response scale that received the highest score was regarded as the respondent's answer. And since there was no "none of the above" option included, this has probably resulted in a higher proportion of correct identification than if this option had been included.

A fifth weakness is that no member of the research team instructing the models during the photo shoots, and selecting images for validation was certified according to the Facial Action Coding System (FACS) [30]. FACS is a guide to the categorization of facial movements according to the muscles used in producing them. But this would not have been entirely satisfactory either, as using FACS to create images has resulted in American "dialects" of facial expressions [10]. Not using FACS could be regarded as an advantage when the images are to be used with untrained participants.

Finally, the instruction not to wear make-up was not followed by all participants, which may bias the interpretation of the images. However, the resulting images may more closely resemble the facial expressions seen in real life.

Conclusion

The goal of creating the Umeå University Database of Facial Expressions was to provide the scientific community with an online database for scientific experiments. The database consists of a large and contemporary set of images showing models across a spectrum of age, ethnicity, and gender. The Internet-based validity study obtained a larger number of ratings for each image compared to previous validation studies, and it has a higher proportion of correct identification compared to many existing databases. However, the validity of the Umeå University Database of Facial Expressions needs to be tested by further validation studies of similar or different design. Finally, we invite the scientific community to help expand the database by allowing inclusion of additional models to provide a more representative sample of populations. Obviously any added faces would first need to be validated to ensure high standards.

Acknowledgments

The study was funded by a grant from the Swedish Council for Social Research and the Swedish Council for Work Life Research (2009-0222). The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript. Alexander Alasjö is thanked for excellent web programming and Hans Pettersson for statistical help.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplemental Tables 1-7. Factors associated with images.

[PDF File (Adobe PDF File), 622KB - jmir_v14i5e136_app1.pdf]

References

1. Tottenham N, Tanaka JW, Leon AC, McCarry T, Nurse M, Hare TA, et al. The NimStim set of facial expressions: judgments from untrained research participants. Psychiatry Res 2009 Aug 15;168(3):242-249. [doi: 10.1016/j.psychres.2008.05.006] [Medline: 19564050]
2. Goeleven E, De Raedt R, Leyman L, Verschuere B. The Karolinska Directed Emotional Faces: A validation study. Cognition & Emotion 2008;22(6):1094-1118. [doi: 10.1080/02699930701626582]
3. Hakamata Y, Lissek S, Bar-Haim Y, Britton JC, Fox NA, Leibenluft E, et al. Attention bias modification treatment: a meta-analysis toward the establishment of novel treatment for anxiety. Biol Psychiatry 2010 Dec 1;68(11):982-990. [doi: 10.1016/j.biopsych.2010.07.021] [Medline: 20887977]
4. Nathanson DL. Shame and pride: affect, sex, and the birth of the self. New York: Norton; 1992.
5. Dimberg U. Facial electromyography and the experience of emotion. Journal of Psychophysiology 1988;2(4):277-282.

6. Ekman P, Friesen VW. Unmasking the face. A guide to recognizing emotions from facial expressions. Cambridge, MA: Malor Books; 2003.
7. Tracy JL, Robins RW, Schriber RA. Development of a FACS-verified set of basic and self-conscious emotion expressions. Emotion 2009 Aug;9(4):554-559. [doi: 10.1037/a0015766] [Medline: 19653779]
8. Langner O, Dotsch R, Bijlstra G, Wigboldus DHJ, Hawk ST, van Knippenberg A. Presentation and validation of the Radboud Faces Database. Cognition & Emotion 2010;24(8):1377-1388. [doi: 10.1080/02699930903485076]
9. Ebner NC, Riediger M, Lindenberger U. FACES--a database of facial expressions in young, middle-aged, and older women and men: development and validation. Behav Res Methods 2010 Feb;42(1):351-362. [doi: 10.3758/BRM.42.1.351] [Medline: 20160315]
10. Elfenbein HA, Ambady N. On the universality and cultural specificity of emotion recognition: a meta-analysis. Psychol Bull 2002 Mar;128(2):203-235. [Medline: 11931516]
11. Hugenberg K, Bodenhausen GV. Facing Prejudice. Psychological Science 2003;14(6):640-643. [doi: 10.1046/j.0956-7976.2003.psci_1478.x]
12. Biehl M, Matsumoto D, Ekman P. Matsumoto and Ekman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability Data and Cross-National Differences. Journal of Nonverbal Behavior 1997;21(1):3-21. [doi: 10.1023/a:1024902500935]
13. Beaupré MG, Hess U. Cross-Cultural Emotion Recognition among Canadian Ethnic Groups. Journal of Cross-Cultural Psychology 2005;36(3):355-370.
14. Ekman P. Are there basic emotions? Psychol Rev 1992 Jul;99(3):550-553. [Medline: 1344638]
15. Herba C, Phillips M. Annotation: Development of facial expression recognition from childhood to adolescence: behavioural and neurological perspectives. J Child Psychol Psychiatry 2004 Oct;45(7):1185-1198. [doi: 10.1111/j.1469-7610.2004.00316.x] [Medline: 15335339]
16. McClure EB. A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. Psychol Bull 2000 May;126(3):424-453. [Medline: 10825784]
17. Plant EA, Kling KC, Smith GL. The Influence of Gender and Social Role on the Interpretation of Facial Expressions. Sex Roles 2004;51(3):187-196. [doi: 10.1023/b:sers.0000037762.10349.13]
18. Vigil JM. A socio-relational framework of sex differences in the expression of emotion. Behav Brain Sci 2009 Oct;32(5):375-90; discussion 391. [doi: 10.1017/S0140525X09991075] [Medline: 19825246]
19. Williams LM, Mathersul D, Palmer DM, Gur RC, Gur RE, Gordon E. Explicit identification and implicit recognition of facial emotions: I. Age effects in males and females across 10 decades. J Clin Exp Neuropsychol 2009 Apr;31(3):257-277. [doi: 10.1080/13803390802255635] [Medline: 18720177]
20. Brosgole L, Weisman J. Mood recognition across the ages. Int J Neurosci 1995 Jun;82(3-4):169-189. [Medline: 7558648]
21. Ruffman T, Henry JD, Livingstone V, Phillips LH. A meta-analytic review of emotion recognition and aging: implications for neuropsychological models of aging. Neurosci Biobehav Rev 2008;32(4):863-881. [doi: 10.1016/j.neubiorev.2008.01.001] [Medline: 18276008]
22. Russell JA. Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. Psychol Bull 1994 Jan;115(1):102-141. [Medline: 8202574]
23. Barrett LF, Mesquita B, Gendron M. Context in Emotion Perception. Current Directions in Psychological Science 2011;20(5):286-290.
24. Ekman P. Emotions revealed: Recognizing faces and feelings to improve communication and emotional life. New York: St Martin's Griffin; 2007.
25. Ekman P, Friesen WV. Pictures of facial affect. In: Slides. Palo Alto, CA: Consulting Psychologists Press; 1976.
26. Hall JA, Matsumoto D. Gender differences in judgments of multiple emotions from facial expressions. Emotion 2004 Jun;4(2):201-206. [doi: 10.1037/1528-3542.4.2.201] [Medline: 15222856]
27. Riediger M, Voelkle MC, Ebner NC, Lindenberger U. Beyond "happy, angry, or sad?": age-of-poser and age-of-rater effects on multi-dimensional emotion perception. Cogn Emot 2011 Sep;25(6):968-982. [doi: 10.1080/02699931.2010.540812] [Medline: 21432636]
28. Elfenbein HA, Ambady N. When familiarity breeds accuracy: cultural exposure and facial emotion recognition. J Pers Soc Psychol 2003 Aug;85(2):276-290. [Medline: 12916570]
29. Gur RC, Sara R, Hagendoorn M, Marom O, Hughett P, Macy L, et al. A method for obtaining 3-dimensional facial expressions and its standardization for use in neurocognitive studies. J Neurosci Methods 2002 Apr 15;115(2):137-143. [Medline: 11992665]
30. Ekman P, Friesen WV, Hager JC. Facial action coding system - The manual on cd-rom. Salt Lake City, UT: Research Nexus Division of Network Information Research Corporation; 2002.

Edited by G Eysenbach; submitted 01.06.12; peer-reviewed by S Blackwell, T Berger, P Enock; comments to author 19.06.12; revised version received 20.06.12; accepted 27.08.12; published 09.10.12.

Please cite as:
Samuelsson H, Jarnvik K, Henningsson H, Andersson J, Carlbring P
The Umeå University Database of Facial Expressions: A Validation Study
J Med Internet Res 2012;14(5):e136
URL: http://www.jmir.org/2012/5/e136/
doi:10.2196/jmir.2196
PMID:23047935

Original Paper

A Panel Analysis of the Strategic Association Between Information and Communication Technology and Public Health Delivery

Sarah Jinhui Wu[1], PhD; Wullianallur Raghupathi[2], PhD

[1]Management Systems, The Gabelli School of Business, Fordham University, Bronx, NY, United States

[2]Graduate School of Business, Information and Communication Systems, Fordham University, New York, NY, United States

Corresponding Author:
Wullianallur Raghupathi, PhD
Graduate School of Business
Information and Communication Systems
Fordham University
113 W. 60th Street
New York, NY, 10023
United States
Phone: 1 212 636 7230
Fax: 1 212 765 5573
Email: wraghupathi@gmail.com

Abstract

Background: In this exploratory research, we use panel data analysis to examine the correlation between Information and Communication Technology (ICTs) and public health delivery at the country level.

Objective: The goal of this exploratory research is to examine the strategic association over time between ICTs and country-level public health.

Methods: Using data from the World Development Indicators, we construct a panel data set of countries of five different income levels and look closely at the period from 2000 to 2008. The panel data analysis allows us to explore this dynamic relationship under the control for unobserved country-specific effects by using a fixed-effects estimation method. In particular,, we examine the association of five ICT factors with five public health indicators: adolescent fertility rate, child immunization coverage, tuberculosis case detected, life expectancy, and adult mortality rate.

Results: First, overall ICTs' factors substantially improve a country's public health delivery on the top of wealth effect. Second, among all the ICTs' factors, accessibility is the only one that is associated with improvements in all aspects of public health delivery, while the contributions from the usage, quality, and applications are negligible. ICTs' accessibility factor is associated with a considerable extension to life expectancy and reduced adult mortality rate. Third, all entity-specific factors are significant in each model, indicating that countries' economic development level does influence their public health delivery.

Conclusions: Our results indicate that ICT accessibility has a strong association with effective delivery of public health. There are others, but the key strategic applications are eHealth and mHealth. The findings of this study will help government officials and public health policy makers to formulate strategic decisions regarding the best ICT investments and deployment. For example, the study shows that providing accessibility should be a critical focus.

(J Med Internet Res 2012;14(5):e147) doi:10.2196/jmir.2242

KEYWORDS

Association; Correlation; Information and communication technologies (ICTs); Panel data; Public health delivery

Introduction

Despite the slow diffusion of information and communication technology (ICT) to support public health delivery worldwide as well as the high cost of implementation, global health policy makers and public health officials must focus on key strategic applications of ICTs that deliver high-quality public health care at lower costs. ICTs can be characterized by such factors as access, quality, and applications among others. Specifically, the applications may include such categories as eHealth, mHealth, situation awareness systems, and "smart" health systems [1-13] among others. eHealth refers to the use of ICTs such as computers, mobile phones, and satellite and wireless communication systems for health information and services [1,6,12,14]. mHealth indicates the use of mobile communications such as personal digital assistants (PDAs) and mobile phones to provide remote health information and services [15]. Situation awareness systems enable us to prepare for and respond to disease outbreaks and other public health emergencies by automating the surveillance processes and the collection, analysis, and reporting of the health data [5]. Smart health systems refer to the use of ICTs for better connection, higher-quality data, and real-time analysis of the data. Doctors, patients, insurers, and government officials can share health information seamlessly and efficiently.

Understanding the association between ICTs and the nature of public health delivery is important [16,17] since ICTs have the potential to accelerate the improvement in public health. ICTs provide the infrastructure and resources for the development of large-scale population-level applications such as health information networks, surveillance systems, and telemedicine. ICTs can improve health outcomes and combat diseases. For example, in rural Niger, following the introduction of a radio ambulance system, the number of emergency evacuations from outlying health centers to the district hospital rose from 10 to 197. Good communications and information sharing help to deliver diagnostic information and drugs and to spread information on reproductive health and communicable diseases. Through the Global Media acquired immunodeficiency syndrome (AIDS) initiative, more than 50 broadcast networks are promoting AIDS prevention messages. In the fight against malaria, satellite monitoring identifies and targets mosquito breeding areas for control [18]. Other examples of ICT-based systems at work (reported in [6]) include telemedicine networks in Bangladesh; e-pharmacy projects in Malaysia; low-cost sustainable electronic medical records for human immunodeficiency virus infection (HIV)/AIDS patients in Kenya; and web-based communication tools to address material and child deaths in Peru [6].

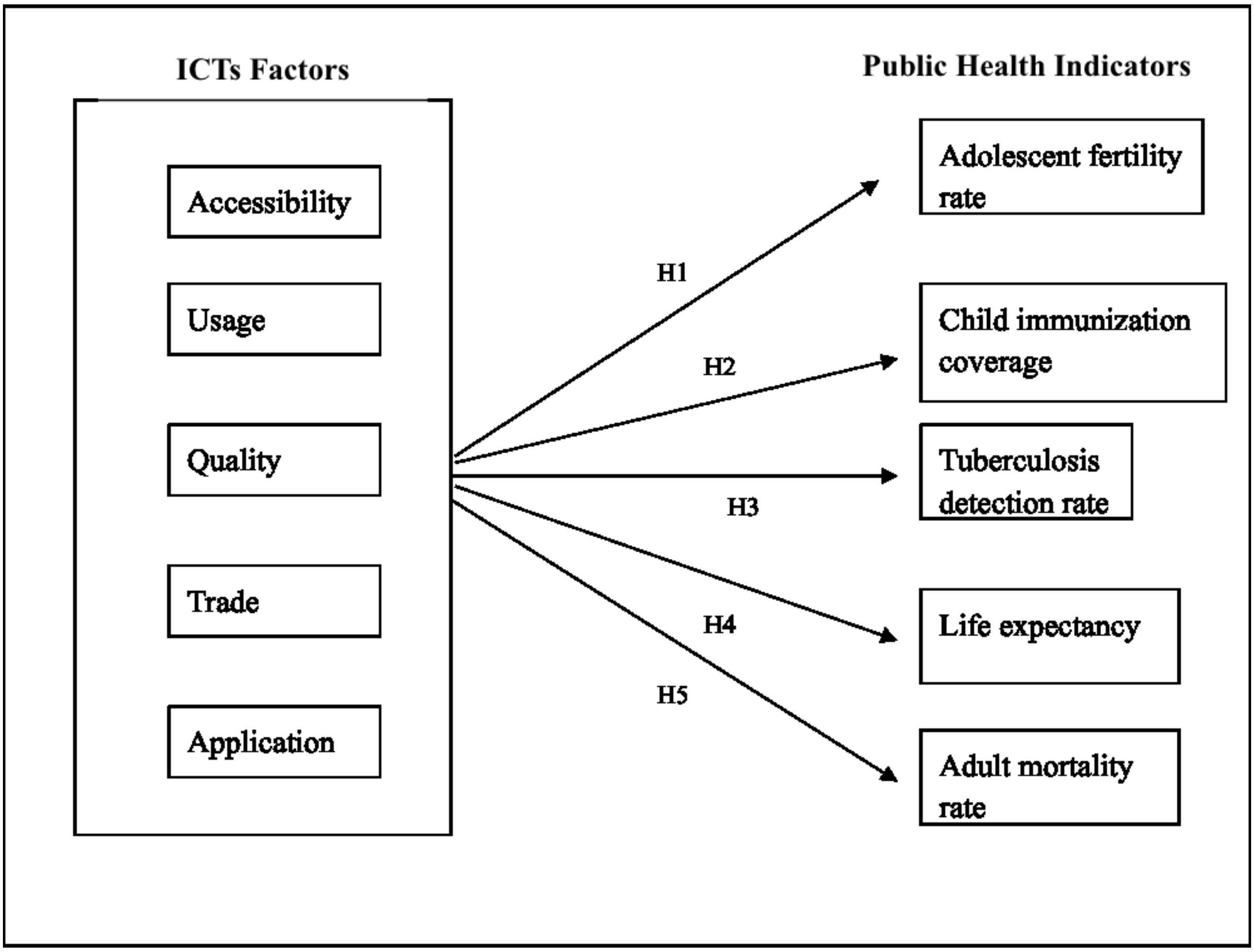
In addition, public health informatics (PHI) was defined as the “systematic application of information and computer science and technology to public health practice, research and learning that integrates public health and information technology” [19]. PHI needs to use a “systematic and informed approach to the application of information science and technology in order to take full advantage of its potential to enhance and facilitate public health activities” [20].

Given the potential linkage between ICTs and public health, the purpose of this paper is to empirically evaluate the two related research questions: (1) Do ICTs have a strong association with a country’s public health delivery indicators over time?; (2) Which of the key ICT factors are associated with improvement in public health delivery?

The framework in this paper is an adaptation of the conceptual framework developed in an earlier paper (see [17] for comprehensive literature review) as well as the publications at the International Telecommunication Union (ITU) [8-10], World Health Organization (WHO) [13,18], and World Bank [21]. We also draw on the substantive health information technology literature [17] and the recent special issues of Journal of the Association for Information Systems (JAIS) [22] and Journal of Strategic Information Systems (JSIS) [23]. Both anecdotal evidence and various surveys indicate a potential association between ICTs and country-level public health delivery as previously discussed. As such, we propose a positive association between ICTs (independent variables) and public health variables (dependent variables) (Figure 1). The ICT factors are derived from the World Bank ICT at-a-Glance Tables [24] that include accessibility, quality, usage, affordability, trade, and application. We chose for our macro-level study the significant aggregate indicators related to public health [25] that contain adolescent fertility rates, child immunization coverage, tuberculosis detection rates, life expectancy, and adult mortality rates.

While proposing these associations we recognize the reality that high-income countries would have better ICT infrastructure and better public health. That is, ICTs may not have an association with global public health delivery; wealth may provide the association. To obviate this possibility, the wealth effect is controlled so as to obtain accurate and reliable estimates. We discuss next the public health indicators in context of the hypotheses.

Figure 1. Conceptual framework.



Public Health Indicators

Adolescent Fertility Rate

A high number for this indicates that fewer children are going to school; additionally, early childbirth has its own complications with less access to post-childbirth health care. ICTs, in aggregate can help in educating this population to take preventive steps, assist in delivering health information via eHealth and mHealth, and generally assist in reducing the rate. Therefore, we formulate hypothesis 1 as: H1. Information and communication technology (ICT) is associated with a reduction in the adolescent fertility rate over time.

Child Immunization Rate

A high number for this variable would indicate better quality preventive health care and thereby improved public health. The different ICT factors can influence the child immunization rate in positive ways such as enabling the communication and dissemination of immunization information, educating the public about the benefits of immunization in reducing instances of preventable diseases, and facilitating the delivery of immunization by health workers via setting up a database of records, tracking, and generating reminders for follow-up. So we formulate hypothesis 2 as: H2. Information and communication technology (ICT) is associated with an increase in child immunization coverage over time.

Tuberculosis Detection Rate

This represents other diseases such as malaria, HIV/AIDs, and other in terms of detection, monitoring, and prevention. ICTs can help in several ways to increase the detection rate. For one, advanced ICTs can help in the analysis of the blood work and the intakes from the tests and procedures. The collected information can be organized and analyzed to gain insight to make informed decisions regarding the disease status. Additionally, ICTs can help communicate the results and follow-up guidelines to the patients (eg, via texting or mobile phone calls) for prompt action. Therefore, hypothesis 3 is formulated as: H3. Information and communication technology (ICT) is associated with an increase in the tuberculosis case detection rate over time.

Life Expectancy

A key goal of development as well as health is to increase the life expectancy of a country's citizens so human beings can achieve their full potential. The different ICT factors can contribute overall to the achievement of this goal via providing health-related information to the public, for example on hygiene, sanitation, and other good practices on living. Additionally, health workers can be provided with ICT-enabled accessories to educate the public; identify, gather, and document information about health status; and to perform follow-up activities. Educating the public is necessary to improved public health, and ICTs can play a critical role in this regard. We formulate hypothesis 4 as follows: H4. Information and communication technology (ICT) is associated with an increase in the life expectancy over time.

Adult Mortality Rate

This again represents the overall human condition status of a country. The goal is to reduce the adult mortality rate. This is possible by focusing on good public health and preventive measures. ICTs can play a significant role in many ways including education for preventive care, delivery of health care via telemedicine, etc, and the monitoring of health conditions such as spread of diseases, pandemics, etc, via the use of mHealth, Internet, and radio and television. By reducing the adult mortality rate, a country can take advantage of a healthy adult population to engage in productive work. Therefore, hypothesis 5 is formulated as such: H5. Information and communication technology (ICT) is associated with a reduction in the adult mortality rate over time.

Method

In a prior cross-sectional study of 200 countries using secondary data from World Bank, it was found that in most cases cumulated investment in ICTs is strongly associated with a country's public health delivery [17]. Yet, the nature of the dataset limited the ability to make strong inferences between the two even though the wealth effect was controlled in the analysis. To estimate the extent of the association between ICTs and improved country public health delivery, this study looks at a panel data set of countries representing five different income levels from 2000 to 2008. It has been argued that this association also leads to economic growth and poverty reduction. The panel data set allows us to explore these dynamic relationships while controlling for unobserved country-specific effects. We use a fixed-effects estimation method to control for unobserved country factors within an income level that may be correlated with countries' public health delivery. We used this method because any unobserved country factor is likely to bias estimated coefficients in a traditional regression model. This method also allows us to examine the wealth effect of different income levels on countries' public health delivery. Overall, this panel analysis delineates the impacts of income level and ICT factors on countries' public health delivery, thus providing more accurate evaluation of the contribution of ICTs to public health delivery.

Data Collection and Measurements

The data used to test our hypotheses came from two World Bank databases for more than 200 countries during 2000 to 2008. The World Bank ICT at-a-Glance database [24] classifies the major ICT sector performance variables into six factors, namely accessibility, usage, quality, affordability, trade, and applications. The detailed measurement variables are summarized in Table 1. Some measurement variables suffered from a high proportion of missing values and were not included in the later analysis. Unfortunately, all three measurement variables for the affordability factor had high missing values, and this factor was not included in the later analysis. Since the measurement variables were in different scales, they were first standardized with a mean of zero and standard deviation of one and then averaged to extract the corresponding factor scores.

Table 1. Measurement items for ICT sector performance.

ICT factors	Measurement variables
Accessibility	
	Telephone lines (per 100 people)
	Mobile cellular subscriptions (per 100 people)
	Fixed Internet subscribers (per 100 people)
	Personal computers (per 100 people)
	[a]Households with a television set (%)
Usage	
	[a]International voice traffic (minutes/person/month)
	Mobile telephone usage (minutes/user/month)
	Internet users (per 100 people)
Quality	
	[a]Population covered by mobile cellular network (%)
	Fixed broadband subscribers (% of total Internet subscribers)
	International Internet bandwidth (bits/second/person)
Affordability	
	[a]Residential fixed line tariff (US$/month)
	[a]Mobile cellular prepaid tariff (US$/month)
	[a]Fixed broadband Internet access tariff (US$/month)
Trade	
	ICT goods exports (% of total goods exports)
	ICT goods imports (% of total goods imports)
	ICT service exports (% of total service exports)
Applications	
	ICT expenditure (% of GDP)
	[a]E-government web measure index
	Secure Internet servers (per 1 million people)

[a] Measurement items are not included in the later analysis due to high missing values.

Data for the public health indicators were extracted from the World Development Indicators database [25], which includes adolescent fertility rate, child immunization, tuberculosis case detection rate, life expectancy, and adult mortality rate. The definitions of these indicators are summarized in Table 2. Child immunization consists of two variables: immunization against DPT and against measles. Likewise, adult mortality rate contains two variables: mortality rates for females and males. All the variables were normalized, and the values of child immunization and adult mortality rate were calculated as the averages of their two measurement variables respectively.

Table 2. Measurement items for public health indicators.

Variables	Measurement variables
Adolescent fertility rate	The number of births per 1000 women ages 15−19.
Child immunization	The percentage of children ages 12−23 months who received vaccination against diphtheria, pertussis, and tetanus (DPT) as well as measles.
Tuberculosis case detection rate	The ratio of newly notified tuberculosis cases (including relapses) to estimated incident cases (case detection, all forms).
Life expectancy	The number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life.
Adult mortality rate	The probability of dying between the ages of 15 and 60 (per 1000 female adults); that is, the probability of 15-year-old dying before reaching age 60, if subject to current age-specific mortality rates between those ages.

The dataset of ICT factors and public health indicators were extracted based on five income groups defined by World Bank (ie, high-income countries, upper high-income countries, middle-income countries, lower middle-income countries, and low-income countries) and then merged based on the income groups. The final dataset contained five country categories for nine years (from 2000−2008).

Analysis Method

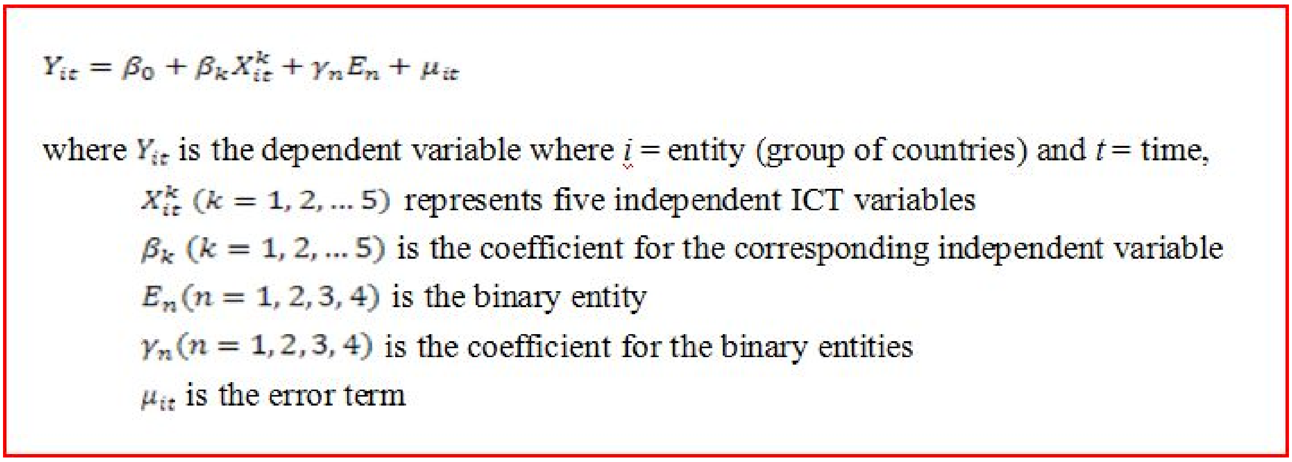
The primary analysis method in this study is panel data analysis, which is an increasingly popular form of longitudinal data analysis among social and behavioral science researchers. A panel is a cross-section or group of people who are surveyed periodically over time. With repeated observations of enough cross-sections, panel analysis permits the researcher to study the dynamics of change with short time series. The combination of time series with cross-sections can enhance the quality and quantity of data in ways that would be impossible using only one of these two dimensions. Specifically, the panel data allow us to control for variables that we cannot observe or measure in each group (eg, culture) and also help to control for unobservable variables that change over time but not across entities (eg, economic development level).

With panel data one can include variables at different levels of analysis. In this study, we could choose each country as the level of analysis, but we set the level of analysis at an aggregate level—groups of countries based on their income. The primary reason for this choice is that the missing values of all the interested variables across 200 countries result in a substantial reduction of the dataset. Following World Bank classification, all the countries are divided into five income groups: high-income group, upper middle-income group, middle-income group, lower middle-income group, and low-income group.

The following fixed effect model is set up to explore the relationship between ICT factors and public health delivery indicators within each group of countries, as we assume that something within a country group may impact or bias the predictor or outcome variables and that this needs to be controlled in the model. The key insight is that if the unobserved variables do not change over time, then any changes in the outcome variable must be caused by influences other than these fixed characteristics. As such, once the effect of those time-invariant characteristics from the predictor variables is removed, we can assess the predictors' net effect on outcome variables.

There are two ways to build the fixed effect model, and we chose to use the one with binary variables because we could separate the association of ICT factors and income levels from public health delivery indicators. Since we have five entities (five groups of countries) to generate binary (dummy) variables, only four entities are included in the model (see Figure 2).

Figure 2. Fixed effect model.



$$Y_{it} = \beta_0 + \beta_k X_{it}^k + \gamma_n E_n + \mu_{it}$$

where Y_{it} is the dependent variable where i = entity (group of countries) and t = time,

X_{it}^k $(k = 1, 2, \dots 5)$ represents five independent ICT variables

β_k $(k = 1, 2, \dots 5)$ is the coefficient for the corresponding independent variable

$E_n (n = 1, 2, 3, 4)$ is the binary entity

$\gamma_n (n = 1, 2, 3, 4)$ is the coefficient for the binary entities

μ_{it} is the error term

Results

The data analysis was conducted in STATA. The association between the ICT factors and the five public health indicators are summarized in Tables 3 and 4. Generally, we found support for all the hypotheses. That is, the application of ICTs overall could reduce adolescent fertility rates (H1) and adult mortality rates (H5), increases the coverage of child immunization (H2) and the chances of tuberculosis detection (H3), as well as lengthens life expectancy (H4).

Second, among all the ICT factors, accessibility is the only one that is associated with improvements in all aspects of public health delivery, while the contributions from the usage, quality, and applications are negligible. This is possibly due to lack of longitudinal data or the fact that some aspects of the application of ICTs in public health are only now emerging.

Third, all entity-specific factors are significant in each model, indicating that countries' economic development level does influence their public health delivery. To an extent, developed countries do have a base level of public health delivery. This is because of overall higher standard of living, better education, and less poverty. Figures 3 to 7 plot the ICT accessibility factor against each public health indicator for each group of countries. Overall, high-income countries enjoy higher ICT accessibility and better health conditions, whereas low-income countries have low accessibility and poor health conditions. Thus, wealth effect is confirmed.

Table 3. Associations between ICT factors and public health delivery indicators (adolescent fertility rate, child immunization, and tuberculosis detection).

Independent variables	Adolescent fertility rate		Child immunization		Tuberculosis detection	
	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value
Accessibility	-0.28	<.001	0.64	<.001	0.86	<.001
Usage	-0.03	.32	0.14	.16	0.20	.07
Quality	<0.001	.05	<-0.001	.21	<-0.001	.06
Trade	-0.15	.01	0.90	<.001	1.37	<.001
Application	-0.05	.15	-0.06	.61	0.04	.76
Lower middle-income countries	-1.33	<.001	-3.73	<.001	-5.13	<.001
Middle-income countries	-1.39	<.001	-2.67	<.001	-3.82	<.001
Upper middle-income countries	-1.18	<.001	-0.70	.09	-2.06	<.001
High-income countries	-1.98	<.001	-1.94	.005	-3.45	<.001
Constant	1.12	<.001	2.10	<.001	3.31	<.001
F test that all U_i=0	76.86	<.001	38.51	<.001	19.05	<.001

Table 4. Associations between ICT factors and public health delivery indicators (life expectancy, adult mortality rate).

Independent variables	Life expectancy		Adult mortality rate	
	Coefficient	*P* value	Coefficient	*P* value
Accessibility	0.30	.03	-0.35	<.001
Usage	-0.004	.89	-0.04	.30
Quality	<-0.001	.13	<0.001	.84
Trade	-0.07	.52	0.01	.88
Application	0.14	.15	0.05	.35
Lower middle-income countries	1.54	.01	-1.72	<.001
Middle-income countries	1.57	.002	-1.63	<.001
Upper middle-income countries	1.78	<.001	-1.28	<.001
High-income countries	2.31	.01	-2.40	<.001
Constant	-1.46	.003	1.38	<.001
F test that all U_i=0	113.91	<.001	38.36	<.001

Figure 3. Impact of ICT accessibility on adolescent fertility rate.

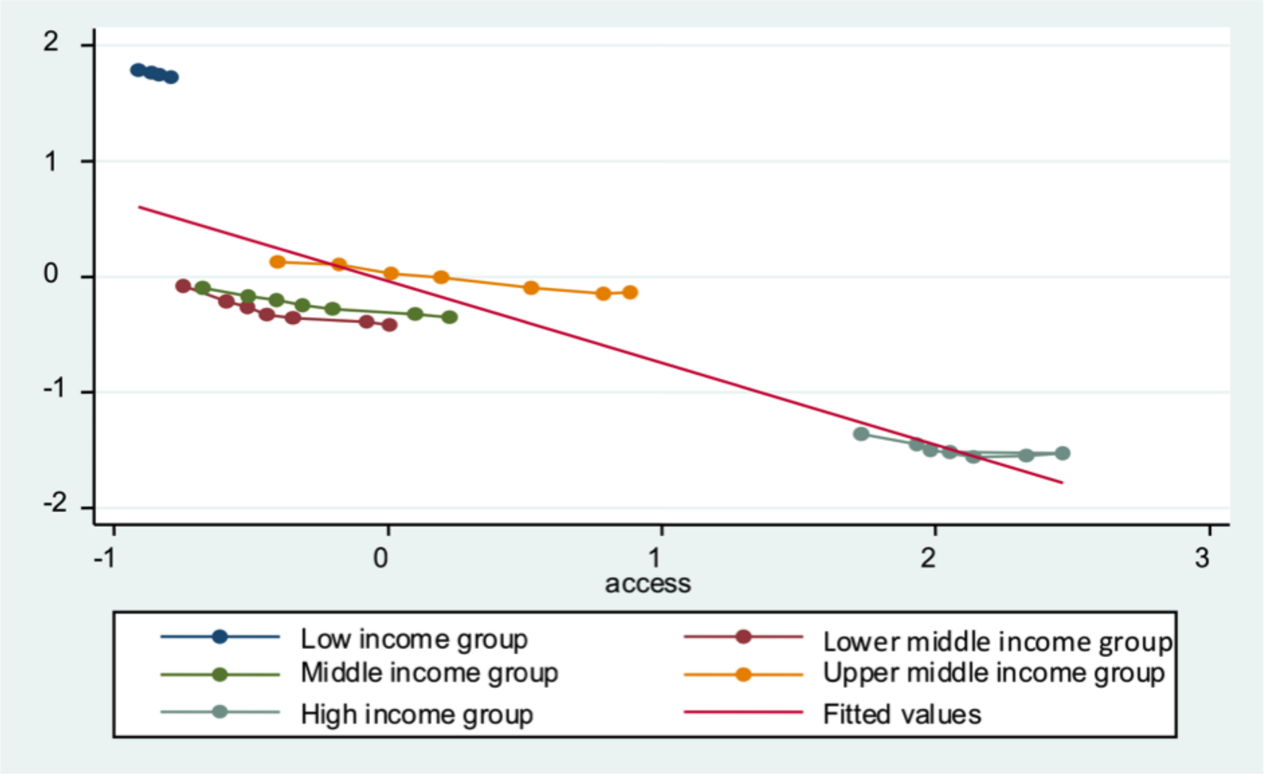


Figure 4. Impact of ICT accessibility on child immunization coverage.

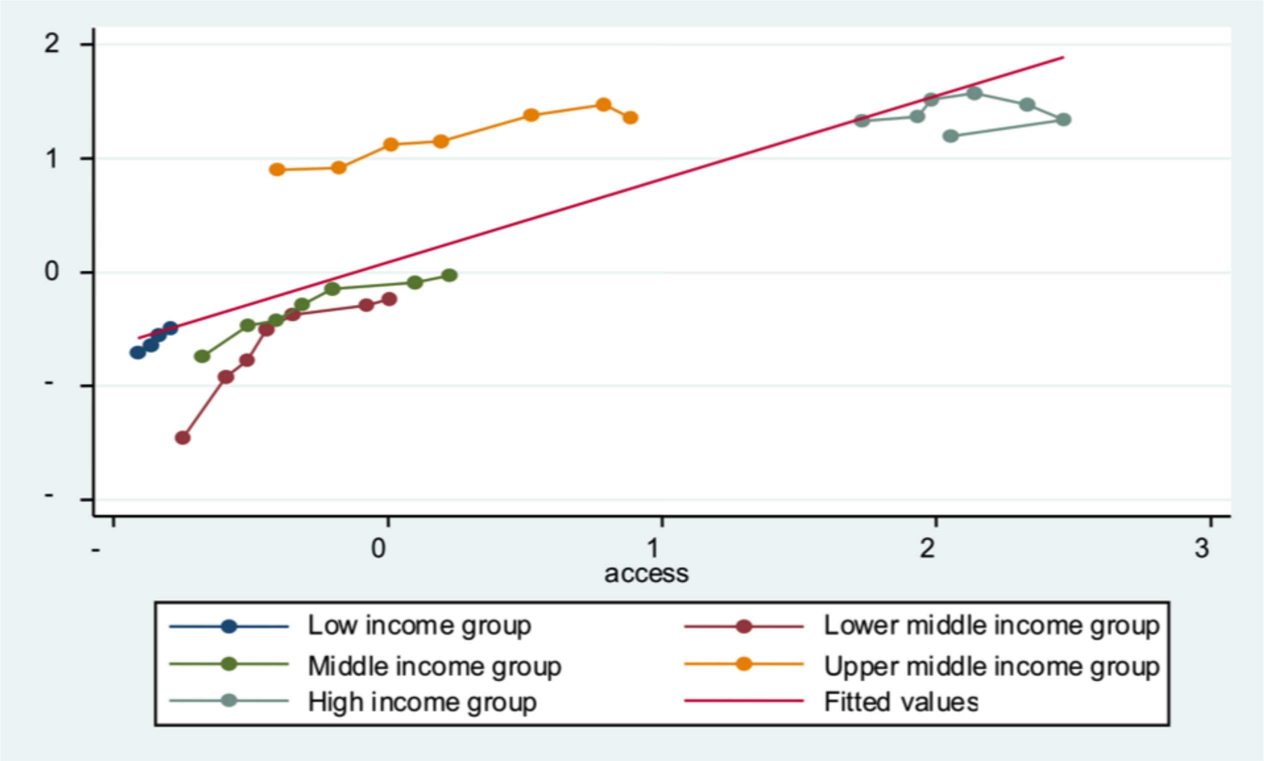


Figure 5. Impact of ICT accessibility on tuberculosis case detection rate.

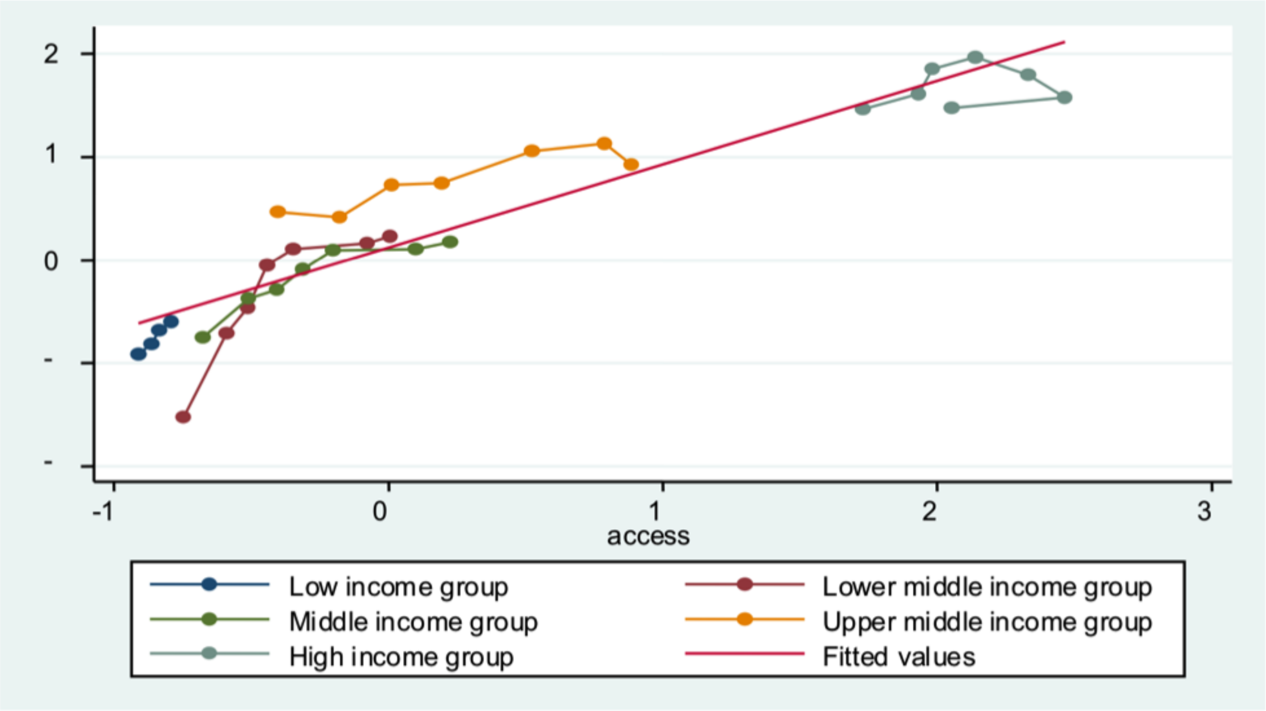


Figure 6. Impact of ICT accessibility on life expectancy.

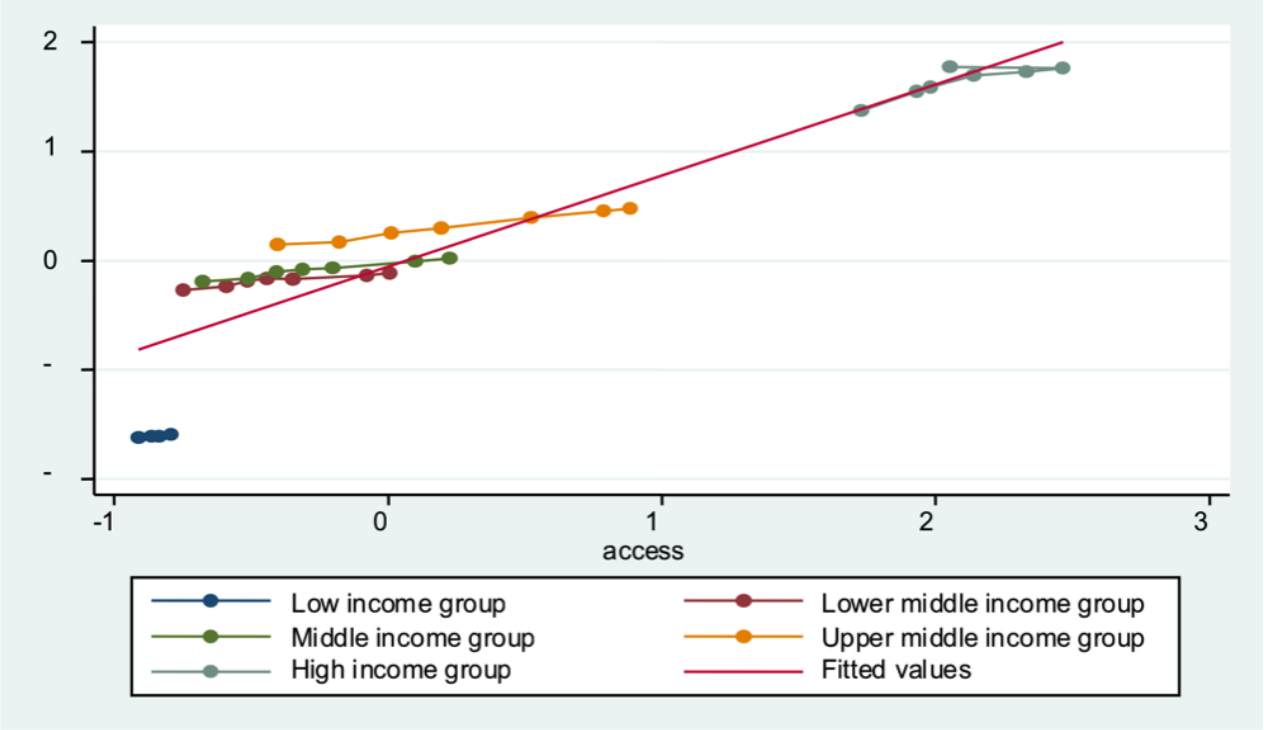
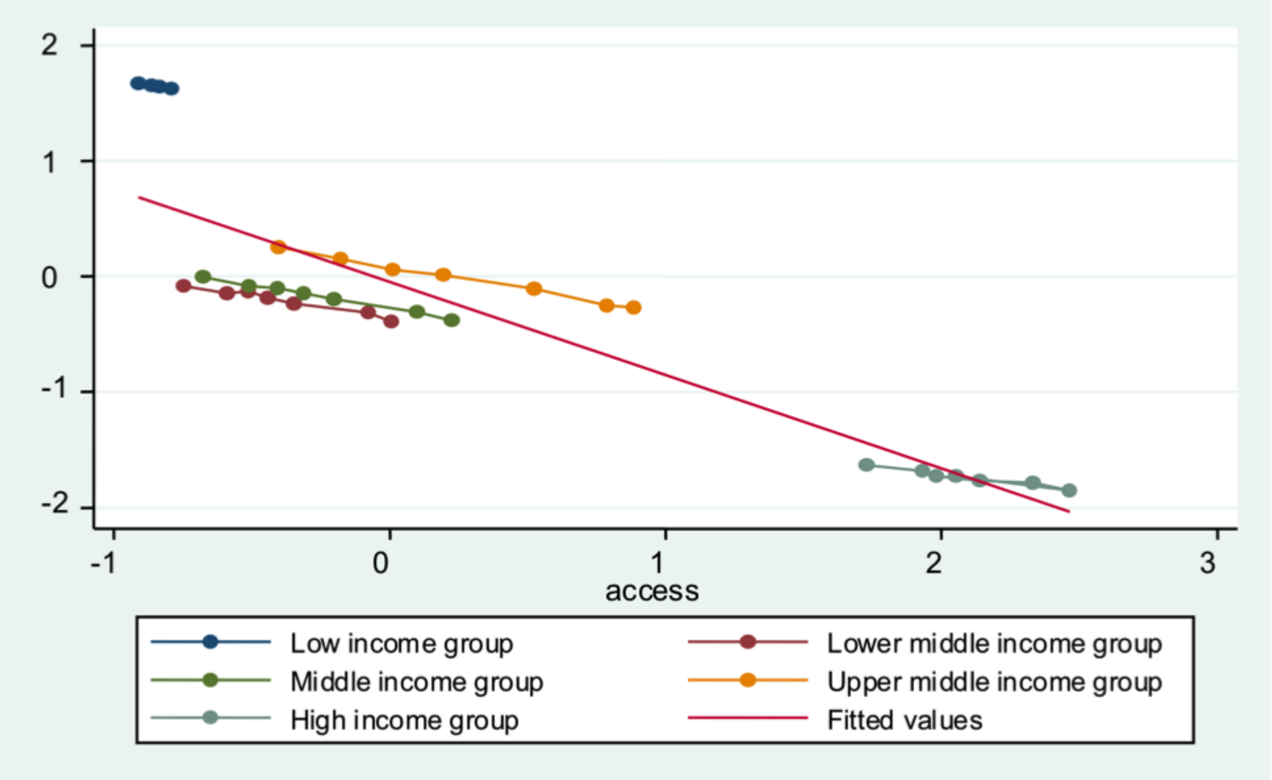


Figure 7. Impact of ICT accessibility on adult mortality rate.



Discussion

Key Results

First, our results indicate that overall ICT factors substantially improve a country's public health delivery on top of the wealth effect. As mentioned previously, ICTs are not vaccines or medications or health protocols, rather ICTs are technologies that enable the processing of health information more efficiently and effectively. This has implications for health policy, particularly to developing countries.

In developing countries, lack of access to health care and inefficient health care delivery methods are the norm. Poor health care is characterized by high infant, female (pregnancy & birth mortality), and adult mortalities, low immunization rates, death from diseases, and low life expectancies. The shortage of health care is worsened by the inability of governments to devote adequate funding to their respective health care sectors, and the shortage is compounded by the multiple crises in finance, food, and energy. Unsuccessful reform efforts, combined with little funding, have left billions without the ability to tap into basic health care services. The investment in ICTs is channelized for use in health care via education, remote delivery of health care, and monitoring and tracking, for example.

Second, not all the ICT factors generate similar effects on public health indicators—some produce significant effects while others contribute little. As such, the government of a country can focus on ICT investment to achieve its specific goal in public health. For instance, ICT accessibility factors are associated with a considerable extension to life expectancy and reduced adult mortality rate. This is possibly due to the use of ICTs in the key strategies of education and preventive care enabled by ICTs. These two variables are also linked to poverty alleviation and economic development. Additionally, the access and trade factors of ICTs result in reduced adolescent fertility rates. This can be rationalized by the observation that having access to ICTs and the economic development activity of ICT trade indicate a higher level of influence of education and awareness of health consequences. In others, ICTs make the population more sophisticated in the access to health information. Likewise, access and trade improve child immunization coverage. This again indicates the fact that access to mobile and other ICTs by both the general population and health workers leads to increased awareness of immunization benefits, vaccination schedules, and follow-up processes leading to an increase in coverage. However, only access factors appear to have an impact on tuberculosis case detection rates, life expectancy, and adult mortality rates. But access is a key ICT factor, and it is not critical that every ICT factor have some bearing on public health delivery. There are possible variations in the use of ICTs across geographic locations and cultures. Additionally, life expectancy and adult mortality rates have to be examined over a very long period of time.

Finally, the analysis results also suggest that in almost all areas studied, countries in the higher-income groups have progressed further in the adoption of actions and provision of services than those in the lower-income groups. This finding is not surprising; it confirms that the "digital divide" includes eHealth. Our common goal should be, therefore, to lessen this divide with concerted action [18]. First, by investing in ICT's targeting of public health, the delivery of public health is enhanced. Second, although the effect of wealth is acknowledged, the cumulative investment in ICTs will improve public health delivery regardless of the development stage of a country. Third, accessibility has the greatest association with positive delivery of public health vis-à-vis the other ICT factors. Therefore, enabling accessibility ought to be the top priority in ICT infrastructure investment, particularly in developing countries.

It is also important to note that the study does not show a direct causality between ICTs and public health delivery. In other words, the study does not say that by using ICTs, health improves. Rather, we conclude that there is a strong association in the relationship between certain ICTs and how public health is delivered. In this regard, the findings support the anecdotal and case studies evidence reported in the literature. Further longitudinal studies in the linkages between specific ICT applications (eg, introduction of a telemedicine project) and health indicators (eg, reduction in the number of hospital visits) may address causal relationships. But large "big data" sets are needed for this type of study.

Conclusion

The contributions of this study are several. For one, by further studying the association between ICTs and public health indicators empirically using panel data analysis, we contribute to the sparse literature in this particular area. Note that we acknowledge the portfolio of general research that has emerged in health information technology—see special issues of JAIS [22] and JSIS [23]. The findings can help global policy makers strategize on ICT health resource allocation and invest in technologies that would maximize the population health benefits.

ICTs have the potential to radically transform the delivery of public health. There are others, but the key strategic applications are eHealth and mHealth. The findings of this study will help government officials and public health policy makers to formulate strategic decisions regarding the best ICT investments and deployment. For example, the study shows that providing accessibility ought to be a critical focus. While the use of ICTs is one among several strategies in improving global public health, the delivery of public health cannot be replaced or underplayed. The actual carrying out of immunizations and vaccinations and preventive health measures, health education, hygiene, clean water, and sanitation are vital factors in promoting public health. In the long term, better public health can improve productivity, contribute to poverty alleviation, and enhance quality of life overall.

Several research directions are possible. The emerging field of health analytics can be applied to "big health data" sets for more robust analysis resulting in richer insight and informed decision making. While most governmental and NGOs (eg, World Bank, WHO, Rockefeller Foundation) report and publish surveys, there is a paucity of quantitative analysis of the raw data, or aggregated data. The application of health analytics techniques will enrich findings. While this is a study done at the aggregate level, it is important to continue case studies of specific applications and deployments in various countries to gain insight into the sociotechnical dimensions of the role of ICTs in health care. These include social, cultural, and political aspects of what works and what does not. The lessons learned can be shared with characteristically similar countries so as to avoid costly investment mistakes. Thus, knowledge creation and sharing is important. Also, future research may continually monitor and update the variables and data sets in real time and perform cross-country and cross-regional research as well as additional longitudinal studies. Other variables and development indicators may reveal additional associations and effects. An important issue to address is the "reverse effect" or "rebound effect". What are the negative consequences of using or overusing technology? Carpel tunnel syndrome, for instance, is symptomatic of repeated use of the keyboard. Excessive cell phone use or constant staring at electronic screens or monitors also may have adverse effects. Going beyond quantitative studies, one may research the diffusion of ICTs for public health purposes and consider additional concepts and models, such as the introduction of key "disruptive technologies" into health care, the "leap frog" effect, the alleviation of the "digital divide" with regard to health care, "local innovation", and new health care delivery designs and models. Additionally, new resource allocation and investment models and strategies of ICTs may emerge. Their adoption and assimilation may be studied. Globally, the application of ICTs such as eHealth and mHealth is rapidly, if unevenly, proliferating. Much needs to be done to provide universal and equitable health care to all.

Conflicts of Interest

None declared.

References

1. Blaya JA, Fraser HS, Holt B. E-health technologies show promise in developing countries. Health Aff (Millwood) 2010 Feb;29(2):244-251. [doi: 10.1377/hlthaff.2009.0894] [Medline: 20348068]
2. Bonander J, Gates S. Public health in an era of personal health records: opportunities for innovation and new partnerships. J Med Internet Res 2010;12(3):e33 [FREE Full text] [doi: 10.2196/jmir.1346] [Medline: 20699216]
3. Borde A, Fromm C, Kapadia F, Molla DS, Sherwood E, Sorenson JB. White Paper, Global Alliance for ICT and Development, The New School University. 2010. Information & communication technologies for development: health URL: http://german-practice-collection.org/en/download-centre/doc.../900 [WebCite Cache ID 6AbVnKvRI]

4. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
5. GAO. Public health information technology - additional strategic planning needed to guide HHS's efforts to establish electronic situational awareness capabilities. 2010 URL: http://www.gao.gov/assets/320/314167.pdf [WebCite Cache ID 6AaZqGpye]
6. Gerber T, Olazabal V, Brown K, Pablos-Mendez A. An agenda for action on global e-health. Health Aff (Millwood) 2010 Feb;29(2):233-236. [doi: 10.1377/hlthaff.2009.0934] [Medline: 20348066]
7. Heinzelmann PJ, Lugn NE, Kvedar JC. Telemedicine in the future. J Telemed Telecare 2005;11(8):384-390. [Medline: 16356311]
8. International Telecommunication Union. ITU Internet Reports - The Internet of Things. 2005. 2005 URL: http://www.itu.int/osg/spu/publications/internetofthings/InternetofThings_summary.pdf [WebCite Cache ID 6Ac2yJE5A]
9. International Telecommunication Union. Implementing eHealth in Developing Countries: Guidance and Principles. 2008 URL: http://www.itu.int/ITU-D/cyb/app/docs/e-Health_prefinal_15092008.PDF [WebCite Cache ID 6Ac1dAsZf]
10. International Telecommunication Union. Question 14 -2/2 (Final Report) - Mobile eHealth Solutions for Developing Countries. 2010 URL: http://www.itu.int/dms_pub/itu-d/opb/stg/D-STG-SG02.14.2-2010-PDF-E.pdf [WebCite Cache ID 6Ac1pN1YG]
11. Kahn JG, Yang JS, Kahn JS. 'Mobile' health needs and opportunities in developing countries. Health Aff (Millwood) 2010 Feb;29(2):252-258. [doi: 10.1377/hlthaff.2009.0965] [Medline: 20348069]
12. Mars M, Scott RE. Global e-health policy: a work in progress. Health Aff (Millwood) 2010 Feb;29(2):237-243. [doi: 10.1377/hlthaff.2009.0945] [Medline: 20348067]
13. : 237-243. . [doi: 10.1377/hlthaff.2009.0945]
14. World Health Organizatio 2011. Global Observatory for eHealth Series, 3 Geneva. mHealth-new horizons for health through mobile technologies URL: http://www.who.int/goe/publications/goe_mhealth_web.pdf [WebCite Cache ID 6AaZMYhQ8]
15. Dentzer S. E-health's promise for the developing world. Health Aff (Millwood) 2010 Feb;29(2):229. [doi: 10.1377/hlthaff.2010.0006] [Medline: 20348063]
16. Curioso WH, Mechael PN. Enhancing 'M-health' with south-to-south collaborations. Health Aff (Millwood) 2010 Feb;29(2):264-267. [doi: 10.1377/hlthaff.2009.1057] [Medline: 20348071]
17. Raghupathi W, Tan J. Information Systems and Healthcare XXX: charting a strategic path for health information technology. Communications of the Association for Information Systems Dec 2008;23:501-522 [FREE Full text]
18. Raghupathi W, Wu SJ. The relationship between information and communication technologies and the delivery of public health: a country level study. Communications of the Association for Information Systems Feb 2011;28(1):99-116 [FREE Full text]
19. World Health Organization 2005. Report for the World summit on the Information Society. Connecting for health: global vision, local insight URL: http://www.who.int/ehealth/publications/WSISReport_Connecting_for_Health.pdf [WebCite Cache ID 6AaZD5JeY]
20. O'Carroll PW. Public health informatics and information systems. New York: Springer; 2003.
21. Yasnoff WA, O'Carroll PW, Koo D, Linkins RW, Kilbourne EM. Public health informatics: improving and transforming public health in the information age. J Public Health Manag Pract 2000 Nov;6(6):67-75. [Medline: 18019962]
22. World Bank. World View. 2010 URL: http://data.worldbank.org/sites/default/files/section1.pdf [WebCite Cache ID 6AaXfpv56]
23. Payton FC, Pare G, Le Rouge CM, Reddy M. Health care IT: process, people, patients and interdisciplinary considerations. Journal of the Association for Information Systems 2011;12(2) [FREE Full text]
24. Newell S. Editorial: special section on healthcare information systems. Journal of Strategic Information Systems 2011;20(2):158-160.
25. World Bank ICT at-a-glance database. Archived by WebCite® at URL: http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTINFORMATIONANDCOMMUNICATIONANDTECHNOLOGIES/0,,contentMDK:21756274~menuPK:5053110~pagePK:210058~piPK:210062~theSitePK:282823,00.html [WebCite Cache ID 6AbU7Uv5C]
26. World Bank world development indicators database. Archived by WebCite® at. URL: http://data.worldbank.org/data-catalog/world-development-indicators [WebCite Cache ID 6AbTbQfGO]

Edited by G Eysenbach; submitted 26.06.12; peer-reviewed by J Tan, S Kesh, N Ashrafi, M Scotch, B Massoudi; comments to author 20.07.12; revised version received 09.08.12; accepted 25.09.12; published 22.10.12.

Please cite as:
Wu SJ, Raghupathi W
A Panel Analysis of the Strategic Association Between Information and Communication Technology and Public Health Delivery
J Med Internet Res 2012;14(5):e147
URL: http://www.jmir.org/2012/5/e147/
doi:10.2196/jmir.2242
PMID:23089193

Original Paper

A Web-Based Multidrug-Resistant Organisms Surveillance and Outbreak Detection System with Rule-Based Classification and Clustering

Yi-Ju Tseng[1], BS; Jung-Hsuan Wu[2], PhD, MS; Xiao-Ou Ping[3], MS; Hui-Chi Lin[4], BS; Ying-Yu Chen[5], MS; Rung-Ji Shang[6], PhD; Ming-Yuan Chen[6], MS; Feipei Lai[1,2,3], PhD; Yee-Chun Chen[4,7], MD, PhD

[1]Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei, Taiwan
[2]Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan
[3]Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan
[4]Center for Infection Control, National Taiwan University Hospital, Taipei, Taiwan
[5]Graduate Institute of Networking and Multimedia, National Taiwan University, Taipei, Taiwan
[6]Information Systems Office, National Taiwan University Hospital, Taipei, Taiwan
[7]Department of Internal Medicine, National Taiwan University Hospital and College of Medicine, Taipei, Taiwan

Corresponding Author:
Yee-Chun Chen, MD, PhD
Department of Internal Medicine
National Taiwan University Hospital and College of Medicine
No. 7 Chung-Shan South Road
Taipei, 100
Taiwan
Phone: 886 2 23123456 ext 65054
Fax: 886 2 23971412
Email: yeechunchen@gmail.com

Abstract

Background: The emergence and spread of multidrug-resistant organisms (MDROs) are causing a global crisis. Combating antimicrobial resistance requires prevention of transmission of resistant organisms and improved use of antimicrobials.

Objectives: To develop a Web-based information system for automatic integration, analysis, and interpretation of the antimicrobial susceptibility of all clinical isolates that incorporates rule-based classification and cluster analysis of MDROs and implements control chart analysis to facilitate outbreak detection.

Methods: Electronic microbiological data from a 2200-bed teaching hospital in Taiwan were classified according to predefined criteria of MDROs. The numbers of organisms, patients, and incident patients in each MDRO pattern were presented graphically to describe spatial and time information in a Web-based user interface. Hierarchical clustering with 7 upper control limits (UCL) was used to detect suspicious outbreaks. The system's performance in outbreak detection was evaluated based on vancomycin-resistant enterococcal outbreaks determined by a hospital-wide prospective active surveillance database compiled by infection control personnel.

Results: The optimal UCL for MDRO outbreak detection was the upper 90% confidence interval (CI) using germ criterion with clustering (area under ROC curve (AUC) 0.93, 95% CI 0.91 to 0.95), upper 85% CI using patient criterion (AUC 0.87, 95% CI 0.80 to 0.93), and one standard deviation using incident patient criterion (AUC 0.84, 95% CI 0.75 to 0.92). The performance indicators of each UCL were statistically significantly higher with clustering than those without clustering in germ criterion ($P < .001$), patient criterion ($P = .04$), and incident patient criterion ($P < .001$).

Conclusion: This system automatically identifies MDROs and accurately detects suspicious outbreaks of MDROs based on the antimicrobial susceptibility of all clinical isolates.

(J Med Internet Res 2012;14(5):e131) doi:10.2196/jmir.2056

KEYWORDS

multidrug resistance; surveillance; infection control; information systems; cluster analysis; Web-based services

Introduction

During the last few decades, the nonspecific nature and overlapping spectra of early phase infection caused by bacteria and other pathogens have resulted in the overuse of antimicrobials [1]. This parallels a relentless increase in the number and types of microorganisms resistant to these medicines. Patients infected with resistant organisms are more likely to receive inappropriate initial therapy and are, thus, associated with higher mortality, morbidity, and medical costs [2-3]. The emergence and spread of multidrug-resistant organisms (MDROs) are considered to be causing a global crisis [1,4]. To resolve this problem, the WHO has made antimicrobial resistance an organization-wide priority and the focus of the 2011 World Health Day.

Patients can become infected, or a carrier, by exposure to an MDRO-contaminated environment (including medical devices), close contact with a carrier, or following the use of antimicrobial agents. Multidrug-resistant organisms may spread further, resulting in outbreaks and compromising patient safety [5-10]. Combating antimicrobial resistance requires prevention of transmission of resistant organisms and improved use of antimicrobials [1]. The core components of infection control strategies for MDRO, therefore, include early identification, monitoring, and prevention of spread [2]. One or more hospital staff, usually infection control personnel, is responsible for identifying and tracking down carriers or patients with infections caused by epidemiologically important MDROs and preventing further spread. This is usually conducted by reviewing the laboratory reports of all clinical isolates, on a daily or weekly basis, to recognize any unusual clustering or increases in the numbers of specific bacteria or by identifying antimicrobial susceptibility testing results (antibiogram) at indicated units within a period. The key information from MDRO surveillance consists of the species of MDRO, antibiogram, space (unit), and time. The identification of potential MDRO outbreak is a complex process and, therefore, extremely challenging for infection control personnel at a large teaching hospital.

Information technology is expected to improve efficiency in automated surveillance and infection control [11-13]. Previous studies have devised and implemented computer-assisted infection control surveillance or outbreak detection systems, and these have proved beneficial in MDRO surveillance [14-17]. However, prior attempts to establish a Web-based MDRO surveillance system that allows automated real-time integration, analysis, and interpretation have been few. In such a system, visualization methods are also important for the clear presentation of the proximity of time and space, and the species of MDRO, and to facilitate data-driven decision-making.

The present study developed a Web-based MDRO surveillance and outbreak detection information system at a teaching hospital in Taiwan. The system adheres to a Service-Oriented Architecture (SOA) and to Health Level Seven (HL7). It incorporates rule-based classification and cluster analysis of all reported antibiogram profiles, implements control charts with surveillance rules and hierarchical clustering for data analysis, and provides useful information to facilitate the timely targeting of the correct unit by infection control personnel for appropriate intervention. The study includes evaluation of the system performance.

Methods

Hospital Setting and Infection Control Program

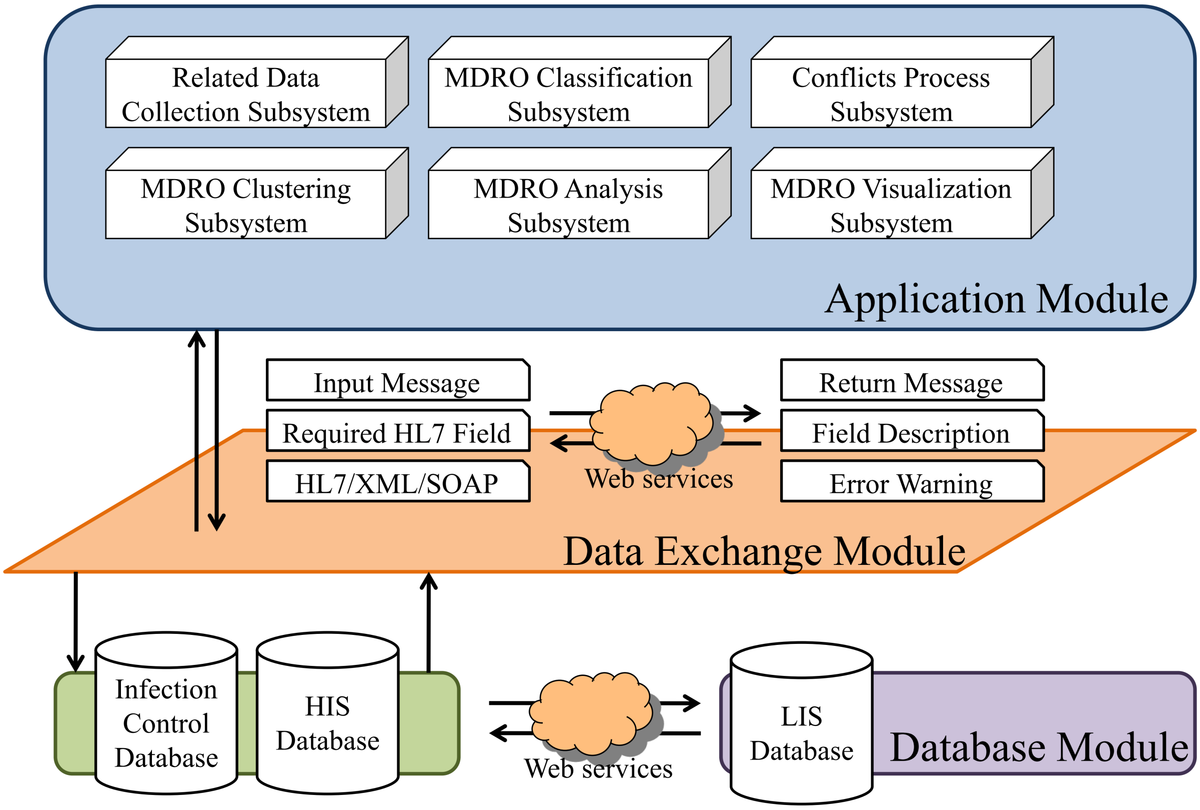
National Taiwan University Hospital (NTUH), a 2200-bed major teaching hospital in Taiwan, provides both primary and tertiary medical care. In 2010, it served 87, 559 inpatients and 2 ,181, 764 outpatients and received 106, 090 emergency visits. A total of 248, 362 specimens were sent for bacterial isolation and identification. There were 20,472 MDROs, and other surveillance target organisms (described in Appendix 1) were identified.

Prospective, hospital-wide on-site surveillance of health care-associated infection was conducted, from its initiation in 1981, by way of weekly visits by infection control personnel to all inpatient units [18]. In addition, infection control personnel monitored culture results from the clinical microbiology laboratory, on a daily basis, to identify any clustering of epidemiologically important MDROs. Oral reminders and formal feedback were provided to the hospital units to strengthen infection control measures. Site visits, audits, and investigations were conducted periodically and if necessary.

A Web-based MDRO surveillance system was developed to automatically and instantaneously detect and monitor the hospitalized patients with MDRO carriage. It has been executed routinely since October 2010. The following sections describe the system architecture and software components. The performance of the system was also evaluated. System Architecture

The MDRO surveillance system includes an application module, a data exchange module, and a database module (Figure 1). The database module consists of the infection control database, health information system (HIS) database, and laboratory information system (LIS) database. The data exchange module connects the application services to the database services through an SOA (service oriented architecture). The HL7-embedded Extensible Markup Language (XML) formatted data are implemented in the data exchange module [19-20] and support message management, routing, mapping, and database access. The application module consists of software components. The MDRO surveillance services can easily be provided to other heterogeneous information systems because of their adherence to SOA and HL7 standards.

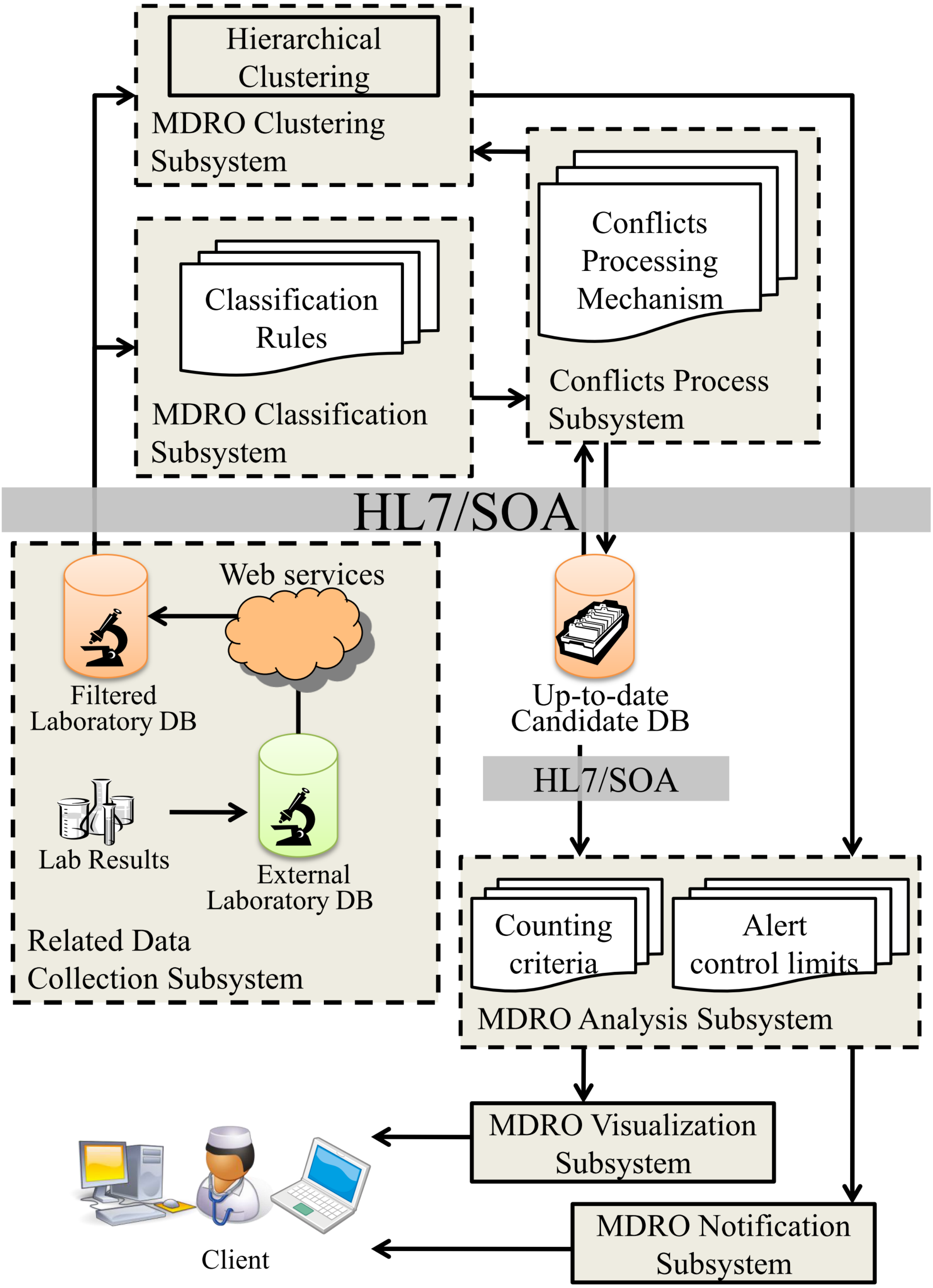
Figure 1. The system architecture of the Web-based multidrug-resistant organisms (MDRO) surveillance system.



Software components

The software components of the application module consist of 7 subsystems for data collection, conflict processing, MDRO classification and clustering, analysis, visualization, and notification (Figure 2). Data from the clinical microbiology laboratory are collected by Web service from the LIS, mapped, and classified according to the predefined criteria of MDRO and other surveillance target organisms (Appendix 1). Multidrug-resistant organisms are stored in the up-to-date candidate database after processing of the conflicts between preliminary and final reports. Meanwhile, the clinical microbiology laboratory data in the filtered laboratory database are grouped by cluster analysis. The MDRO candidates are then analyzed by counting criteria and alert upper control limits (UCLs). The results of analysis are displayed in a Web-based user interface. This Web-based MDRO surveillance system monitors MDROs on a daily basis, and every hour if indicated, and is in conjunction with the HIS, which offers a single entry point to the Web-based interface by way of a browser. The following subsections detail each subsystem. Components of the system, including the related data collection, MDRO classification, conflict process, and notification subsystem have been described previously [21] (Appendix 2). The target organisms for surveillance and definition of the multidrug-resistant organisms are shown in Appendix 1, classified into 5 categories according to classification logics.

Figure 2. The software components of the application module consist of 7 subsystems.



MDRO Clustering Subsystem

The MDRO clustering subsystem automatically compares and analyzes the antimicrobial susceptibility testing results among organisms. For each organism against an antimicrobial agent, there are 4 possible results: R (resistant), I (intermediate), S (sensitive), and missing (when no data are available). For example, ten antimicrobial agents were evaluated for one organism: AN, ATM, CAZ, CIP, FEP, GM, LVX, MEM, SAM, and TIM. The results were I, R, R, I, R, R, S, S, S, and S, respectively. The result was considered an ordered sequence, IRRIRRSSSS, which was analyzed, and the similarities were calculated by hierarchical clustering. Different from K-means clustering, hierarchical clustering does not require determining the number of clusters, and clusters are separated based on their distance [22]. More importantly, the distance concept of hierarchical clustering meets the clinical meaning of organism clustering, and infection control personnel could easily understand the degree of distance between organism clusters.

First, Euclidean distance between each organism was calculated. The distance between two 1-by-n vectors x_p and x_q in one dimension was the absolute value of the difference between their coordinates (defined as in Figure 3). The result of this equation is commonly known as a distance matrix.

The organisms were then grouped into a binary hierarchical cluster tree, which is a multilevel hierarchy, where clusters at one level are joined as clusters in the following level [23]. In this step, pairs of objects that were in proximity were linked using the single linkage. The single linkage uses the smallest distance generated in the previous step between different objects and produces long clusters with large diameters [22,24]. That is, two clusters were merged according to the minimum distance, and it was the property that was needed in this research to cluster each organism under less strict rules (Figure 4).

Following the pairing of objects into binary clusters, the newly formed clusters were grouped into larger clusters until a hierarchical tree was formed, thus using agglomerative methods. Agglomerative hierarchical clustering is a bottom-up clustering method that has been studied and used extensively [25]. In the infection control personnel's view, this bottom-up strategy is more similar to the concept of organism clustering than divisive (top-down) strategy.

The clustering flowchart is shown in Figure 5. With clustering, the MDROs of each class from the rule-based MDRO classification subsystem were clustered in the MDRO clustering subsystem by R [23,26]. The components of a cluster changed by altering the cutting Euclidean distance: the smaller the cutting distance used, the more clusters that were obtained. Different clusters were analyzed separately in the MDRO analysis subsystem. Without clustering, the MDROs of each class were analyzed directly after MDRO classification.

Figure 3. Equation 1.

$$d(p,q)=\sqrt{(p_1-q_1)^2+(p_2-q_2)^2+\cdots+(p_n-q_n)^2}$$

Figure 4. Equation 2.

$$d(C_i,C_j)=\min_{a\in C_i,b\in C_j} d(a,b)$$

MDRO Analysis Subsystem

To assist infection control personnel with judgment and decision making, the MDRO candidates were analyzed based on the temporal and spatial distribution of patients with MDRO colonization or infection. The system used three counting criteria: germ criterion, patient criterion, and incident patient criterion (Table 1). For the germ criterion, the system counted the numbers of positive results in MDRO culture reports in a given period (in this system, 1 week). One patient may have one or more reports (from different body sites or the same body sites at different times). For the patient criterion, the system counted the numbers of patients who have MDRO specimen reports in a given period. When a patient had more than one MDRO specimen report (eg, MRSA isolated from a blood sample, the tip of the central venous catheter, and pus from a bed sore) in a given period, they were counted only once. The data, thus, present the disease burden at a given time. For the incident patient criterion, this system counted the numbers of patients who were newly colonized or infected by an MDRO if they did not have MDRO culture reports in the 30 days previous to the release of the current MDRO culture report. The data were approximately, although not equal to, the number of patients with newly acquired MDRO during their hospital stay.

Figure 5. Flowchart of MDRO cluster analysis.

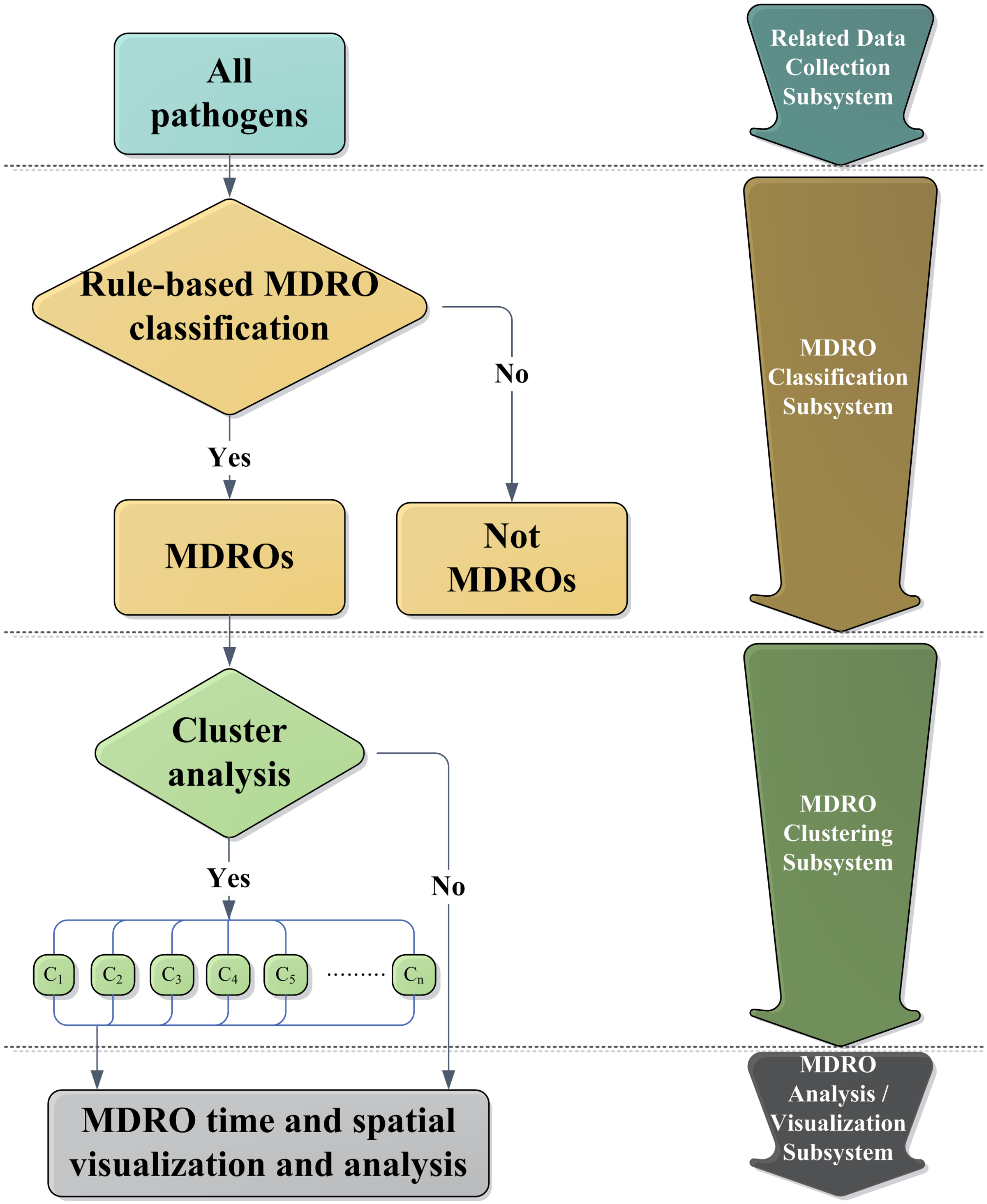


Table 1. The definition and rationale of three counting criteria for multidrug-resistant organism surveillance and outbreak detection system.

Criteria	Definition	Rationale
Germ criterion	The numbers of positive results in MDRO culture reports.	More MDROs isolated from an individual or a group of patients may represent the higher probability of spreading.
Patient criterion	The numbers of patients who have MDRO specimen reports.	The data present the disease burden at a given time. This data may help for resource allocation, such as the use of single room isolation versus cohorting of more than one patients with the same MDRO colonization/infection in the same room.
Incident patient criterion	The numbers of patients who were newly colonized or infected by MDRO if they did not have MDRO culture reports in the 30 days previous to the release of the current MDRO culture report.	The number of patients with newly acquired MDRO during their hospital stay. The higher incident of patients represents the poorer performance of infection control practice. More infection control personnel are likely required to remind, audit, or practice another intervention.

This system also included alert UCLs to identify suspicious outbreaks:

n SD: With this UCL, an alert is defined as *n* standard deviations (*n* SD) above the mean (central line).

m% CI: With this UCL, an alert is defined as lying outside of the *m*% confidence interval (CI) [27]. The value of *m*, defined as the confidence coefficient, could be assigned by users.

These alert UCLs were calculated based on a defined observation period in the past. For example, if the surveillance month was December 2008 and the defined observation period was 1 year, these UCLs were calculated based on the data from December 2007 to November 2008. Due to 3 SD UCL being stricter than 2 SD and 1 SD UCL, there are fewer events flagged as outbreaks by 3 SD UCL. It is anticipated that events that were flagged outside of 2 SD include those defined by 3 SD. Events that were flagged outside of the 95% confidence interval include the 99% confidence interval. The optimal UCL will be determined by the impact (severity or chance for control) of the incident event or the disease surveyed. That is, the more severe the outcome, the more that the lower UCL is preferred.

MDRO Visualization Subsystem

After the MDRO analysis process, a line chart was used to describe time trends of the MDRO count (Figure 6). For the outliers, the clustering results were further presented in a bubble chart to illustrate the spatial distribution of the MDROs, the similarities of the antimicrobial susceptibility testing results, and the MDRO counts in different areas (Figure 7).

The bubble chart shown in Figure 7 represents the clustering results with spatial distribution. The x-axis represents the branch or building of the ward where the patient with MDRO stayed; the y-axis of the bubble chart represents the floor of the ward. The bubble size represents the number of MDROs in a specific ward, which is defined by the x and y axes. The different colors of bubbles indicate that the MRDOs belonged to different clusters. The line chart, describing the time trends of clustering results, supplements the bubble chart (Figure 8).

In addition to time trends and spatial distribution of numbers of MDROs, medical staff can also retrieve detailed information on specimen reports that interest them. With the 2-level embedded function, they can rapidly respond to control the further spread of MDROs. This system facilitates the stream processing of the occurrence of MDROs in the population levels, as well as the identification of patients with MDRO carriage and those in need of special attention.

Figure 6. Control chart for visualization of the time series of the MDRO classification results in a defined patient population (eg, the whole hospital). This chart displays the numbers of vancomycin-resistant Enterococcus faecium isolates (germ criteria, y-axis) by week from March 9, 2008, to May 31, 2008 (x-axis).

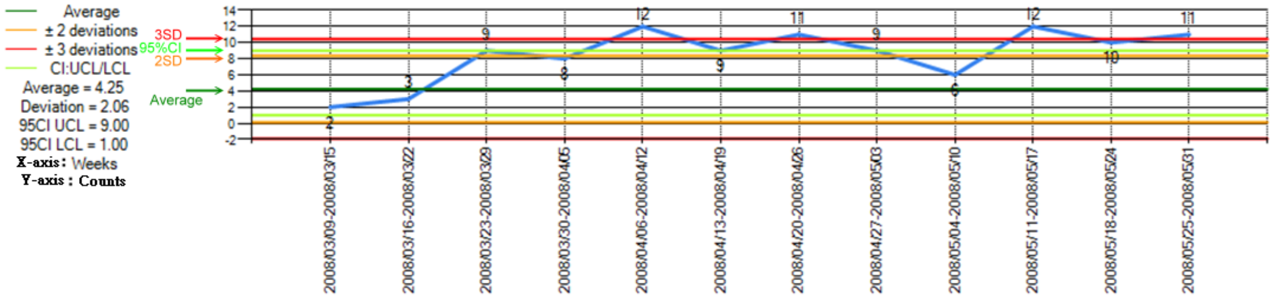


Figure 7. Bubble chart for visualization of the spatial distribution of MDRO clustering results. This chart display the spatial distribution of vancomycin-resistant Enterococcus faecium isolated from April 6, 2008, to April 26, 2008, and clustered with Euclidean distance equal to zero and single linkage.

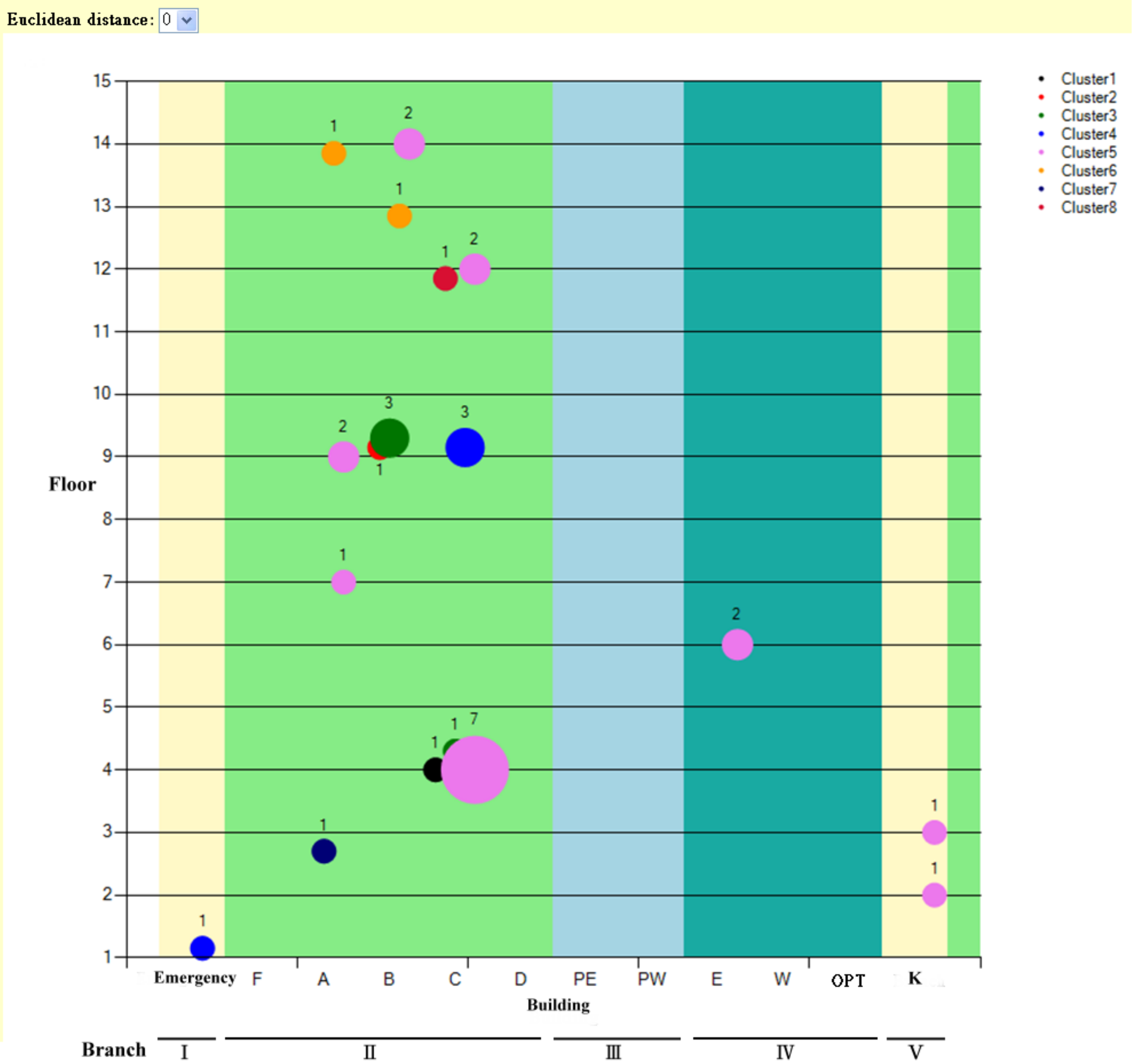
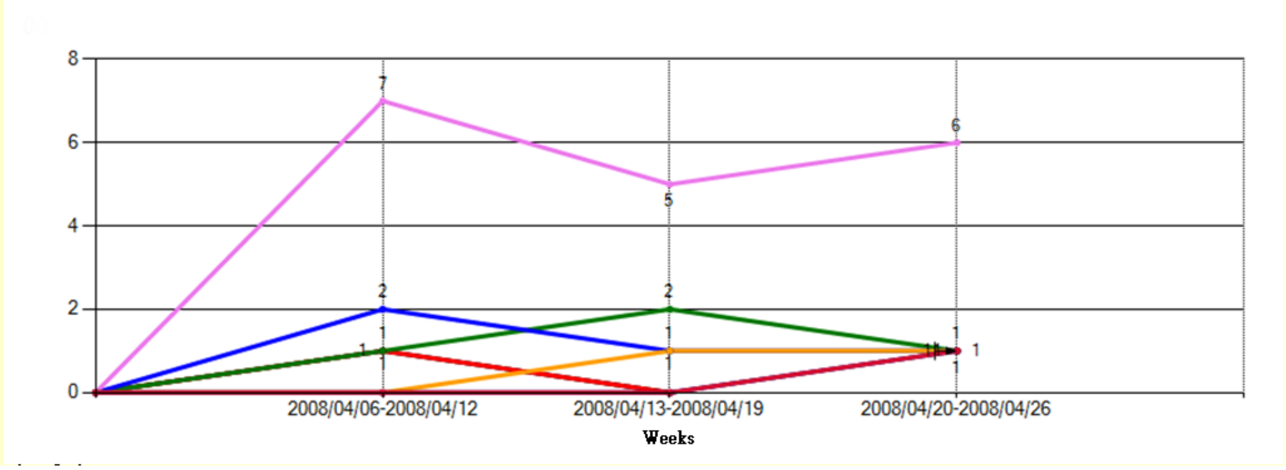


Figure 8. Line chart for visualization of the time trend of MDRO clustering results. This chart display the number of clustered vancomycin-resistant Enterococcus faecium isolates (germ criteria, y-axis), from April 6, 2008, to April 26, 2008, with Euclidean distance equal to zero and single linkage.



Evaluation of Performance

Two levels of performance evaluation were conducted on the Web-based MDRO surveillance system: MDRO detection and classification of specimen reports, and outbreak detection. The performance of the first part of the system described previously [21] (Appendix 2), includes the time cost and accuracy of MDRO detection and classification of specimen reports, and the proportion of patients receiving contact isolation. The proportion of contact isolation orders was defined according to the number of patients with contact isolation orders per 100 patients with MDRO specimen reports. These indicators were compared in the absence of the Web-based MDRO surveillance system (from April 1, 2008, to July 9, 2008) and in the presence of the system (from September 1, 2008, to December 9, 2008). The time cost in the absence of the MDRO system was the average person-minute to identify the MDRO among a set of 100 clinical isolates using antimicrobial susceptibility testing results by 10 hospital staff, multiplied by the total number of clinical isolates per day, and divided by 100.

The MDRO outbreak detection with clustering has been available since April 1, 2011. According to the hospital-wide prospective infection control surveillance data, vancomycin-resistant *Enterococcus* species (VRE) were the leading pathogens to cause outbreaks in 2008. Thus, we used the prospectively defined VRE outbreak data in 2008 (before the implementation of the system) to evaluate the MDRO outbreak detection performance (Figure 9). A suspicious outbreak was defined when a group of patients displayed temporal and spatial clustering of isolations of VRE with identical antibiogram from clinical specimens, and infection control personnel notified the relevant hospital unit to intensify infection control precautions. Only a number of the outbreaks, usually those persisting despite intervention, were investigated further including active microbial surveillance and confirmed by pulsed field gel electrophoresis (PFGE) (Figure 10). A confirmed outbreak was defined by the presence of a predominant clone of VRE from clinical specimens as determined by PFGE among a group of patients with temporal and spatial clustering. The details of the outbreak investigation and the infection control program for VRE were described previously [28].

The stability of the hierarchical clustering algorithm on this dataset was ascertained by clValid R-package and used for comparing different cutting distances of the tree [29]. The package was often used for biomedical clustering [30-32], providing stability measures, including average proportion of non-overlap (APN), average distance (AD), average distance between means (ADM), and figure of merit (FOM). The definitions of these measures were described in previous research [29]. These stability measures compare the results from clustering based on the full data to clustering based on removing each column one at a time.

To detect a suspicious outbreak, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), area under the receiver operating characteristic curve (AUC), and CI of AUC were used to select an optimal UCL for predicting MDRO suspicious outbreaks [33]. The suspicious outbreaks of VRE were identified by 7 UCLs, including the upper 99% CI, upper 95% CI, upper 90% CI, upper 85% UCL, 3 SD, 2 SD, and 1 SD. The UCLs were generated based on data from the previous year.

Figure 9. Data description of suspicious and confirmed outbreaks due to vancomycin-resistant Enterococcus.

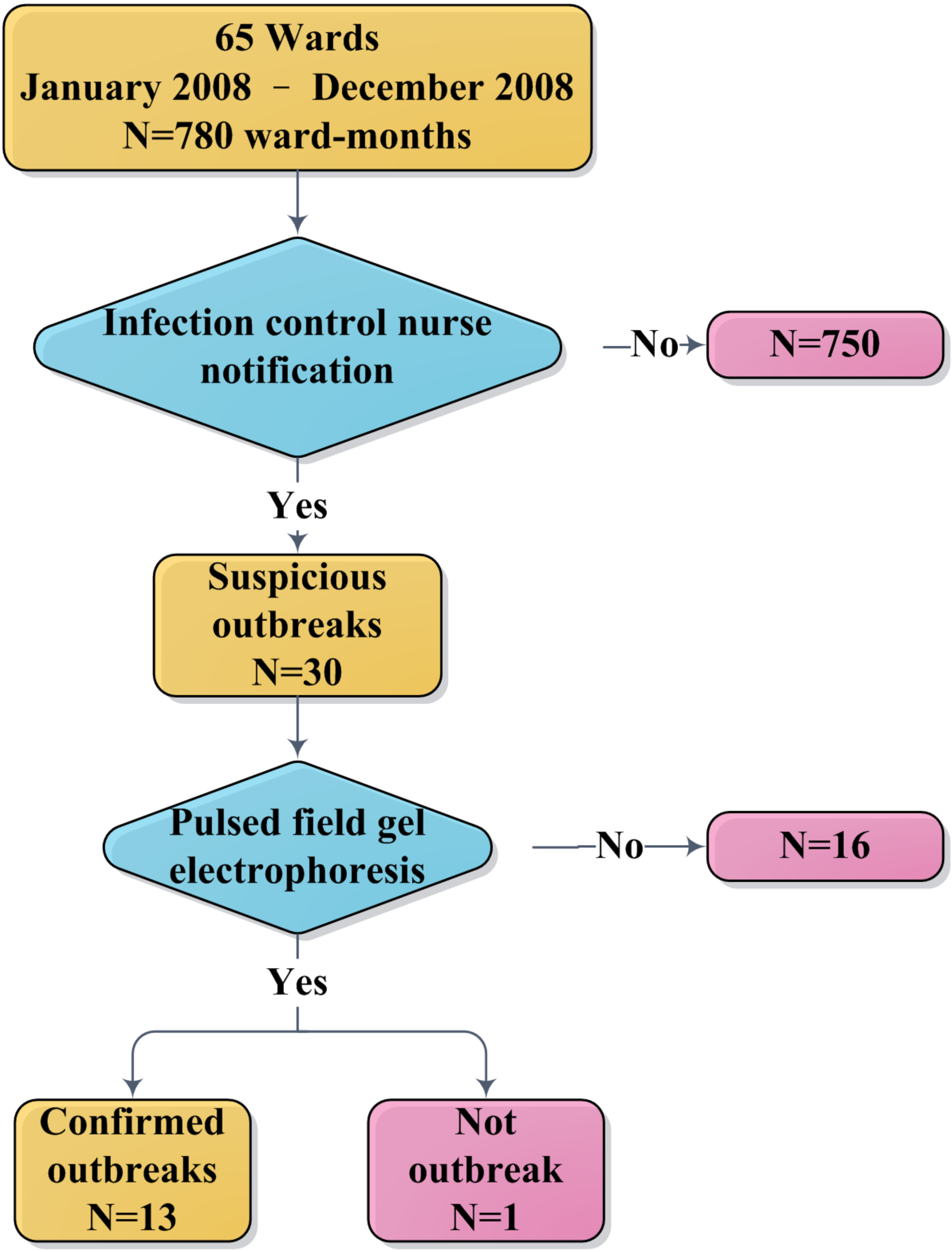
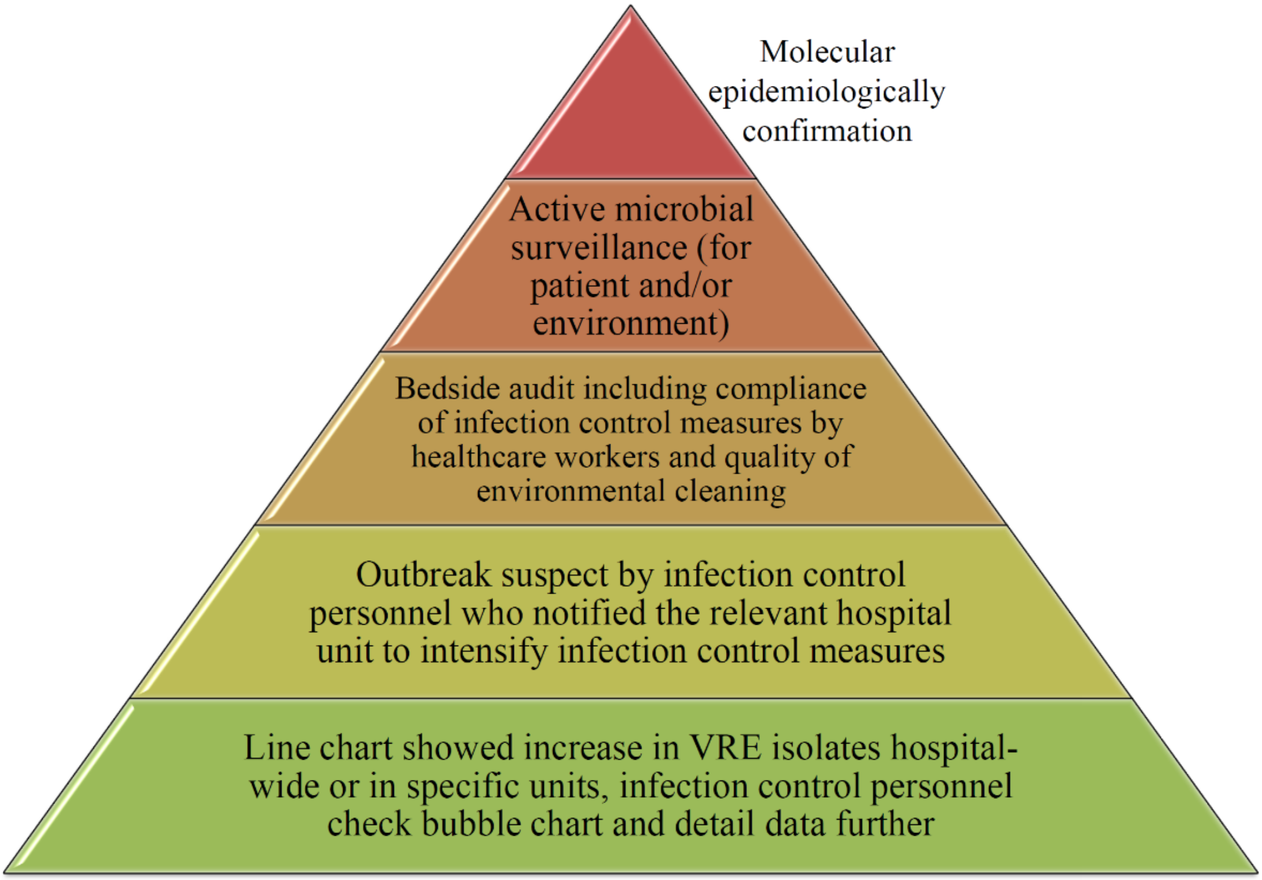


Figure 10. The pyramid of vancomycin-resistant Enterococcus outbreaks identified, investigated, and confirmed, showing the concept of the integrated Web-based surveillance and outbreak detection system and infection control personnel expertise.



Results

The time cost of MDRO detection and classification of specimen reports in the absence of the Web-based MDRO surveillance system was 450 ± 192.2 min per day; in the presence of the system it was 0 min ($P < .001$). That is, implementation of this system may save approximately 1 person-day of the 10 infection control personnel daily. The accuracy of MDRO detection was 63.9 ± 26.4% by infection control personnel; the system was 100% ($P < .001$). The proportion of contact precautions of incident patients increased after implementation of the system (16.5 ± 16.5% versus 25.5 ± 22.1%, $P = .001$) [21] (Appendix 2).

The system evaluated all specimens sent to the routine bacteriology laboratory collected from patients hospitalized in 65 wards during a 12-month study period in 2008. Data from each ward were analyzed by month. Among 780 ward-months evaluated, there were 30 suspicious outbreaks identified by infection control nurses. Of 14 suspicious outbreaks, PFGE study of the VRE isolates confirmed 13 as outbreaks (Figure 9).

Tables 2 to 4 display the system's performance for VRE outbreak detection according to defined UCLs using the germ criterion, patient criterion, or incident patient criterion, with and without clustering. The optimal UCL was upper 90% CI using the germ criterion (number of VRE isolates) with clustering (AUC 0.93, 95% CI 0.91 to 0.95), followed by upper 85% CI using the patient criterion (number of patients with VRE isolated from clinical specimens) (AUC 0.87, 95% CI 0.80 to 0.93), and 1 SD using the incident patient criterion (number of patients with VRE not identified in previous months) (AUC 0.84, 95% CI 0.75 to 0.92). Appendices 3 to 5 display the details of the system's performance for VRE outbreak detection according to each criterion, with and without clustering. The performance indicators of each UCL were statistically significantly higher with cluster analysis than those without cluster analysis in germ criterion ($P < .001$), patient criterion ($P = .04$), and incident patient criterion ($P < .001$).

The stability of clustering is shown in Table 5. AD and FOM measures were optimized in all criteria with cutting Euclidean distance being zero.

Table 2. Performance in outbreak detection according to germ criterion and a upper control limit defined by 90% confidence interval, with and without clustering.

Parameter	Without clustering	With clustering (d=0)[g]
Sensitivity (%)[a]	90.0 (27/30)	100 (30/30)
Specificity (%)[b]	84.0 (630/750)	86.7 (650/750)
PPV (%)[c]	18.4 (27/147)	23.1 (30/130)
NPV (%)[d]	99.5 (630/633)	100 (650/650)
AUC[e] (95% CI[f])	0.87 (0.81-0.94)	0.93 (0.91-0.95)

[a] Sensitivity = TP/(TP+FN) (where TP (true positive) is an outbreak correctly identified as an outbreak, and FN (false negative) is an outbreak wrongly identified as a non-outbreak).

[b] Specificity = TN/(TN+FP) (where TN (true negative) is a non-outbreak correctly identified as a non-outbreak, and FP (false positive) is a non-outbreak wrongly identified as an outbreak).

[c] Positive predictive value (PPV) = TP/(TP+FP) (variables defined above).

[d] Negative predictive value (NPV) = TN/(TN+FN) (variables defined above).

[e] AUC: area under receiver operating characteristic curve.

[f] CI: confidence interval.

[g] d: cutting Euclidean distance.

Table 3. Performance in outbreak detection according to patient criterion and a variety of upper control limits defined by 1 standard deviation and 85% confidence interval, with and without clustering.

	Without clustering	With clustering (d=1)[g]	With clustering (d=0)
	1 standard deviation	1 standard deviation	Upper 85% CI
Sensitivity (%)[a]	86.7 (26/30)	86.7 (26/30)	90.0 (27/30)
Specificity (%)[b]	83.5 (626/750)	83.2 (624/750)	83.6 (627/750)
PPV (%)[c]	17.3 (26/150)	17.1 (26/152)	18.0 (27/150)
NPV (%)[d]	99.4 (626/630)	99.4 (624/628)	99.5 (627/630)
AUC[e] (95% CI[f])	0.85 (0.78-0.92)	0.85 (0.78-0.92)	0.87 (0.80-0.93)

[a-g] as shown in Table 2 footnotes.

Table 4. Performance in outbreak detection according to incident patient criterion and a variety of upper control limits defined by 1 standard deviation and 90% confidence interval, with and without clustering.

	Without clustering	With clustering (d=1)[g]	With clustering (d=0)
	Upper 90% CI	Upper 90% CI	1 standard deviation
Sensitivity (%)[a]	76.7 (23/30)	76.7 (23/30)	80.0 (24/30)
Specificity (%)[b]	83.9 (629/750)	83.9 (629/750)	87.3 (655/750)
PPV (%)[c]	16.0 (23/144)	16.0 (23/144)	20.2 (24/119)
NPV (%)[d]	99.0 (629/636)	99.0 (629/636)	99.1 (655/661)
AUC[e] (95% CI[f])	0.80 (0.72-0.89)	0.80 (0.72-0.89)	0.84 (0.75-0.92)

[a-g] as shown in Table 2 footnotes.

Table 5. Stability of clustering with germ, patient, and incident patient criteria and a variety of cutting Euclidean distance.

Criteria	Cutting distance[a]	APN[b]	AD[c]	ADM[d]	FOM[e]
Germ criterion	0	0.00	0.12	0.12	0.18
Patient criterion	0	0.00	0.12	0.12	0.17
	1	0.00	0.78	0.01	0.20
Incident patient criterion	0	0.00	0.13	0.13	0.18
	1	0.00	0.83	0.01	0.21

[a] d: cutting Euclidean distance.

[b] Average proportion of non-overlap (APN).

[c] Average distance (AD).

[d] Average distance between means (ADM).

[e] Figure of merit (FOM).

Discussion

Principal Results

This study demonstrated the concept that a Web-based MDRO surveillance and outbreak detection information system provides useful information to facilitate the timely targeting of the correct unit by infection control personnel for appropriate intervention and described how to achieve a fine balance in the use of automated cluster detection tools between capturing all statistical clusters versus capturing all clinically meaningful clusters (Figure 10).

This study established a variety of UCLs and clustering methods for the detection and intervention of outbreaks caused by MDROs. The highest area under the ROC curve for detecting VRE outbreaks was 0.93 using a UCL defined by 90% CI with clustering and the germ criterion. In all criteria, the performance indicators of each UCL were statistically significant higher with clustering than those without clustering. For the stability of clustering, although ADM measure was optimized in all criteria with cutting Euclidean distance being 1, AD and FOM measures were optimized in all criteria with cutting Euclidean distance being zero. Average distance (AD) and FOM measures compute distance and variance of observations in the same cluster between original data and single column removed data. Because the observations in a cluster have the same property with cutting Euclidean distance being zero, the effect of removing a single column in AD and FOM measures could be smaller than others. Different from that, ADM computes the distance of the cluster center for observations placed in the same cluster between the original data and single column removed data. The distance between cluster centers could increase as the number of clusters increases. Even though ADM was not optimized with cutting Euclidean distance being zero, it was a relatively small value in the range from zero to infinity.

The optimal UCLs for detecting outbreaks of VRE in 2008 were presented; however, whether these UCLs remain useful in different MDROs warrants further study. Therefore, this system designated confidence coefficients or multiples of standard deviation to provide the different UCLs, within a user-friendly interface, and incorporates a greater number of possible future scenarios. Infection control experts can, therefore, select 1 or more of these integrated UCLs, which are useful in specific scenarios, then manage the suspicious outbreak with minimal delay. In addition, prospectively defined VRE outbreak was recorded by month; therefore, the performance of the system was determined by month. The control chart presents the number of VRE isolates or patients by week. The system could, therefore, provide earlier detection of suspicious outbreak than detection by infection control nurses.

The traditional expert-defined MDRO classification system suffers from blind spots: it can fail to notice undefined new organisms. It classifies MDROs according to expert-defined rules; therefore, if organisms are not within the defined classification rules, they escape detection. To solve this problem, this study's new Web-based MDRO-surveillance system combines cluster analysis and traditional MDRO classification. Unlike the traditional MDRO classification system, it groups MDROs according to the antimicrobial susceptibility profiles in cluster analysis. If the unexpected MDROs have antibiogram profiles, the cluster analysis, therefore, also analyzes their suspicious outbreak.

In accordance with the results of previous studies, including those of Kho et al [14], using a computer-assisted system provided a higher proportion of contact isolations than without computer assistance. A computerized decision support system could further improve risk measurement in other aspects of health care [34-35]. The present computer-assisted MDRO surveillance system can provide precise and consistent support to the health care provider [14,16,36-37] and demonstrates usefulness for infection control. Huang et al implemented an automated statistical software that provided valuable real-time guidance by both identifying otherwise unrecognized outbreaks and preventing the unnecessary implementation of resource-intensive infection control measures that interfere with regular patient care [38]. Different from that, this system is Web-based and can be easily adapted by others because of adherence to SOA and HL7 standards. The SOA can also be seamlessly integrated into the HIS and facilitate infection control personnel in obtaining more information from patients. Furthermore, the Web-based system is not only integrated with a clustering algorithm but also provides a visualization tool and adjustable surveillance parameters within a user-friendly interface for MDRO surveillance.

Limitations

Although our results suggested that the Web-based MDRO surveillance system performs well, it did have several limitations. First, the affected factors of the system include data integrity and instantaneity. The system bases its MDRO classifications and clustering on antibiogram profiles of organisms; thus, it neglects organisms without antimicrobial susceptibility testing results. Second, the performance of the outbreak detection system was evaluated based on the suspicious outbreaks, but not those with PFGE confirmation. Pulsed field gel electrophoresis is currently considered the gold standard for molecular epidemiological characterization of VRE outbreaks [28,39]. However, only some of the outbreaks, usually those persisting despite intervention, were investigated further and confirmed by PFGE (Figure 10). Thus, if limited to those with PFGE, confirmation might lead to underestimation of the occurrence of outbreak.

Third, low PPV risks alert fatigue, particularly if too many alerts are deemed clinically insignificant by infection control personnel, is an important issue. Nevertheless, the important contribution of this system was to provide a user-defined, adoptable, and flexible tool, and to facilitate and assist MDRO surveillance. Fourth, this system has yet to incorporate appropriate statistical analysis for rare events. The rare events in historical data, which defined the central line and UCLs and were used to provide a baseline, were both zero. In 2007, VRE were very rarely isolated from the hospitalized patients (historical data used to define the UCLs). This system is, therefore, always sensitive to new cases, which increases the false positive rate. Fifth, the system did not active alert for any outbreak; all UCL selected outbreaks were only flagged in the line chart and bubble chart waiting for infection control personnel to check, judge, and act accordingly.

Sixth, the contribution of the Web-based surveillance system to any decrease in the rate of incident MDROs detected or the rates of health care-associated infections was unclear. The rate of VRE colonization/infection at this hospital was less than predicted after 2009 [40]. However, hand hygiene promotion and active surveillance also contributed. Finally, the indicators that were used for evaluating MDRO detection and classification of specimen reports were compared over different time periods, which limited the strength of the evidence for the system. Ideally randomization would be used, over the same time period comparing the performance between absence and presence of system, but this isn't always possible or feasible. Similar evaluation methods had been used in previous studies [41-43].

Conclusion

In conclusion, this study presents a Web-based MDRO surveillance system that automatically and efficiently classifies, and accurately clusters, MDROs according to the antimicrobial susceptibility profiles, thus detecting potential MDRO clustering. The data connections optimally represent the full conceptual content of the data, allowing automated integration and data-driven decision making. The results may alert hospital personnel to implement contact precaution measures for patients with MDROs. The system is able to save 1 person-day for 10 infection control nurses and also to instigate outbreak intervention and management. As of May 15, 2012, the system was still used for MDRO surveillance and has become an indispensable tool for infection control personnel's daily work. Implementing this system could, therefore, improve patient safety as well as the quality of medical care in a hospital.

Acknowledgments

The authors would like to acknowledge the members of the Center of Infection Control at National Taiwan University Hospital for assistance with system evaluation and also the members of the Information Systems Office in National Taiwan University Hospital for their contributions toward this study.

YC Chen received grants from the Center for Disease Control, Department of Health (DOH 98-DC-1007), and from the Department of Health, Taiwan (DOH100-TD-B-111-001). F Lai received grants from National Science Council, Taiwan (NSC 98-2923-E-002-004-MY3). The funding organizations were not involved in designing or conducting the study, data collection, management, analysis, or interpretation, nor the preparation, review, and approval of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Target organisms for surveillance and definition of the multidrug-resistant organisms. The patterns are classified into 5 categories according to classification logics.

[PDF File (Adobe PDF File), 70KB - jmir_v14i5e131_app1.pdf]

Multimedia Appendix 2

Real-time Automated MDRO Surveillance System (Reference 21).

[PDF File (Adobe PDF File), 161KB - jmir_v14i5e131_app2.pdf]

Multimedia Appendix 3

Performance in outbreak detection according to a variety of germ criterion and upper control limits (UCL), with and without clustering analysis.

[PDF File (Adobe PDF File), 50KB - jmir_v14i5e131_app3.pdf]

Multimedia Appendix 4

Performance in outbreak detection according to patient criterion and a variety of upper control limits (UCL), with and without clustering analysis.

[PDF File (Adobe PDF File), 51KB - jmir_v14i5e131_app4.pdf]

Multimedia Appendix 5

Performance in outbreak detection according to incident patient criterion and a variety of upper control limit (UCL), with and without clustering analysis.

[PDF File (Adobe PDF File), 51KB - jmir_v14i5e131_app5.pdf]

References

1. World Health Organization. Drug resistance. 2011 URL: http://www.who.int/drugresistance/en/ [accessed 2011-05-05] [WebCite Cache ID 5yT8Qvg3e]
2. Centers for Disease Control and Prevention. Management of Multidrug-Resistant Organisms In Healthcare Settings. 2006 URL: http://www.cdc.gov/hicpac/mdro/mdro_toc.html [accessed 2012-01-17] [WebCite Cache ID 64kiNWBAs]
3. Hospital Infection Control Practices Advisory Committee (HICPAC). Recommendations for preventing the spread of vancomycin resistance. MMWR Recomm Rep 1995 Sep 22;44(RR-12):1-13 [FREE Full text] [Medline: 7565541]
4. Kumarasamy KK, Toleman MA, Walsh TR, Bagaria J, Butt F, Balakrishnan R, et al. Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. Lancet Infect Dis 2010 Sep;10(9):597-602. [doi: 10.1016/S1473-3099(10)70143-2] [Medline: 20705517]
5. Grundmann H, Bärwolff S, Tami A, Behnke M, Schwab F, Geffers C, et al. How many infections are caused by patient-to-patient transmission in intensive care units? Crit Care Med 2005 May;33(5):946-951. [Medline: 15891318]
6. Wenzel RP, Brewer TF, Butzler JP. A Guide to Infection Control in the Hospital. In: Guide to Infection Control in the Hospital. Hamilton, Ontario: B.C. Decker; 2002.
7. Burke JP. Infection control - a problem for patient safety. N Engl J Med 2003 Feb 13;348(7):651-656. [doi: 10.1056/NEJMhpr020557] [Medline: 12584377]
8. Sheng WH, Wang JT, Lu DC, Chie WC, Chen YC, Chang SC. Comparative impact of hospital-acquired infections on medical costs, length of hospital stay and outcome between community hospitals and medical centres. J Hosp Infect 2005 Mar;59(3):205-214. [doi: 10.1016/j.jhin.2004.06.003] [Medline: 15694977]
9. Allegranzi B, Storr J, Dziekan G, Leotsakos A, Donaldson L, Pittet D. The First Global Patient Safety Challenge . J Hosp Infect 2007 Jun;65 Suppl 2:115-123. [doi: 10.1016/S0195-6701(07)60027-9] [Medline: 17540254]
10. Lindenauer PK, Remus D, Roman S, Rothberg MB, Benjamin EM, Ma A, et al. Public reporting and pay for performance in hospital quality improvement. N Engl J Med 2007 Feb 1;356(5):486-496 [FREE Full text] [doi: 10.1056/NEJMsa064964] [Medline: 17259444]
11. Wright M. Automated surveillance and infection control: towards a better tomorrow. Am J Infect Control 2008 ;36(S1):S1-S6.
12. Halpin H, Shortell SM, Milstein A, Vanneman M. Hospital adoption of automated surveillance technology and the implementation of infection prevention and control programs. Am J Infect Control 2011 May;39(4):270-276. [doi: 10.1016/j.ajic.2010.10.037] [Medline: 21531272]
13. Hota B, Jones RC, Schwartz DN. Informatics and infectious diseases:what is the connection and efficacy of information technology tools for therapy and health care epidemiology? Am J Infect Control 2008 Apr;36(S3):S47-S56.
14. Kho AN, Dexter PR, Warvel JS, Belsito AW, Commiskey M, Wilson SJ, et al. An effective computerized reminder for contact isolation of patients colonized or infected with resistant organisms. Int J Med Inform 2008 Mar;77(3):194-198. [doi: 10.1016/j.ijmedinf.2007.02.005] [Medline: 17398145]
15. Aboelela SW, Saiman L, Stone P, Lowy FD, Quiros D, Larson E. Effectiveness of barrier precautions and surveillance cultures to control transmission of multidrug-resistant organisms: a systematic review of the literature. Am J Infect Control 2006 Oct;34(8):484-494. [doi: 10.1016/j.ajic.2006.03.008] [Medline: 17015153]
16. Hacek DM, Cordell RL, Noskin GA, Peterson LR. Computer-assisted surveillance for detecting clonal outbreaks of nosocomial infection. J Clin Microbiol 2004 Mar;42(3):1170-1175 [FREE Full text] [Medline: 15004070]
17. Wieland SC, Brownstein JS, Berger B, Mandl KD. Automated real time constant-specificity surveillance for disease outbreaks. BMC Med Inform Decis Mak 2007;7:15 [FREE Full text] [doi: 10.1186/1472-6947-7-15] [Medline: 17567912]

18. Chuang YC, Chen YC, Chang SC, Sun CC, Chang YY, Chen ML, et al. Secular trends of healthcare-associated infections at a teaching hospital in Taiwan, 1981-2007. J Hosp Infect 2010 Oct;76(2):143-149. [doi: 10.1016/j.jhin.2010.05.001] [Medline: 20663585]
19. Ko LF, Lin JC, Chen CH, Chang JS, Lai F, Hsu KP, et al. HL7 middleware framework for healthcare information system. 2006 Presented at: Proceeding of the 8th International Conference on e-Health Networking, Applications and Services; Aug; 2006; New Delhi, India.
20. Hsieh SH, Hsieh SL, Weng YC, Yang TH, Lai F, Cheng PH, et al. Middleware based Inpatient Healthcare Information System. 2007 Presented at: Proceedings of the 7th IEEE International Conference on BioinformaticsBioengineering; Oct; 2007; Boston, MA, USA. [doi: 10.1109/BIBE.2007.4375721]
21. Wu JH, Chen YC, Hsieh SH, Lin HC, Chen YY, Cheng PH, et al. Proceedings. In: Real-time Automated MDRO Surveillance System. Las Vegas, Nevada, USA: CSREA Press; 2009 Presented at: The 2009 International Conference on Bioinformatics & Computational Biology; 2009 July 13-16; Las Vegas, Nevada.
22. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference and prediction. New York, N.Y: Springer; 2009.
23. Lucas A. amap: Another Multidimensional Analysis Package. 2010 URL: http://cran.r-project.org/web/packages/amap/index.html [accessed 2011-05-06] [WebCite Cache ID 5yTr8iSUb]
24. Gower JC, Ross GJS. Minimum Spanning Trees and Single Linkage Cluster Analysis. Applied Statistics 1969;18(1):54-64. [doi: 10.2307/2346439]
25. Han J, Kamber M. Data mining: concepts and techniques. San Francisco, CA: Morgan Kaufmann; 2005.
26. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2011.
27. Scrucca L. qcc: an R package for quality control charting and statistical process control. R News 2004 Jun;4(1) [FREE Full text]
28. Chuang YC, Wang JT, Chen ML, Chen YC. Comparison of an automated repetitive-sequence-based PCR microbial typing system with pulsed-field gel electrophoresis for molecular typing of vancomycin-resistant Enterococcus faecium. J Clin Microbiol 2010 Aug;48(8):2897-2901 [FREE Full text] [doi: 10.1128/JCM.00136-10] [Medline: 20554812]
29. Brock G, Pihur V, Datta S, Datta S. clValid: An R Package for Cluster Validation. Journal of Statistical Software 2008 Mar;25(4).
30. Kalina T, Stuchlý J, Janda A, Hrusák O, Růzicková S, Sedivá A, et al. Profiling of polychromatic flow cytometry data on B-cells reveals patients' clusters in common variable immunodeficiency. Cytometry A 2009 Nov;75(11):902-909 [FREE Full text] [doi: 10.1002/cyto.a.20801] [Medline: 19802875]
31. Ng AY, Rajapakse JC, Welsch RE, Matsudaira PT, Horodincu V, Evans JG. A cell profiling framework for modeling drug responses from HCS imaging. J Biomol Screen 2010 Aug;15(7):858-868. [doi: 10.1177/1087057110372256] [Medline: 20525958]
32. Cooke EJ, Savage RS, Kirk PD, Darkins R, Wild DL. Bayesian hierarchical clustering for microarray time series data with replicates and outlier measurements. BMC Bioinformatics 2011 Oct;12:399 [FREE Full text] [doi: 10.1186/1471-2105-12-399] [Medline: 21995452]
33. Bravata DM, Sundaram V, McDonald KM, Smith WM, Szeto H, Schleinitz MD, et al. Evaluating detection and diagnostic decision support systems for bioterrorism response. Emerg Infect Dis 2004 Jan;10(1):100-108 [FREE Full text] [doi: 10.3201/eid1001.030243] [Medline: 15078604]
34. Peiris DP, Joshi R, Webster RJ, Groenestein P, Usherwood TP, Heeley E, et al. An electronic clinical decision support tool to assist primary care providers in cardiovascular disease risk management: development and mixed methods evaluation. J Med Internet Res 2009;11(4):e51 [FREE Full text] [doi: 10.2196/jmir.1258] [Medline: 20018588]
35. Evans R, Joseph-Williams N, Edwards A, Newcombe RG, Wright P, Kinnersley P, et al. Supporting informed decision making for prostate specific antigen (PSA) testing on the web: an online randomized controlled trial. J Med Internet Res 2010;12(3):e27 [FREE Full text] [doi: 10.2196/jmir.1305] [Medline: 20693148]
36. Steurbaut K, Van Hoecke S, Colpaert K, Lamont K, Taveirne K, Depuydt P, et al. Use of web services for computerized medical decision support, including infection control and antibiotic management, in the intensive care unit. J Telemed Telecare 2010;16(1):25-29. [doi: 10.1258/jtt.2009.001008] [Medline: 20086264]
37. Young J, Stevenson KB. Real-time surveillance and decision-support: Optimizing infection control and antimicrobial choices at the point of care. Am J Infect Control 2008 Apr;36(3, Supplement 1):S67-S74.
38. Huang SS, Yokoe DS, Stelling J, Placzek H, Kulldorff M, Kleinman K, et al. Automated detection of infectious disease outbreaks in hospitals: a retrospective cohort study. PLoS Med 2010 Feb;7(2):e1000238 [FREE Full text] [doi: 10.1371/journal.pmed.1000238] [Medline: 20186274]
39. Turabelidze D, Kotetishvili M, Kreger A, Morris JG, Sulakvelidze A. Improved pulsed-field gel electrophoresis for typing vancomycin-resistant enterococci. J Clin Microbiol 2000 Nov;38(11):4242-4245 [FREE Full text] [Medline: 11060099]
40. Chen YC, Chuang YC. Modelling the impact of infection control measures on the incidence of vancomycin-resistance enterococci: a time-series analysis. BMC Proceedings 2011;5(Suppl 6):P25. [doi: 10.1186/1753-6561-5-S6-P25]

41. Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. JAMA 1998 Oct 21;280(15):1339-1346. [Medline: 9794315]
42. de Vries EN, Prins HA, Crolla RM, den Outer AJ, van Andel G, van Helden SH, SURPASS Collaborative Group. Effect of a comprehensive surgical safety system on patient outcomes. N Engl J Med 2010 Nov 11;363(20):1928-1937 [FREE Full text] [doi: 10.1056/NEJMsa0911535] [Medline: 21067384]
43. Hermsen ED, VanSchooneveld TC, Sayles H, Rupp ME. Implementation of a clinical decision support system for antimicrobial stewardship. Infect Control Hosp Epidemiol 2012 Apr;33(4):412-415. [doi: 10.1086/664762] [Medline: 22418640]

Abbreviations

AD: Average distance
ADM: Average distance between means
APN: Average proportion of non-overlap
AUC: area under the receiver operating characteristic curve
CI: confidence interval
DB: database
FOM: Figure of merit
HIS: health information system
HL7: Health Level Seven
LIS: laboratory information system
MDRO: multidrug-resistant organisms
MRSA: methicillin-resistant Staphylococcus aureus
NPV: negative predictive value
PPV: positive predictive value
ROC: receiver operating characteristic
SD: standard deviation
SOA: service oriented architecture
SOAP: simple object access protocol
UCL: upper control limit
VRE: vancomycin-resistant Enterococcus
XML: extensible markup language

Edited by K El Emam; submitted 06.02.12; peer-reviewed by M Klompas, L Arbuckle; comments to author 03.04.12; revised version received 15.05.12; accepted 12.06.12; published 24.10.12.

Please cite as:
Tseng YJ, Wu JH, Ping XO, Lin HC, Chen YY, Shang RJ, Chen MY, Lai F, Chen YC
A Web-Based Multidrug-Resistant Organisms Surveillance and Outbreak Detection System with Rule-Based Classification and Clustering
J Med Internet Res 2012;14(5):e131
URL: http://www.jmir.org/2012/5/e131/
doi:10.2196/jmir.2056
PMID:23195868

Original Paper

Innovation in Weight Loss Programs: A 3-Dimensional Virtual-World Approach

Jeanne D Johnston[1], PhD; Anne P Massey[2], PhD; Celeste A DeVaneaux[3]

[1]Department of Kinesiology, Indiana University, Bloomington, IN, United States

[2]Operations and Decision Technologies Department, Indiana University, Bloomington, IN, United States

[3]Club One Island, San Francisco, CA, United States

Corresponding Author:
Jeanne D Johnston, PhD
Department of Kinesiology
Indiana University
1025 East 7th Street
HPER 112
Bloomington, IN, 47405
United States
Phone: 1 812 855 5073
Fax: 1 812 856 5222
Email: jdjohnst@indiana.edu

Abstract

Background: The rising trend in obesity calls for innovative weight loss programs. While behavioral-based face-to-face programs have proven to be the most effective, they are expensive and often inaccessible. Internet or Web-based weight loss programs have expanded reach but may lack qualities critical to weight loss and maintenance such as human interaction, social support, and engagement. In contrast to Web technologies, virtual reality technologies offer unique affordances as a behavioral intervention by directly supporting engagement and active learning.

Objective: To explore the effectiveness of a virtual-world weight loss program relative to weight loss and behavior change.

Methods: We collected data from overweight people (N = 54) participating in a face-to-face or a virtual-world weight loss program. Weight, body mass index (BMI), percentage weight change, and health behaviors (ie, weight loss self-efficacy, physical activity self-efficacy, self-reported physical activity, and fruit and vegetable consumption) were assessed before and after the 12-week program. Repeated measures analysis was used to detect differences between groups and across time.

Results: A total of 54 participants with a BMI of 32 (SD 6.05) kg/m^2 enrolled in the study, with a 13% dropout rate for each group (virtual world group: 5/38; face-to-face group: 3/24). Both groups lost a significant amount of weight (virtual world: 3.9 kg, $P < .001$; face-to-face: 2.8 kg, $P = .002$); however, no significant differences between groups were detected ($P = .29$). Compared with baseline, the virtual-world group lost an average of 4.2%, with 33% (11/33) of the participants losing a clinically significant (≥5%) amount of baseline weight. The face-to-face group lost an average of 3.0% of their baseline weight, with 29% (6/21) losing a clinically significant amount. We detected a significant group × time interaction for moderate ($P = .006$) and vigorous physical activity ($P = .008$), physical activity self-efficacy ($P = .04$), fruit and vegetable consumption ($P = .007$), and weight loss self-efficacy ($P < .001$). Post hoc paired *t* tests indicated significant improvements across *all* of the variables for the virtual-world group.

Conclusions: Overall, these results offer positive early evidence that a virtual-world-based weight loss program can be as effective as a face-to-face one relative to biometric changes. In addition, our results suggest that a virtual world may be a *more effective* platform to influence meaningful behavioral changes and improve self-efficacy.

(J Med Internet Res 2012;14(5):e120) doi:10.2196/jmir.2254

KEYWORDS

Virtual world; obesity; weight loss programs; Internet technology; behavior change

Introduction

Recent national survey data indicate that 35.7% of the US population is obese and 68.8% overweight [1]. Obesity can reduce quality of life and increase the risk for many serious chronic diseases and premature death. As little as a 5%-10% reduction in weight can diminish the risk for diseases such as coronary heart disease and type 2 diabetes [2]. While there are a variety of means to achieve weight loss, one approach is a formal weight loss program. Traditionally, these programs have been delivered via face-to-face programming either in the community or in a commercial or clinical setting. They typically emphasize lesser rather than dramatic changes via weekly small group meetings over a multiweek (16-26 weeks) period. On average, these programs produce a weight loss of 0.4 to 0.5 kg per week with an overall loss of 7%-10% [3,4].

Over the last decade or so, Internet or Web-based programming has emerged [5-9]. This is not surprising given that, as of 2012, the global Internet penetration rate was reported as 32.7%, with North America leading all geographic regions at 78.6% [10]. American adults use the Web for an increasing array of activities, including over 80% who report seeking health-related information [11]. In the context of weight loss, Web offerings address many of the challenges of face-to-face programming, including location-neutral access and often lower cost [12-14]. And, for people who find face-to-face programming intimidating, online programs offer some degree of, if not complete, anonymity. However, results of Web-based weight loss interventions have been highly variable and often with small effect sizes [7,9,15-17]. This may be due to the fact that most are intended to inform (eg, they are educational or diagnostic) and are not necessarily designed to elicit behavior change. In addition, two recent reviews have emphasized the importance of including a behavioral change component when using technology-based programs to influence weight loss [8,15].

The importance of behavioral change in face-to-face interventions is well established [18-20]. Specifically, behavioral-based programs aim to change habits though the use of self-monitoring techniques, cognitive restructuring, and social support [3,12]. The premise is that changing daily habits and behaviors (eg, diet and physical activity) is necessary to induce weight loss and longer-term maintenance [21-23]. To induce change, the design of interventions may be informed by an understanding of effective learning environments, including providing plenty of social interaction; making the environment intrinsically rewarding; and ensuring engagement through active learning [24]. For example, social support helps create a learning environment, which provides a context for positive behaviors and the development of self-efficacy [25]. Furthermore, weight loss interventions may benefit from advances in understanding how people learn and the reality that participatory media (eg, mobile and online games, and social media) are increasingly important components of the lives of many people.

Given all this, behavioral change may be advanced by integrating an increased understanding of behavioral learning with innovative technologies. However, despite helping to overcome face-to-face challenges, current Web-based interventions are typically missing some important aspect of human (social) interaction *and* often fall well short of delivering truly engaging experiences—both are key elements of effective learning environments. In contrast to Web technologies, 3-dimensional (3D) virtual reality technologies more directly support engagement and active learning [26]. Based on education and behavioral theories, this should lead to improved outcomes. Early evidence suggests, for example, that simulated experiences via virtual reality can help develop self-efficacy [27]. To date, however, research has largely involved costly and sophisticated virtual reality (eg, head-mounted displays, cave automatic virtual environments) [28]. Interest has recently expanded to 3D virtual worlds (virtual worlds) [7,29-32].

Description of Virtual Worlds

Virtual worlds share many of the strengths of virtual reality technologies, particularly the rendering of 3D spaces. However, they are more accessible to users (via an Internet-connected personal computer) and thus may offer a way to extend the reach of programs to obese and overweight individuals. Virtual worlds also possess affordances that differentiate them (to varying degrees) from virtual reality and Web technologies. Specifically, virtual worlds are *persistent* (continuing to exist even when users are not logged in) and, as multiuser spaces, they support *social interactivity* [7,33]. Since individuals can be influenced by others, social interactivity may promote positive behaviors via emotional (eg, encouragement) or informational (eg, advice or knowledge) support [25,34,35]. In principle, virtual-world users experience a sense of *presence*—the feeling of being there in the virtual place rather than in the physical space where one's body is really located [36]. The notion of *being* there is also enhanced by the possibility of *doing* there [37], a necessary condition for active learning. Participants act within the virtual world through the use of an *avatar*—a digital representation of self, typically customizable so a user can portray an actual or desired self-image [38].

Importantly, the ability to customize one's avatar self and use it to interact with others allows for a new way to assert one's embodied subjectivity. Early research suggests that it is avatar identification that matters so significantly for identity and behavior modification. The effect of one's experience in a virtual world on one's being in the physical world has been termed the Proteus effect [39]. This effect results in users transferring expectations or understandings of their avatar's behavior to their own real-world behavior [40]. This phenomenon may have similarities to how behavior is learned from others in the real world who act as role models, as posited by social cognitive theory [23]; may aid self-efficacy; and, in turn, may lead to improvement in health behaviors [41,42]. To illustrate, Fox et al [40] conducted a study in which participants were exposed (via head-mounted displays) to either a virtual representation of self running on a treadmill, a virtual representation of an other running, or a virtual representation of self loitering. Follow-up surveys 24 hours after the experiment revealed that participants in the virtual representation of self running condition reported significantly higher levels of physical activity than those in the other two conditions.

Based on these ideas, the purpose of the present study was to examine the effectiveness of a virtual-world-based weight loss intervention program in achieving weight loss, behavioral change, and self-efficacy.

Methods

Context of Study

Club One Island is an interactive weight loss community delivered via Linden Lab's Second Life [43], an online virtual world. An island is essentially virtual land that can be built on and customized. Club One Island provides participants with a professional team, education, and specialized tools to help them overcome individual barriers to weight loss. Access to educational components and specialized tools is available 24 hours a day, 7 days a week, along with virtual world and email access to instructors. Club One Island can be described relative to two main, interrelated elements: *island* and *program design*. The design of both was informed by social cognitive theory, concepts from gaming, prior research on weight loss interventions; and emerging research on avatar identification.

First, Club One Island was designed to be visually and functionally engaging. It offers highly interactive 3D spaces (eg, a restaurant, a Mini-Mart convenience store, and an encouragement room), creative educational tools (eg, a nutritional jeopardy game and a fire pit illustrating how the body uses food as fuel), over 30 movement activities (eg, bikes and bike paths, surfing, exercise balls, lap swim, basketball, weight lifting, yoga, dancing, and rock climbing), and numerous healthy habits tools (eg, tracking charts). All elements are intended to engage participants in social networking, play, and learning. For example, the restaurant area was designed for use as a practice area for nutrition planning. It has an interactive menu that displays a full selection of items that a participant might encounter during a restaurant outing. The menu works on a stoplight (red, yellow, green) model, and participants are asked to choose what they believe are the healthiest choices. The menu responds to their selection with the color appropriate to their selection and an explanation as to why this menu item was ranked at that particular color, as well as ways to make the item healthier. Topics discussed are hidden calories, eating out, and portion control. The restaurant area also includes an ice cream counter, dining tables, a dessert bar, and a bar area. In addition, the social support classes were designed to provide an environment within which the participant could reflect on the past week's activities and social bonds. To this end, social support classes took place in the pool, at the camp fire on the beach at sunset, and while performing yoga poses. Overall, Club One Island was designed to provide an environment that closely mirrors the physical world. By setting up learning situations that incorporate practicing new behaviors (eg, throwing away 3D food, addressing the "food pushers" and nonsupportive people in their lives, and doing any physical activity in public), Club One Island is intended to help participants overcome their fears related to weight loss. Figure 1 and Figure 2 show examples of available activities and participants using the restaurant area.

Second, the weight loss program itself was designed to move participants from a diet and exercise cycle of weight loss and gain to a view that they are on a *healthy life path* that does not have a stop and end date but is maintainable for the rest of their lives. The 12-week program was delivered to cohorts of 15–20 participants for a total of 48 instructor hours. Each week, four 1-hour classes (*Nutrition*, *Movement*, *Healthy Habits*, and *Support Group*) were led by certified fitness, nutrition, and support professionals. Each week addressed a common theme (eg, emotion as related to eating) across all 4 classes.

In the program, participants chose how their avatar looked (actual or desired) and made modifications over time, as wanted. In addition, the Nutrition, Movement, and Healthy Habits classes were all designed in such a way that participants were always moving. For example, participants (ie, their avatars) in Movement sessions engaged in 1 to 4 different activities, ranging from roller skating to surf boarding to riding bicycles, to swimming and more. Similarly, during the Nutrition and Healthy Habits classes, participants spent 90% of their class time in (virtual) standing and moving positions. When proceeding from one activity to the next, they would run, bike, roller skate, etc, to get there. These program design elements were intended to encourage avatar identification, leading to the transfer of virtual behaviors to the physical world. Multimedia Appendix 1 provides an overview of Club One Island's virtual spaces and participant activities.

Figure 1. Examples of program activities in Club One Island.

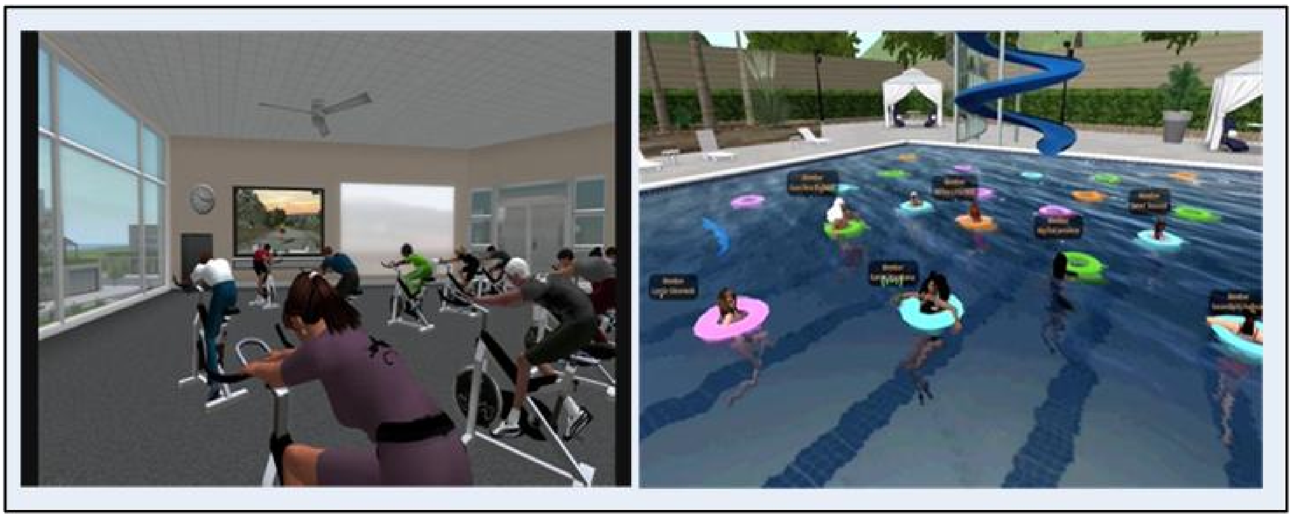


Figure 2. Design of the Club One Island virtual weight loss community.

Participants

We conducted a study of Club One Island’s weight loss program, comparing it with a face-to-face program similar in structure and content offered in a commercial setting (owned by Club One Inc., San Francisco, CA, USA). Specifically, this program included instructor-led weekly educational sessions on nutrition, movement, and habit change, as well as a social support group meeting. Program participants were also able to use club equipment and facilities during normal business hours.

Club One Island’s virtual-world program involved 38 participants recruited via print and online media. Participants had to be over the age of 18 years, with a body mass index (BMI) of 25 kg/m^2 or greater, and have access to an Internet-connected computer. They were told they were helping to assess a new program and, as such, it was free. Of the 38 enrollees, 1 was a current Club One member, 9 were members of competing clubs, and 28 did not belong to any health club. Since the comparison face-to-face group had to attend a real facility, we recruited a convenience sample (via email and newsletters) from Club One’s member base. Enrollees had similar age and BMI requirements, leading to a 24-person face-to-face cohort. Across all enrollees, they had on average previously tried 2 other weight loss programs. Prior to the start of the virtual-world program, participants received technical training and support (eg, computer setup and navigating the island).

Measures

Data were collected by trained professionals at a Club One facility and provided to the researcher in de-identified form. Both objective and self-report measures were captured at baseline (preprogram) and within 1 week of program completion. Objective data (height, weight, and BMI) were recorded using standard techniques. We calculated the percentage change in baseline weight; a reduction of 5% or more of baseline weight results in clinically significant health benefits [44]. The baseline and postsurveys captured data regarding health-related behaviors (ie, general health, sleep, and degree of moderate and vigorous physical activity) and nutrition and eating habits (ie, frequency of breakfast, and number of servings of fruit and vegetables per day). These items were adapted from the US Centers for Disease Control and Prevention’s Behavioral Risk Factor Surveillance System Survey [45]. In addition, self-efficacy with regard to both physical activity and weight management were captured. Items were drawn from the Physical Activity Confidence Scale [46] and the Weight Efficacy Lifestyle Questionnaire (WEL) [47]. We used the global WEL score for this study, where high scores represent more positive beliefs toward the completion of weight management actions [48]. In addition to demographic data, the presurvey also asked questions about attitudes toward exercising at a health club with items adapted from Miller and Miller [49]. Virtual-world participants were also asked about prior experience with Second Life and, on the postsurvey, some questions about their perceptions of the environment. Lastly, virtual-world participants were given an opportunity to offer open-ended responses about the experience. Regarding our surveys, all construct reliabilities exceeded 0.70. Specific survey items and factor analytic results are available, upon request. This study was approved by Indiana University Institutional Review Board.

Statistical Analysis

Descriptive statistics were used to describe baseline characteristics of groups. The analyses tested for group (virtual world and face-to-face) differences in baseline characteristics using a multivariate analysis of variance. We used repeated-measures multivariate analysis to detect group differences, within-participant differences (time), and interactions (group × time) in measured outcomes from baseline to postintervention. Paired *t* tests were used to identify where significant differences occurred pre- to postintervention. A 2-sample *t* test between proportions was performed to determine whether there was a significant difference between groups with respect to the percentage of weight lost. We considered differences to be significant at the $P < .05$ level and used SPSS version 18 (IBM Corporation, Somers, NY, USA) for analysis.

Results

Demographics

Full data sets were available for 33 virtual world and 21 face-to-face participants. Virtual-world participants (n = 5) dropped out in the first 2 weeks due to technical difficulties and personal reasons. Face-to-face participants (n = 3) dropped out midprogram, citing disinterest and personal reasons. Demographically, the virtual-world group was 76% female (25/33), with a mean age of 46.3 (SD 9.6) years; 73% (24/33) held college or advanced degrees; and 76% (25/33) had annual incomes exceeding US $75,000. All 33 reported they were novice users of Second Life (0–3 months). The face-to-face group was 95% (20/21) female, with a mean age of 37.5 (SD 10.6) years; 90% (19/21) held college or advanced degrees; and 71% (15/21) reported incomes over US $75,000. Across these data, there were no significant differences. However, regarding attitudes toward exercising at a real club, virtual-world participants reported a statistically significant ($P = .03$) higher negative attitude (mean score 3.35, SD 1.13, on a scale of 1 to 7, where 1 = strongly disagree; higher scores represent more negative attitudes) than face-to-face participants (2.69, SD 0.97).

Change in Primary Outcome

The mean BMI of the overall sample was 32.0 (SD 6.05) kg/m^2 (virtual world: 33.13, SD 6.13; face-to-face: 30.21, SD 5.62). No significant baseline differences were noted between groups for weight, BMI, general health, fruit and vegetable consumption, breakfast frequency, and physical activity self-efficacy. However, significant baseline differences were observed for moderate physical activity ($P = 0.02$), vigorous physical activity ($P = .001$), sleep ($P = .02$), and WEL ($P = .04$).

Table 1 summarizes baseline and postintervention values for objective measures. No significant group × time interactions were detected; however, both groups lost a significant amount of weight (virtual world: 3.9 kg, $P < .001$; face-to-face: 2.8 kg, $P = .002$). Compared with baseline, the virtual-world group lost an average of 4.3% (range –17.3% to 3.3%), with 33% (11/33) of the participants losing a clinically significant (≥5%) amount of weight. The face-to-face group lost an average of 3.0% (range –11.0% to 2.7%), with 29% (6/21) losing a clinically significant amount. Furthermore, 15.2% (5/33) of the virtual-world and 14.3% (3/21) of the face-to-face groups lost 7% or more of their body weight. Significant differences were not detected between groups for the percentage of weight lost ($P = .34$) or the percentage of participants losing 5% or more of their baseline body weight ($P = .39$).

Table 1. Changes in body weight from baseline to postintervention.

Outcome variable, mean (SD)	Virtual world		Face-to-face		*P* value	
	Pre	Post	Pre	Post	Group × time interaction	Time main effect
Weight (kg)	92.1 (23.2)	88.2 (21.6)	83.9 (16.1)	81.1 (14.5)	.29	.001
Body mass index (kg/m^2)	33.13 (6.13)	31.71 (5.51)	30.21 (5.6)	29.13 (4.96)	.39	.001

Change in Secondary Outcomes

Table 2 summarizes baseline and postintervention values for self-reported measures of behavioral change and self-efficacy. The group × time interaction was significant for pre- to postintervention general health, moderate and vigorous physical activity, physical activity self-efficacy, fruit and vegetable consumption, and WEL. Post hoc paired *t* tests indicated significant improvements across *all* of the variables for the virtual-world group, while the face-to-face group had nonsignificant improvements in self-efficacy for physical activity and WEL, as well as for fruit and vegetable consumption. The face-to-face group reported decreases in moderate and vigorous physical activity that were nonsignificant. Lastly, we noted a significant time effect for general health and breakfast consumption. Paired-samples *t* tests indicated a significant improvement in perceptions of general health ($P < .001$) and an increase in the number of days the virtual-world group ate breakfast ($P = .003$); in contrast, there were no significant changes in these variables for the face-to-face group.

Table 2. Changes in health behaviors and self-efficacy from baseline to postintervention.

Outcome variable, mean (SD)	Virtual-world group		Face-to-face group		*P* value	
	Pre	Post	Pre	Post	Group × time interaction	Time main effect
Health (1 = poor to 5 = excellent)	3.0 (1.1)	3.5 (1.0)	3.2 (0.7)	3.4 (0.9)	.12	.001
Sleep (1 = <6 hours to 5 = ≥9 hours)	2.82 (0.88)	2.87 (0.78)	3.38 (0.81)	3.23 (0.89)	.30	.62
Moderate PA[a] (no. days/week)	2.8 (2.2)	4.2 (1.7)	4.2 (2.1)	3.9 (1.9)	.006	.08
Vigorous PA (no. days/week)	1.4 (1.8)	2.5 (2.0)	3.2 (1.9)	3.0 (1.8)	.008	.04
PA self-efficacy (1 = not to 5 = extremely confident)	2.92 (0.80)	3.42 (0.80)	3.39 (0.92)	3.41 (0.80)	.04	.02
Fruit and vegetables (1 = 0 to 5 = >5/day)	2.79 (0.93)	3.37 (0.89)	2.81 (0.81)	2.90 (0.77)	.007	<.001
Breakfast (no. days/week)	6.15 (1.48)	6.78 (0.60)	6.52 (0.98)	6.67 (0.66)	.13	.02
Weight efficacy (scale 19 to 133)	80.70 (20.1)	108.7 (16.7)	92.0 (17.2)	97.3 (16.7)	<.001	<.001

[a] Physical activity.

Discussion

The purpose of this study was to explore the potential of a virtual-world-based weight loss program relative to weight loss, behavioral change, and self-efficacy. We used a face-to-face program that was similar in structure and content for comparative purposes. We found both groups significantly benefitted from their respective interventions in terms of weight loss and BMI reduction (see Table 1). Importantly, the average weight loss results of virtual-world participants (3.9 kg) compares favorably with studies of other short-term (12- to 16-week) programs [13,14,48,50-52]. In addition, the percentage weight lost and percentage of virtual-world participants who lost 5% or more of their body weight is also consistent with or greater than those in previous investigations [8,13,14,52]. Interestingly, virtual-world participants exhibited significant improvements in nearly all indicators (except sleep) of behavioral change, while the face-to-face participants showed *no* marked improvements in any indicator. Additionally, virtual-world participants' self-efficacy regarding their abilities to engage in physical activity and to resist eating (WEL) both increased significantly (see Table 2), while again the face-to-face group exhbited no significant changes. In fact, the change in the WEL score (+28) for Club One Island participants was higher than has been noted in previous face-to-face and Web-based intervention studies [48,53].

As we consider these results, it is important to emphasize the tight integration of the Club One Island program and island designs. The design of both aspects allowed for motivational reinforcement, practice-oriented instruction and active learning, and social support, collectively serving to create a learning environment that fostered desired outcomes. In contrast, physical environmental limitations experienced by face-to-face participants seemed to have been less supportive of behavioral change. For example, the dynamic 3D spaces allowed virtual-world participants opportunities to test both positive and negative behaviors like navigating complex food situations, such as at a party. Our results suggest that behavioral change and increased self-efficacy were likely influenced by virtual-world participants observing their avatars engage in healthy behaviors. Virtual-world participants overwhelmingly created avatars that reflected real depictions of themselves. As they lost weight in the real world, they made changes to their avatar's appearance to reflect this. Through this environment, participants were able—many for the first time in their lives—to have a positive experience related to physical activity, and to test both positive and negative behaviors such as navigating complex food situations. Sample comments by virtual-world participants provide evidence of avatar identification and the transfer of virtual-world behaviors to the real world, as well as the development of self-efficacy and the role of social support (for additional comments from participant interviews, see Multimedia Appendix 2):

> *During the workday, I remember my avatar sipping from a (3D) water bottle and I'd grab my own (real) bottle. Having the bottle and drinking animation has led directly to a change in my behavior.*
>
> *Usually when I'm on the treadmill at the gym, I walk for 5 minutes and run 1 minute, which is really challenging. This last time, I pictured my avatar running and I felt like my avatar and it made me feel stonger. I ran for 2 minutes easily.*
>
> *I'm integrating skills I didn't know how to use. I usually read nutritional labels but didn't really know what to do with the information. Now I have more confidence and know how to maintain my weight loss.*
>
> *This was the best part of any class, when the other members talked about their experiences. It was just good to know other people out there are struggling with the same issues.*

Overall, our results offer positive early evidence that a virtual-world-based weight loss program can be as effective as a face-to-face one relative to biometric changes. Our results also suggest that a virtual world may be a *more effective* platform to influence meaningful behavioral changes and improve self-efficacy. Nearly 70% (23/33) of participants agreed or strongly agreed that the virtual-world program worked better than other things they had tried in the past. The fact that all were novice virtual-world users suggests that deep technical skills are not a prerequisite for success.

While we do not direct compare them here, our results are also promising relative to those of prior studies of Web-based interventions. One indicator is that, despite the weekly time commitment, the dropout rate within the virtual-world program was low (15%) compared with other Web-based programs [54,55]. This suggests a greater level of engagement and less boredom with the virtual-world program, both factors that have led to attrition in Web-based programs. More specifically, a review by Lodama et al [56] suggests that the lack of face-to-face interaction and static pages are two major limitations of Web-based programs. Harvey-Berino et al [4] also found significant differences in perceived social support when comparing Web-based versus face-to-face programs, partially explaining lower and significantly different total weight lost. They concluded that “strategies to facilitate and enhance a sense of group cohesion online” are needed (p. 127). In response, recent research has focused on enhancing Web-based interventions by incorporating elements (eg individualized programs, chat sessions, forums, and more dynamic material) to increase engagement. However, to date, results of such efforts have not led to significant improvements in weight loss [54,55]. Since engagement results from a synergistic interaction between motivation and active learning [57] (p. 8), the very nature of virtual worlds may offer significant advantages over earlier technologies. As illustrated in this study, virtual worlds are inherently dynamic, and they offer significant opportunities for social interactivity and support. Importantly, through avatar self-embodiment, virtual worlds allow for a functional equivalent to face-to-face interactions, thus helping to address the above-noted shortcomings of Web-based interventions.

Limitations and Directions

The sample in this study provided both strengths and weaknesses. Given our research focus and the interests of our industry partner, the demographics of our sample were appropriate. However, this is a possibility of a self-selection bias, as participants in both groups volunteered to participant in the study in response to a recruitment advertisement. Thus, their motivation and other possible variables may not be representative of the general population. Future research should involve a broader recruitment strategy and controlled assignment to an intervention. While our study followed past practice by using a face-to-face group for comparative purposes, the next step would also be to compare a virtual-world versus a Web-based intervention. This may allow, for example, deeper analysis of the relationship between avatar identification and outcomes.

We also focused our attention largely on outcomes to compare the effectiveness of the virtual-world and face-to-face weight loss programs. While we found positive benefits immediately following program completion, a longitudinal assessment is needed to determine whether behavioral changes are sustained over time for virtual-world participants. We have some evidence that avatar identification, virtual- to real-world behavior transfer, and social support components relate to outcomes. Yet, as with earlier studies [8], how these elements actually work together or whether relationships to outcomes are correlational or causal is not yet fully understood. Similarly, the weight loss program itself involved multiple educational components along with social support. The relative importance of each to the observed outcomes warrants further exploration. In addition, research is needed regarding the relationship between individual characteristics (eg, obesity classication, learning style, and immersive tendencies) and outcomes. For example, the degree to which one is receptive to immersive experiences during media exposure (such as a virtual world) may influence outcomes. Similarly, as a visually oriented environment, a virtual world may be more attractive to individuals with a better spatial sense and who prefer learning through images and objects, rather than text.

Lastly, from a human–computer interaction design perspective, a deeper understanding of how design elements (eg, environmental or avatar realism) interact with individual characteristics (eg, a preferred representational system) is essential to advancing the adoption of virtual worlds. Moreover, while a virtual presence is considered central to the utility of a virtual world, attempts to design spaces to maximize presence are premature without solid evidence that it relates to desired outcomes. In this study, we also explored the efficacy of a specific virtual-world platform (Second Life). Further work should examine alternative platforms possessing varied sociotechnical capabilities.

In sum, further empirical testing in both controlled experiments and field studies would help to develop a richer understanding of the value of virtual-world-based interventions to address obesity and overcome the challenges of existing approaches.

Conclusion

The rising rates, high prevalence, and adverse consequences of obesity call for the development and testing of innovative approaches that address the cited barriers and bring needed help to those most affected. While more research is needed into their use in medical and health contexts [31], as shown here, virtual worlds may offer an environment within which participants can engage in experiential learning and simulate what-if scenarios without serious repercussions. The affordances of simulated 3D experiences, anonymity, embodied personal representation in the form of an avatar, and rich social interaction constitute the potential for a virtual world to have a strong effect as a weight loss intervention platform. While use and research is in its infancy, there are encouraging signs regarding application to a variety of medical, health, educational, and other contexts. We hope that the ideas and findings offered here contribute to this growing area.

Conflicts of Interest

J. Johnston and A. Massey have no conflicts of interest to declare. C. DeVaneaux is an industry partner who was blind to the analysis and the results prior to the completion of the paper.

Multimedia Appendix 1

Club One Island Virtual Spaces and Activities.

[FLV File, 24MB - jmir_v14i5e120_app1.flv]

Multimedia Appendix 2

Participant Interview.

[FLV File, 29MB - jmir_v14i5e120_app2.flv]

References

1. Flegal KM, Carroll MD, Kit BK, Ogden CL. Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999-2010. JAMA 2012 Feb 1;307(5):491-497. [doi: 10.1001/jama.2012.39] [Medline: 22253363]
2. Klein S, Burke LE, Bray GA, Blair S, Allison DB, Pi-Sunyer X, American Heart Association Council on Nutrition, Physical Activity, and Metabolism. Clinical implications of obesity with specific focus on cardiovascular disease: a statement for professionals from the American Heart Association Council on Nutrition, Physical Activity, and Metabolism: endorsed by the American College of Cardiology Foundation. Circulation 2004 Nov 2;110(18):2952-2967 [FREE Full text] [doi: 10.1161/01.CIR.0000145546.97738.1E] [Medline: 15509809]
3. Wadden TA, Crerand CE, Brock J. Behavioral treatment of obesity. Psychiatr Clin North Am 2005 Mar;28(1):151-70, ix. [doi: 10.1016/j.psc.2004.09.008] [Medline: 15733617]
4. Harvey-Berino J, West D, Krukowski R, Prewitt E, VanBiervliet A, Ashikaga T, et al. Internet delivered behavioral obesity treatment. Prev Med 2010 Aug;51(2):123-128. [doi: 10.1016/j.ypmed.2010.04.018] [Medline: 20478333]
5. Strecher V. Internet methods for delivering behavioral and health-related interventions (eHealth). Annu Rev Clin Psychol 2007;3:53-76. [doi: 10.1146/annurev.clinpsy.3.022806.091428] [Medline: 17716048]
6. Wing RR, Crane MM, Thomas JG, Kumar R, Weinberg B. Improving weight loss outcomes of community interventions by incorporating behavioral strategies. Am J Public Health 2010 Dec;100(12):2513-2519. [doi: 10.2105/AJPH.2009.183616] [Medline: 20966375]
7. Morie JF, Chance E. Extending the reach of health care for obesity and diabetes using virtual worlds. J Diabetes Sci Technol 2011 Mar;5(2):272-276 [FREE Full text] [Medline: 21527093]
8. Neve M, Morgan PJ, Jones PR, Collins CE. Effectiveness of web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: a systematic review with meta-analysis. Obes Rev 2010 Apr;11(4):306-321. [doi: 10.1111/j.1467-789X.2009.00646.x] [Medline: 19754633]
9. Carter-Edwards L, Bastian LA, Schultz M, Amamoo MA, Østbye T. An internet-based weight loss intervention initiated by a newspaper. Prev Chronic Dis 2009 Jul;6(3):A101 [FREE Full text] [Medline: 19527573]
10. Internet World Stats. Miniwatts Marketing Group. 2012 Jun 19. Internet Users in the World: Distribution by World Regions URL: http://www.internetworldstats.com/stats.htm [accessed 2012-06-21] [WebCite Cache ID 68bH9Lw8W]
11. Trend data (adults). Washington, DC: Pew Internet & American Life Project; 2012 Feb. URL: http://www.pewinternet.org/Trend-Data-(Adults)/Online-Activites-Total.aspx [accessed 2012-06-22] [WebCite Cache ID 68c0eVbiV]
12. Gold BC, Burke S, Pintauro S, Buzzell P, Harvey-Berino J. Weight loss on the web: A pilot study comparing a structured behavioral intervention to a commercial program. Obesity (Silver Spring) 2007 Jan;15(1):155-164 [FREE Full text] [doi: 10.1038/oby.2007.520] [Medline: 17228043]
13. Polzien KM, Jakicic JM, Tate DF, Otto AD. The efficacy of a technology-based system in a short-term behavioral weight loss intervention. Obesity (Silver Spring) 2007 Apr;15(4):825-830 [FREE Full text] [doi: 10.1038/oby.2007.584] [Medline: 17426316]
14. Tate DF, Jackvony EH, Wing RR. A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an Internet weight loss program. Arch Intern Med 2006;166(15):1620-1625. [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]
15. 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]
16. Khaylis A, Yiaslas T, Bergstrom J, Gore-Felton C. A review of efficacious technology-based weight-loss interventions: five key components. Telemed J E Health 2010 Nov;16(9):931-938. [doi: 10.1089/tmj.2010.0065] [Medline: 21091286]

17. Norman GJ, Zabinski MF, Adams MA, Rosenberg DE, Yaroch AL, Atienza AA. A review of eHealth interventions for physical activity and dietary behavior change. Am J Prev Med 2007 Oct;33(4):336-345. [doi: 10.1016/j.amepre.2007.05.007] [Medline: 17888860]
18. Berkel LA, Poston WS, Reeves RS, Foreyt JP. Behavioral interventions for obesity. J Am Diet Assoc 2005 May;105(5 Suppl 1):S35-S43. [doi: 10.1016/j.jada.2005.02.031] [Medline: 15867894]
19. Linde JA, Rothman AJ, Baldwin AS, Jeffery RW. The impact of self-efficacy on behavior change and weight change among overweight participants in a weight loss trial. Health Psychol 2006 May;25(3):282-291. [doi: 10.1037/0278-6133.25.3.282] [Medline: 16719599]
20. Warziski MT, Sereika SM, Styn MA, Music E, Burke LE. Changes in self-efficacy and dietary adherence: the impact on weight loss in the PREFER study. J Behav Med 2008 Feb;31(1):81-92. [doi: 10.1007/s10865-007-9135-2] [Medline: 17963038]
21. Brownell KD, Kramer FM. Behavioral management of obesity. Med Clin North Am 1989 Jan;73(1):185-201. [Medline: 2643003]
22. Bandura A. Social cognitive theory of mass communication. Media Psychol 2001;3:265-299. [doi: 10.1207/S1532785XMEP0303_03]
23. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice-Hall; 1986.
24. Straumanis J. The science of learning: breaking news. J Diabetes Sci Technol 2011 Mar;5(2):251-255 [FREE Full text] [Medline: 21527090]
25. Vrazel J, Saunders RP, Wilcox S. An overview and proposed framework of social-environmental influences on the physical-activity behavior of women. Am J Health Promot 2008;23(1):2-12. [doi: 10.4278/ajhp.06070999] [Medline: 18785368]
26. Hansen MM. Versatile, immersive, creative and dynamic virtual 3-D healthcare learning environments: a review of the literature. J Med Internet Res 2008;10(3):e26 [FREE Full text] [doi: 10.2196/jmir.1051] [Medline: 18762473]
27. Spires HA, Lee JK, Turner KA, Johnson J. Having our say: middle grade student perspectives on school, technologies, and academic engagement. J Res Technol Ed 2008;40:497-515.
28. Persky S. Application of virtual reality methods to obesity prevention and management research. J Diabetes Sci Technol 2011 Mar;5(2):333-339 [FREE Full text] [Medline: 21527102]
29. Riva G. The key to unlocking the virtual body: virtual reality in the treatment of obesity and eating disorders. J Diabetes Sci Technol 2011 Mar;5(2):283-292 [FREE Full text] [Medline: 21527095]
30. Siddiqi S, Mama SK, Lee RE. Developing obesity prevention intervention in networked virtual environments: the international health challenge in Second Life. J Virtual Worlds 2010;3(3):1-26.
31. Boulos MN, Hetherington L, Wheeler S. Second Life: an overview of the potential of 3-D virtual worlds in medical and health education. Health Info Libr J 2007 Dec;24(4):233-245. [doi: 10.1111/j.1471-1842.2007.00733.x] [Medline: 18005298]
32. Ershow AG, Peterson CM, Riley WT, Rizzo AS, Wansink B. Virtual reality technologies for research and education in obesity and diabetes: research needs and opportunities. J Diabetes Sci Technol 2011 Mar;5(2):212-224 [FREE Full text] [Medline: 21527084]
33. Montoya MM, Massey AP, Lockwood NS. 3D collaborative virtual environments: exploring the link between collaborative behaviors and team performance. Decis Sci 2011;42:451-476. [doi: 10.1111/j.1540-5915.2011.00318.x]
34. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007 Jul 26;357(4):370-379 [FREE Full text] [doi: 10.1056/NEJMsa066082] [Medline: 17652652]
35. Beard L, Wilson K, Morra D, Keelan J. A survey of health-related activities on second life. J Med Internet Res 2009;11(2):e17 [FREE Full text] [doi: 10.2196/jmir.1192] [Medline: 19632971]
36. Draper JV, Kaber DB, Usher JM. Telepresence. Hum Factors 1998 Sep;40(3):354-375. [Medline: 9849099]
37. Zahorik P, Jenison RL. Presence as being-in-the-world. Presence 1998;7:78-89. [doi: 10.1162/105474698565541]
38. Bartle RA. Designing Virtual Sorlds. Indianapolis, IN: New Riders; 2004.
39. Yee N, Bailenson J. The Proteus effect: the effect of transformed self-representation on behavior. Hum Commun Res 2007;33:271-290.
40. Fox J, Bailenson J, Binney J. Virtual experiences, physical behaviors: the effect of presence on imitation of an eating avatar. Presence 2009;18:294-303. [doi: 10.1162/pres.18.4.294]
41. Rogers RW. Cognitivephysiological processes in fear appealsattitude change: a revised theory of protection motivation. In: Cacioppo J, Petty R. editors. Social Psychophysiology. New York, NY: Guilford Press; 1983.
42. Rosenstock IM. Historical origins of the health belief model. Health Ed Monogr 1974;2(4):328-335.
43. Club One Island. 2011. Club One Island URL: http://www.cluboneisland.com/ [accessed 2012-06-22] [WebCite Cache ID 68c1mYlPo]
44. Goldstein DJ. Beneficial health effects of modest weight loss. Int J Obes Relat Metab Disord 1992 Jun;16(6):397-415. [Medline: 1322866]

45. Office of Surveillance, Epidemiology, and Laboratory Services. US Department of Health and Human Services, Centers for Disease Control and Prevention. 2009. Behavioral Risk Factor Surveillance System Survey Data URL: http://www.cdc.gov/brfss/ [accessed 2012-09-03] [WebCite Cache ID 6APGHXokY]
46. Marcus BH, Selby VC, Niaura RS, Rossi JS. Self-efficacy and the stages of exercise behavior change. Res Q Exerc Sport 1992 Mar;63(1):60-66. [Medline: 1574662]
47. Clark MM, Abrams DB, Niaura RS, Eaton CA, Rossi JS. Self-efficacy in weight management. J Consult Clin Psychol 1991 Oct;59(5):739-744. [Medline: 1955608]
48. Palmeira AL, Teixeira PJ, Branco TL. 2007. Predicting short-term weight loss using four leading health behavior change theories. Int J Behav Nutr Phys Act 2007;4:14. [Medline: 17448248]
49. Miller WC, Miller TA. Attitudes of overweight and normal weight adults regarding exercise at a health club. J Nutr Educ Behav 2010;42(1):2-9. [doi: 10.1016/j.jneb.2008.08.005] [Medline: 19914871]
50. Webber KH, Gabriele JM, Tate DF, Dignan MB. The effect of a motivational intervention on weight loss is moderated by level of baseline controlled motivation. Int J Behav Nutr Phys Act 2010;7:4 [FREE Full text] [doi: 10.1186/1479-5868-7-4] [Medline: 20157441]
51. Tate DF, Wing RR, Winett RA. Using Internet technology to deliver a behavioral weight loss program. JAMA 2001 Mar 7;285(9):1172-1177. [Medline: 11231746]
52. Morgan PJ, Collins CE, Plotnikoff RC, Cook AT, Berthon B, Mitchell S, et al. Efficacy of a workplace-based weight loss program for overweight male shift workers: the Workplace POWER (Preventing Obesity Without Eating like a Rabbit) randomized controlled trial. Prev Med 2011 May;52(5):317-325. [doi: 10.1016/j.ypmed.2011.01.031] [Medline: 21300083]
53. Annesi JJ. Behaviorally supported exercise predicts weight loss in obese adults through improvements in mood, self-efficacy, and self-regulation, rather than by caloric expenditure. Perm J 2011;15(1):23-27. [Medline: 21505614]
54. van Genugten L, van Empelen P, Boon B, Borsboom G, Visscher T, Oenema A. Results from an online computer-tailored weight management intervention for overweight adults: randomized controlled trial. J Med Internet Res 2012;14(2):e44 [FREE Full text] [doi: 10.2196/jmir.1901] [Medline: 22417813]
55. Collins CE, Morgan PJ, Jones P, Fletcher K, Martin J, Aguiar EJ, et al. A 12-week commercial web-based weight-loss program for overweight and obese adults: randomized controlled trial comparing basic versus enhanced features. J Med Internet Res 2012;14(2):e57 [FREE Full text] [doi: 10.2196/jmir.1980] [Medline: 22555246]
56. Kodama S, Saito K, Tanaka S, Horikawa C, Fujiwara K, Hirasawa R, et al. Effect of Web-based lifestyle modification on weight control: a meta-analysis. Int J Obes (Lond) 2012 May;36(5):675-685. [doi: 10.1038/ijo.2011.121] [Medline: 21694698]
57. Barkley EF. Student Engagement Techniques: A Handbook for Faculty. San Francisco, CA: Jossey-Bass; 2010.

Abbreviations

3D: 3-dimensional
BMI: body mass index
WEL: Weight Efficacy Lifestyle Questionnaire

Edited by G Eysenbach; submitted 04.07.12; peer-reviewed by G Gay, D Taylor; comments to author 25.07.12; revised version received 10.08.12; accepted 10.08.12; published 20.09.12.

Please cite as:
Johnston JD, Massey AP, DeVaneaux CA
Innovation in Weight Loss Programs: A 3-Dimensional Virtual-World Approach
J Med Internet Res 2012;14(5):e120
URL: http://www.jmir.org/2012/5/e120/
doi:10.2196/jmir.2254
PMID:22995535

Original Paper

Patients' Ratings of Family Physician Practices on the Internet: Usage and Associations With Conventional Measures of Quality in the English National Health Service

Felix Greaves[1], BM, BCh; Utz J Pape[1], PhD; Henry Lee[2], MB, ChB; Dianna M Smith[1], PhD; Ara Darzi[2], MD; Azeem Majeed[1], MD; Christopher Millett[1], PhD

[1]Department of Primary Care and Public Health, Imperial College London, London, United Kingdom

[2]Centre for Health Policy, Institute of Global Health Innovation, Imperial College London, London, United Kingdom

Corresponding Author:
Felix Greaves, BM, BCh
Department of Primary Care and Public Health
Imperial College London
Reynolds Building
Charing Cross Hospital
London, W6 8RP
United Kingdom
Phone: 44 7866551172
Fax: 44 2075940854
Email: felixgreaves@gmail.com

Abstract

Background: Patients are increasingly rating their family physicians on the Internet in the same way as they might rate a hotel on TripAdvisor or a seller on eBay, despite physicians' concerns about this process.

Objective: This study aims to examine the usage of NHS Choices, a government website that encourages patients to rate the quality of family practices in England, and associations between web-based patient ratings and conventional measures of patient experience and clinical quality in primary care.

Methods: We obtained all (16,952) ratings of family practices posted on NHS Choices between October 2009 and December 2010. We examined associations between patient ratings and family practice and population characteristics. Associations between ratings and survey measures of patient experience and clinical outcomes were examined.

Results: 61% of the 8089 family practices in England were rated, and 69% of ratings would recommend their family practice. Practices serving younger, less deprived, and more densely populated areas were more likely to be rated. There were moderate associations with survey measures of patient experience (Spearman ρ 0.37−0.48, $P<.001$ for all 5 variables), but only weak associations with measures of clinical process and outcome (Spearman ρ less than ±0.18, $P<.001$ for 6 of 7 variables).

Conclusion: The frequency of patients rating their family physicians on the Internet is variable in England, but the ratings are generally positive and are moderately associated with other measures of patient experience and weakly associated with clinical quality. Although potentially flawed, patient ratings on the Internet may provide an opportunity for organizational learning and, as it becomes more common, another lens to look at the quality of primary care.

(J Med Internet Res 2012;14(5):e146) doi:10.2196/jmir.2280

KEYWORDS

Patient Experience; Primary Care; Internet; Quality

Introduction

Consumers are using the Internet to rate services and products, for example when they stay in a hotel or buy a product online. This increasingly applies to health care, particularly in the US and UK. A number of websites allowing patients to rate their care have been developed by health care payers and the commercial sector, such as RateMDs or Angie's List, in an effort to increase transparency and responsiveness of health systems and to help patients choose between providers [1-4].

The value of patient rating websites has been criticized, particularly by family physicians [5], who are concerned that patient ratings are not representative, are unduly negative, may not contain information relevant to the quality of health care provided, and may harm the doctor-patient relationship [6]. However, we also know that physicians may misjudge patients' care experience and that physician performance can be improved by feedback [7], so these websites may provide useful information for patients and health care workers.

NHS Choices is a government-run website that serves as a comprehensive directory of health care providers in England and includes comparative information on performance of family practices [8]. One of its functions is to allow the public to rate the quality of care they received at their family practice, both by leaving ratings on specific aspects of care and by making comments.

Little is known about who uses patient-rating websites and how unsolicited ratings by patients relate to more conventional measures of patient experience and clinical quality. While some patients may be interested in reviewing other patients' ratings as a window into family physicians' interpersonal styles and office amenities, the amount of agreement between these ratings and conventional measures of quality such as patient experience captured in surveys or clinical process and outcome measures will be important in understanding the usefulness of these ratings in quality measurement.

This paper seeks to examine usage patterns of patients' ratings of family physicians on the Internet. We describe how frequency and nature of rating vary with practice and population characteristics, and we present comparisons between web-based ratings and conventional measures of patient experience and clinical quality.

Method

Data Sources and Measures

Online Ratings

All patient ratings of family practices posted on the NHS Choices website between October 14, 2009 (the date the function started), and December 31, 2010, were obtained from the Department of Health, aggregated to the practice level. The NHS Choices website allows patients to indicate whether they would recommend a family practice to a friend (yes/no) and rate practices on a scale for four specific domains of quality: whether they were able to get through to the practice by telephone; whether they were involved in decisions about care; whether they were able to get an appointment when they wanted one; and whether they were treated with dignity and respect by staff. Data on the individual characteristics of those leaving ratings online were not available.

Traditional Patient Surveys

Survey measures of patient experience were obtained from the national General Practice Patient Survey at the practice level. This is a large mail-based survey sent to 5.7 million patients in 2009/10, with more than 2.1 million responses received (a response rate of 37%) [9].

Population and practice characteristics were obtained from the NHS Information Centre. Practice population variables were: the proportion of patients over 65 years, the index of multiple deprivation (IMD) score (an area-based measure of socioeconomic status including components of income, employment, health, education, crime, and housing), the proportion of the practice population who reported their ethnicity as "white", the population density of the practice (measured as people per square kilometre), and the practice list size. Variables describing the practice were: whether it was a singlehanded practice, a training practice, and the type of contract it had with the health care payer. The type of contract was listed as either Personal Medical Services (PMS) or General Medical Services (GMS). The GMS contract is nationally agreed, while PMS is locally agreed.

We obtained data on clinical quality of family practices from the NHS Information Centre [10] and from NHS Comparators [11], which are both central repositories of NHS process and outcome data. We selected clinical outcome measures based on three criteria: (1) the measures are commonly used in public reporting, (2) there is known to be variation in practice performance on the measures, and (3) they represented the breadth of activities taking place in family practice. The measures chosen were: proportion of patients with diabetes receiving flu vaccinations, proportion of hypertensive patients with controlled blood pressure (systolic/diastolic less than 150/90 mm Hg), proportion of diabetic patients with controlled HbA1C (less than 7%), percentage of low-cost statin prescribing, cervical screening rate, admission rates for ambulatory care sensitive conditions, and the proportion of achieved clinical Quality and Outcomes Framework (QOF) points from available points. The QOF is the NHS's pay-for-performance scheme in primary care, which awards points for a variety of clinical activities across acute and chronic disease management and disease prevention.

Data Linkage

In order to create a complete set of family practices in England, we took the NHS's list of 8381 practices [12]. This excluded walk-in centers, out-of-hours services and prison health centers. We then excluded 165 practices with a list size less than 1000 and 127 military practices. The total number of practices used in the study is 8089 (96.5% of the total number in England).

In order to link data about the practices from NHS Choices with other practice performance and demographic data, we matched practices by postcode using Excel. We created a computer program that extracted postcodes from the NHS Choices website using Python programming language. This match was checked manually by one person to ensure appropriate linking. Where more than one practice is located at the same postcode, they were manually checked to ensure the correct practice was listed using the names of the practice physicians.

We compared the NHS Choices rating data against patient survey and outcomes data from the 2009/10 financial year. We obtained only incomplete practice demographic descriptive data for 163 practices, but these practices were still included in the analysis as we still had information from the patient survey and NHS Choices. Excluded from this analysis were 24 GP practices (of 5362 with ratings) on NHS Choices that could not be matched to the official list of practices.

Statistical Analysis

We examined associations between whether a family practice was rated and the practice and population characteristics using logistic multivariate regression. A least square regression analysis was used to examine associations between the proportion of patients willing to recommend the practice and practice and population characteristics.

In order to compare ratings with survey measures of patient experience, we compared ratings for questions on NHS Choices with results of the national General Practice Patient Survey using Spearman rank coefficient of correlation. We selected questions from the national survey that most closely matched the ratings questions on the NHS Choice website. In order to compare ratings with measures of clinical quality, we compared the proportion of people who would recommend the practice on NHS Choices with the traditional quality indicators described above using Spearman rank coefficient of correlation. For ratings on a scale, the mean rating for each practice was calculated. Analysis was done with Stata 11 software.

Mapping

The level of NHS Choices usage to rate family practices online was mapped by ArcMap 9.3 software, using an Inverse Distance Weighted algorithm. The location of each practice was geocoded [13] and mapped with the corresponding data about ratings usage. Rate of using NHS Choices was measured as the number of ratings or comments divided by practice registered population, expressed as ratings per 1000 people. Where multiple practices share a postcode, the mean value of the rate was used.

Results

Descriptive Statistics

Of the 8089 practices included in the study, 4950 (61%) had been rated on the NHS Choices website. There were 16,952 ratings of these family practices. The mean number of ratings of each practice was 2.1; the median was 1. The range was from 0 to 149 ratings. A histogram of frequency of rating is shown in Figure 1. For those practices that had been rated, average ratings at the practice level for each of the questions by the rating website are shown in Table 1.

Figure 1. Histogram of rating frequency for all GP practices.

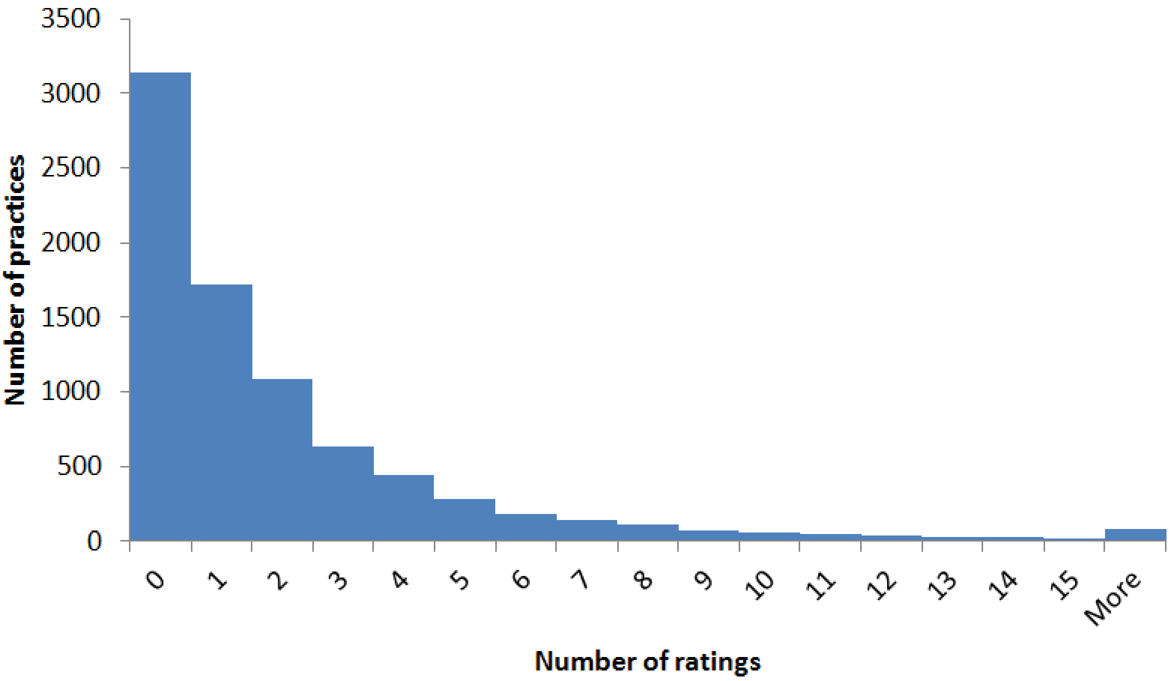


Table 1. Mean patient ratings posted for questions on NHS Choices.

Question asked	Mean rating	Interquartile range
I would recommend this GP practice to a friend.	64.0 %	33.3–100%
I am able to get through to the practice by telephone.	4.2 out of 5 (1 is lowest, 5 is highest)	3.7–5.0
This GP practice involves me in decisions about my care and treatment.	4.1 out of 5 (1 is lowest, 5 is highest)	3.5–5.0
I am able to get an appointment when I want one.	3.6 out of 5 (1 is lowest, 5 is highest)	3.0–4.3
I am treated with dignity and respect by the staff.	4.1 out of 5 (1 is lowest, 5 is highest)	3.5–5.0

Associations Between Population and Practice Characteristics and Whether a Practice is Rated

The logistic regression model showed that a larger practice size, a lower proportion of older patients, lower deprivation, higher population density, and not being a singlehanded practice were all positively associated with the likelihood of a practice being rated (Table 2). Whether the practice was a training practice, the proportion of white patients, and the nature of the practice contract did not appear to be associated with being rated. These results remained similar when the nonsignificant variables were consequentially excluded from the model.

Table 2. Associations between whether a practice is rated with population and practice characteristics.

Independent variable	Practices that have been rated on NHS Choices	Practices that have not been rated on NHS Choices	Z statistic	*P* value
Practice population size (Number of registered patients)	7587	5554	15.38	<.001
IMD score of patients (higher is more deprived)	25.1	28.2	-7.82	<.001
Population density (people/km^2)	458	403	6.72	<.001
Singlehander (% of practices which are singlehanders)	10	20.7	-4.50	<.001
Percentage of population aged over 65 years	15.1	15.6	-3.88	<.001
Percentage of population who are white	87.3	88.3	-1.58	.11
Type of contract (% with PMS contract)	42.7	39.6	-0.71	.48
Training practice (% that are training practices)	32.3	22.0	0.35	.73

Associations Between Population and Practice Characteristics and the Proportion Recommending

The least square regression model showed that smaller practice size, a higher proportion of white patients, lower population density, lower deprivation, being a training practice, and not being a singlehanded practice were all positively associated with higher levels of recommendation (Table 3). The type of practice contract and the age distribution of the patients were not associated with different recommendation levels. These results remained similar when the nonsignificant variables were consequentially excluded from the model.

Table 3. Associations between the proportion of patients recommending a practice with population and practice characteristics.

Independent variable	T statistic	*P* value
Proportion of population who are white	7.39	<.001
Training practice	6.61	<.001
Practice population size	-4.61	<.001
Population density	-4.17	<.001
IMD score of patients	-3.92	<.001
Singlehander	-3.76	<.001
Proportion of population aged over 65 years	1.78	.075
Type of contract	0.62	.53

Association Between Ratings and Conventional Quality Metrics

Associations between patient ratings and patient experience measures from the family practice survey had Spearman ρ values of between 0.36 and 0.48 and were all significant at the *P*<.001 level (Table 4). A Spearman ρ value of 1 represents a perfect rank correlation, 0 represents no correlation, and -1 represents a perfect negative rank correlation. Comparison between patient ratings and clinical quality indicators showed associations between a better rating on NHS Choices and better quality care for six of the seven variables (*P*<.001); however, Spearman ρ values were all less than ±0.2 (Table 5). There is a very weak negative correlation between ratings and low-cost statin prescriptions.

Table 4. Associations between web-based patient ratings and conventional survey measures of patient experience.

Web-based patient rating	NHS General Practice Patient Survey question on patient experience	Spearman ρ	*P* value
I would recommend this GP practice to a friend.	Recommending GP practice to someone who has moved to the local area—% yes	0.48	<.001
I am able to get through to the practice by telephone.	Ease of getting through on the phone—% easy	0.43	<.001
This GP practice involves me in decisions about my care and treatment.	Rating of doctor involving you in decisions about your care—% good	0.38	<.001
I am able to get an appointment when I want one.	Able to book ahead for an appointment with a doctor in the past 6 months—% yes	0.37	<.001
I am treated with dignity and respect by the staff.	Rating of doctor treating you with care and concern—Good	0.39	<.001

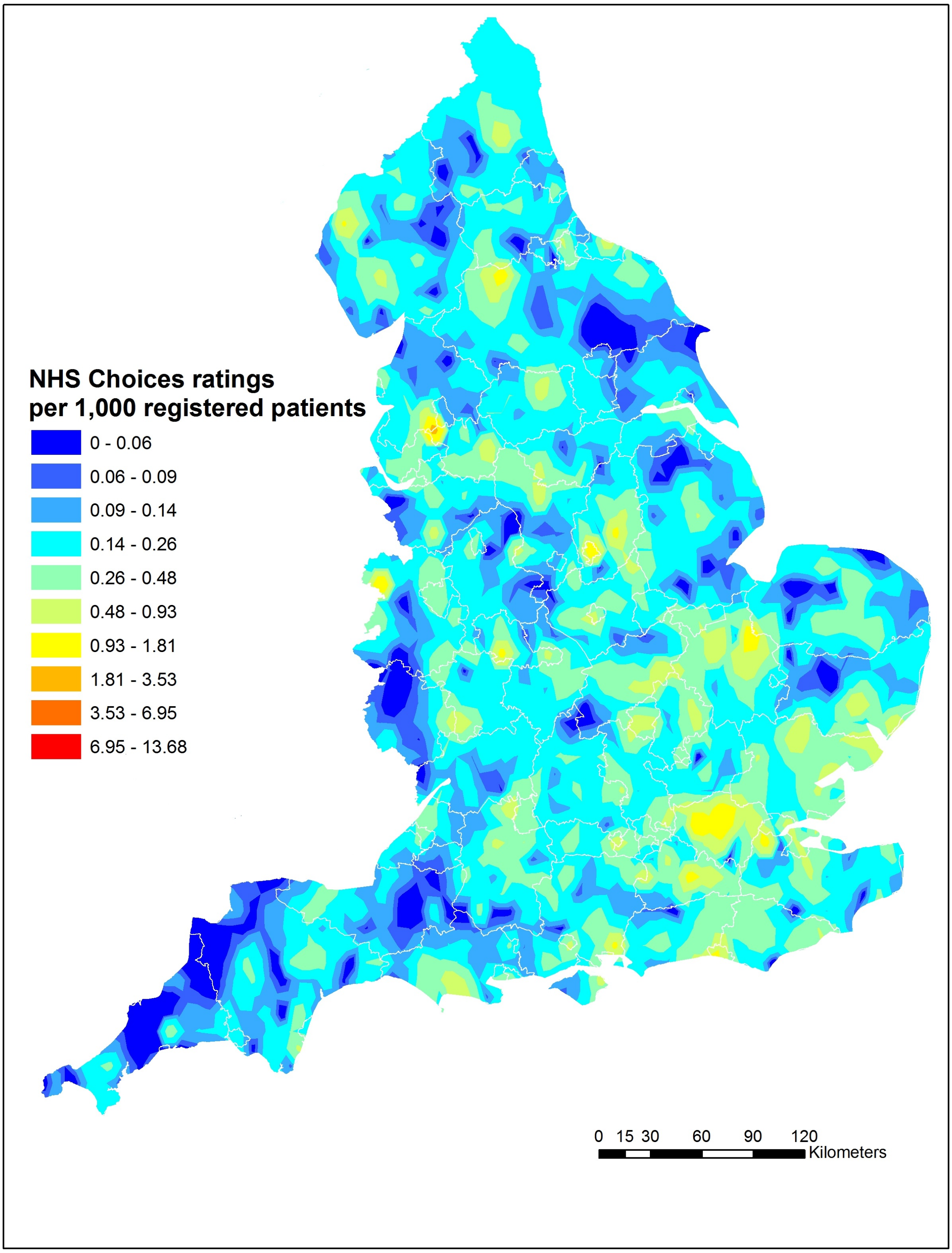
Table 5. Associations between web-based patient ratings of whether the user would recommend the GP practice to a friend and various clinical quality indicators.

Quality indicator	Spearman ρ	*P* value
Proportion of patients with diabetes receiving flu vaccination	0.07	<.001
Controlled blood pressure in hypertensive patients (systolic/diastolic less than 150/90 mm Hg)	0.07	<.001
Controlled HbA1C in patients with diabetes (less than 7%)	0.06	<.001
% low-cost statin prescribing	-0.03	.02
Cervical screening rate	0.18	<.001
Admission rates for ambulatory care sensitive conditions	-0.15	<.001
Total clinical QOF points available	0.11	<.001

Mapping

Figure 2 illustrates the spatial variation in the use of online ratings by GP practice. The map indicates that in urban areas usage of NHS Choices tends to be higher, particularly around London. Rates are lower in rural areas, the southwest, and northeast.

Figure 2. Map of frequency of primary care ratings on the NHS Choices website.



Discussion

The positive nature of responses (with 64% recommending) on the NHS Choices website is in line with other studies of patient-reported Internet feedback in acute hospitals in the UK (where 68% recommend) [14] but is lower than ratings of individual physicians in the USA (88% positive) [1]. These results suggest concerns that online feedback mechanisms will be used only by disgruntled patients hoping to complain are not true. However, as the recommendation level online is lower than in the patient survey (where 82% recommend), the results indicate that there may be a selection bias towards less satisfied patients compared to when patients views are randomly selected. It is also possible that the nature of the NHS Choices website, funded by the government rather than comparable websites in the US that are privately owned, may create a selection bias towards less satisfied patients.

The results demonstrate that usage of patient ratings of family practices via the NHS Choices website has been variable with some practices having more than 100 ratings, but many having none, and an average of only two ratings per practice over the 15-month period covered. A sizable minority (39%) of practices never received a single rating. The data analyzed here represent the initial period of the rating function being available, so usage may well increase as it becomes more popular. However, this usage of online ratings is higher than noted in other settings. A study in Boston, MA, found only 81 of a random sample of 300 physicians had been rated online [1], and a study found only 16% of primary care physicians had been rated on the *RateMDs* website in the US [15]. When compared to the approximately 300 million primary care appointments in England each year [16], the number of responses looks rather small (this corresponds to 0.005% of GP consultations being rated online).

The results suggest that the level of usage of online ratings is different in different communities. Practices serving younger, more urban, and less deprived communities were more likely to be rated. This is in line with previous work showing more usage of the Internet as a health resource in those groups [17].

The demonstration of an association between practice population characteristics and rating usage does not prove a link between individual characteristics and usage; there is a risk of the ecological fallacy, in which an incorrect inference can be drawn about individuals based on aggregated statistics about a group of people. However, these results do confirm that usage rates are variable around the country and suggest that individual characteristics may have a role in influencing usage. Further studies, using individual level data, are required to understand the characteristics of those using ratings websites. The results also show that different practice characteristics are associated with different levels of satisfaction with service, measured as willingness to recommend. This is in line with a wide body of literature on patient experiences, which notes that ethnicity and different socioeconomic factors are associated with satisfaction with medical care [18,19]. These findings may suggest that ratings websites might want to consider ways to broaden their appeal beyond certain groups of users, potentially marketing themselves towards older or rural populations.

Moderate associations between patients' ratings left on the web and more conventional surveys of patient experiences were found (Spearman ρ values of between 0.38 and 0.48). Due to the large number of practices in the analysis, these associations are all highly significant.

The association between ratings and clinical outcomes is less convincing. Although the results do show increasing levels of recommendation associated with better care across a number of indicators, for many of them the strength of the association is weak and significant only because of the large number of data points. This may be because there is a genuine tension between patient experience and some aspects of technical quality; it is possible to do all of the right things technically in health care and still get a bad outcome. It is also possible that patients' personal values may differ from the public health perspective captured in the quality metrics. Both the associations with clinical outcomes and survey measures of experience are consistent with similar findings at the hospital level [14,20,21].

These data suggest that we can be more confident in the use of online rating data as a measure of patient experience, despite the many fundamental biases that are an inevitable consequence of this sort of data. However, the extent to which online ratings reflect the technical quality of clinical primary care is less clear. As momentum develops around the need to capture the patient's experience, these ratings represent a potentially valuable source of information about quality of care when taken with other more conventional measures. Therefore, as such ratings become more common, systems should be developed to allow patients to examine these data alongside other more traditional outcome data, presented in a digestible and accessible way. Similarly, family physicians and practices should develop strategies to respond to these comments in constructive ways [22]. Further research to understand how practices use online feedback to improve the care they provide would be useful. Also, as many practices have few or no ratings online, work is needed to determine the numbers of ratings required over a defined period of time for a patient to obtain a reasonably accurate appreciation of the strengths and weaknesses of a family practice.

Study Limitations and Strengths

Our study has a number of limitations. We removed practices with a size below 1000 patients, as has been done in other analyses of practice performance [23], as smaller practices are often atypical, such as those recently opened or being closed or serving very specialized populations. However, as these exclusions represent less than 2% of practices in our sample, this is unlikely to have a major bearing on our findings. In addition, the timing of the ratings and the other outcome and experience measures do not match entirely, but we have chosen the available data with the most overlap. The use of practice population characteristics leaves some of our findings prone to the ecological fallacy, in which we make an inference about the nature of individuals based on the characteristics of the aggregated population. In addition, the clinical quality measures used may not reflect true variations in quality between family practices. As with other online rating systems, there is potential for ratings on NHS Choices to be "gamed" by organizations or for fake or multiple entries to be left by individuals. We also note that the some of the outcome metrics chosen here for quality of care are not entirely in the doctor's hands, but contingent on both the actions of the physician and the patient. This study also found many practices with no or few ratings.

Our study also has strengths. It uses a novel, largely complete national set of online ratings. The unique nature of the NHS's performance related pay system provides detailed information on clinical performance at the practice level, allowing detailed comparisons to be made.

Conclusions

These findings represent the first national review of an online ratings system in primary care. Our findings lend support to some of the arguments against online rating systems, particularly that they may have a selection bias. However, they also suggest some positive aspects. They are not just used as a mechanism to complain, as more often than not they are positive, and there are moderate associations with more traditional measures of patient satisfaction. As the numbers of rating per practice are low, they may be less useful as a measure of quality in family practice but they may provide a novel route for organizational learning. However, as numbers of ratings rise, it is likely that they will become a more useful tool for patients to make informed choices about where to receive their care.

Acknowledgments

The Department of Primary Care & Public Health at Imperial College London is grateful for support from the NIHR Collaboration for Leadership in Applied Health Research & Care (CLAHRC) Scheme, the NIHR Biomedical Research Centre scheme, and the Imperial Centre for Patient Safety and Service Quality. CM is funded by the Higher Education Funding Council for England and the National Institute for Health Research.

We are grateful to the staff at NHS Choices for providing the data for this analysis, in particular John Robinson.

Conflicts of Interest

Professor Darzi was Parliamentary Under-Secretary of State (Lords) in the United Kingdom Department of Health from 2007 to 2009. Professor Majeed has received funding for research from Dr Foster Intelligence, a commercial health intelligence company. Professor Majeed is also a General Practitioner in the NHS general practice of Dr Curran & Partners, which has a page on NHS Choices. For the other authors, no conflicts are declared.

References

1. Lagu T, Hannon NS, Rothberg MB, Lindenauer PK. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010 Sep;25(9):942-946. [doi: 10.1007/s11606-010-1383-0] [Medline: 20464523]
2. Lagu T, Lindenauer PK. Putting the public back in public reporting of health care quality. JAMA 2010 Oct 20;304(15):1711-1712. [doi: 10.1001/jama.2010.1499] [Medline: 20959582]
3. Kadry B, Chu LF, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. J Med Internet Res 2011;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960] [Medline: 22088924]
4. Greaves F, Millett C. Consistently increasing numbers of online ratings of healthcare in England. J Med Internet Res 2012;14(3):e94 [FREE Full text] [doi: 10.2196/jmir.2157] [Medline: 22742977]
5. Boffey D. Guardian. GPs' fury over 'poison pen' online tirades URL: http://www.guardian.co.uk/society/2011/dec/03/gps-poison-pen-online [accessed 2012-07-23] [WebCite Cache ID 69N3PvT5q]
6. McCartney M. Will doctor rating sites improve the quality of care? No. BMJ 2009;338:b1033. [Medline: 19293224]
7. Veloski J, Boex JR, Grasberger MJ, Evans A, Wolfson DB. Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. Med Teach 2006 Mar;28(2):117-128. [doi: 10.1080/01421590600622665] [Medline: 16707292]
8. Department of Health. 2012. NHS Choices website URL: http://www.nhs.uk/Pages/HomePage.aspx [accessed 2012-07-23] [WebCite Cache ID 69N0psewl]
9. 2009/10 Annual Report. Technical Annex for GP Patient Survey URL: http://www.gp-patient.co.uk/results/download/Y4Q4/Y4Q4_AnnualTechnical_unweighted.pdf [accessed 2012-10-08] [WebCite Cache ID 6BGpU4Xfx]
10. NHS Information Centre. 2012. Quality and Outcomes Framework: Online GP practice results database URL: http://www.qof.ic.nhs.uk/ [accessed 2012-07-23] [WebCite Cache ID 69N1IxOKJ]
11. NHS Information Centre. 2012. NHS Comparators URL: https://www.nhscomparators.nhs.uk/NHSComparators/Login.aspx [accessed 2012-07-23] [WebCite Cache ID 69N1LRso9]
12. NHS Connecting For Health. 2012. Organisation Data Service URL: http://www.connectingforhealth.nhs.uk/systemsandservices/data/ods/genmedpracs [accessed 2012-07-23] [WebCite Cache ID 69N1NRzvq]
13. UK Borders boundary dataset. (Archived by WebCite® at. 2012 URL: http://edina.ac.uk/ukborders/ [accessed 2012-07-23] [WebCite Cache ID 69N1PMYJV]
14. Greaves F, Pape UJ, King D, Darzi A, Majeed A, Wachter RM, et al. Associations between Web-based patient ratings and objective measures of hospital quality. Arch Intern Med 2012 Mar 12;172(5):435-436. [doi: 10.1001/archinternmed.2011.1675] [Medline: 22331980]
15. Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012;14(1):e38 [FREE Full text] [doi: 10.2196/jmir.2003] [Medline: 22366336]

16. Hippisley-Cox J, Vinogradova Y. The NHS Information Centre: Leeds. Trends in Consultation Rates in General Practice 1995 to 2008 URL: http://www.ic.nhs.uk/pubs/gpcons95-09 [accessed 2012-07-23] [WebCite Cache ID 69N3YjnNC]
17. Mead N, Varnam R, Rogers A, Roland M. What predicts patients' interest in the Internet as a health resource in primary care in England? J Health Serv Res Policy 2003 Jan;8(1):33-39. [doi: 10.1258/13558190360468209] [Medline: 12683432]
18. Salisbury C, Wallace M, Montgomery AA. Patients' experience and satisfaction in primary care: secondary analysis using multilevel modelling. BMJ 2010 Oct;341(120):c5004-c5003 [FREE Full text] [doi: 10.1136/bmj.c5004] [Medline: 20940212]
19. Hall JA, Dornan MC. Patient sociodemographic characteristics as predictors of satisfaction with medical care: a meta-analysis. Soc Sci Med 1990;30(7):811-818. [Medline: 2138357]
20. Greaves F, Pape UJ, King D, Darzi A, Majeed A, Wachter RM, et al. Associations between Internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. BMJ Qual Saf 2012 Jul;21(7):600-605. [doi: 10.1136/bmjqs-2012-000906] [Medline: 22523318]
21. Segal J, Sacopulos M, Sheets V, Thurston I, Brooks K, Puccia R. Online doctor reviews: do they track surgeon volume, a proxy for quality of care? J Med Internet Res 2012;14(2):e50 [FREE Full text] [doi: 10.2196/jmir.2005] [Medline: 22491423]
22. Londonwide Local Medical Committees. 2012. Guidance on negative comments posted on the NHS Choices website URL: http://www.lmc.org.uk/article.php?group_id=83 [accessed 2012-07-23] [WebCite Cache ID 69N1Ul20D]
23. Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U, et al. Pay-for-performance programs in family practices in the United Kingdom. N Engl J Med 2006 Jul 27;355(4):375-384. [doi: 10.1056/NEJMsa055505] [Medline: 16870916]

Edited by G Eysenbach; submitted 27.07.12; peer-reviewed by G Gao, J Segal, C Schaefer; comments to author 22.08.12; revised version received 23.08.12; accepted 25.08.12; published 17.10.12.

Please cite as:
Greaves F, Pape UJ, Lee H, Smith DM, Darzi A, Majeed A, Millett C
Patients' Ratings of Family Physician Practices on the Internet: Usage and Associations With Conventional Measures of Quality in the English National Health Service
J Med Internet Res 2012;14(5):e146
URL: http://www.jmir.org/2012/5/e146/
doi:10.2196/jmir.2280
PMID:23076301

Original Paper

Use and Acceptance of Electronic Communication by Patients With Multiple Sclerosis: A Multicenter Questionnaire Study

Rocco Haase[1]; Thorsten Schultheiss[1], MD; Raimar Kempcke[1], BA; Katja Thomas[1], MD; Tjalf Ziemssen[1], MD

Multiple Sclerosis Centre Dresden, Department of Neurology, University of Technology Dresden, Dresden, Germany

Corresponding Author:
Rocco Haase
Multiple Sclerosis Centre Dresden
Department of Neurology
University of Technology Dresden
Fetscherstrasse 74
Dresden, 01307
Germany
Phone: 49 351 458 4465
Fax: 49 351 458 5717
Email: Rocco.Haase@uniklinikum-dresden.de

Abstract

Background: The number of multiple sclerosis (MS) information websites, online communities, and Web-based health education programs has been increasing. However, MS patients' willingness to use new ways of communication, such as websites, mobile phone application, short message service, or email with their physician, remains unknown.

Objectives: We designed a questionnaire to evaluate the a priori use of electronic communication methods by MS patients and to assess their acceptance of such tools for communication with their health care providers.

Methods: We received complete data from 586 MS patients aged between 17 and 73 years. Respondents were surveyed in outpatient clinics across Germany using a novel paper-and-pencil questionnaire. In addition to demographics, the survey items queried frequency of use of, familiarity with, and comfort with using computers, websites, email, and mobile phones.

Results: About 90% of all MS patients used a personal computer (534/586) and the Internet (527/586) at least once a week, 87.0% (510/586) communicated by email, and 85.6% (488/570) communicated by mobile phone. When asked about their comfort with using electronic communication methods for communication with health care providers, 20.5% (120/586) accepted communication by mobile Internet application or short message service via mobile phone, 41.0% (240/586) by websites, 54.3% (318/586) by email service, and 67.8% (397/586) by at least one type of electronic communication. The level of a priori use was the best predictor for the acceptance of electronic communication with health care providers. Patients who reported already searching online for health information (odds ratio 2.4, *P* < .001) and who had already communicated with a physician through a website (odds ratio 3.3, *P* = .03) reported higher acceptance for Web-based communication. Patients who already scheduled appointments with their mobile phones (odds ratio 2.1, *P* = .002) were more likely to accept the use of mobile phone applications or short message service for communicating with their physician.

Conclusions: The majority of MS patients seen at specialist centers already use modern communication technology regularly. New forms of electronic communication appear to have high levels of acceptance for exchanging information about MS between patients and health care providers. Such methods should be integrated into eHealth services such as electronic health records and patient relationship management systems.

(J Med Internet Res 2012;14(5):e135) doi:10.2196/jmir.2133

KEYWORDS

Multiple sclerosis; computers; email; mobile phone; health information seeking; disease management; communications media; health information technology

Introduction

The way people retrieve health information has changed due to an enormous proliferation of new media technologies and a tremendous growth in health information being available online [1-3]. The Pew Internet & American Life Project subsequently found that 56% to 64% of Americans used online sources of health information [4]. In 2003, 23% of Germans searched online for health-related topics sometimes or often [5]. This trend may be increasing, as studies reported an increase from 44% to 57% in use of the Internet for health purposes in Germany between 2005 and 2007, with the highest use among young women and city dwellers [3,6-9]. Due to the high incidence of multiple sclerosis (MS) in young women (about 70%) showing their first symptoms between the ages of 20 and 40 years, MS patients may be early adopters of emerging eHealth trends [10-13]. To date, research has primarily focused on what the Internet offers to MS patients and how existing health information resources are being accessed. However, studies examining the acceptability of new media such as websites or mobile phone applications for communication with health care providers by MS patients are lacking.

Two studies investigating online information sources about MS in English and German reported variable quality [14,15]. In addition to accessing traditional static information websites, patients can also generate and share their health information with peers and participate in electronic health records for the self-management of disease, with mainly positive results [4,16-21]. Such online communities can also support research, such as the development of a self-report questionnaire to quantify MS patients' adherence to treatment [22]. Other studies reported the benefits of testing cognitive functions of MS patients online, including better availability and accessibility than with traditional methods [23-25].

Research suggests that MS patients retrieve information about their disease and their physicians online before and after their medical visit, especially before the initial consultation due to potentially high information needs in the early course of the disease [11,26,27]. This may influence patients' adherence to treatment and their coping styles, and may lead to greater patient empowerment [3,11,28-31], thereby altering the physician-patient relationship [28,32,33]. The observed benefits of new media may have raised the acceptance of eHealth in medicine, but concerns remain such as the digital divide (inequality of access to new communication technologies) and concerns about the security and confidentiality of sharing health data online.

To determine the needs of and future options for patient education and patient–physician communication via new media, we sought to examine the current use of new types of communication (eg, mobile phone applications) and willingness to use these for medical management among MS patients. We hypothesized that patients who were already familiar with electronic communication devices would be more likely to adopt them for communication with health care providers and that the majority of MS patients would already be using an array of electronic communication devices.

Methods

Participants

The present study was a multicenter survey conducted in the Multiple Sclerosis Center Dresden (Dresden, Germany), the Multiple Sclerosis Center Stuttgart (Stuttgart, Germany), and several neurological outpatient centers of NeuroTransData GmbH (Neuburg, Germany) between 2009 and 2010. Patients with clinically definite MS according to the McDonald criteria were asked to attend the survey in each participating center over a period of 2 weeks [34]. We targeted recruitment for a minimum of 500 fully completed questionnaires. Overall, 591 of 650 (90.9%) ambulatory MS patients who were asked to participate enrolled in the study. They were surveyed by the final version of the paper-based, self-developed questionnaire during outpatient clinic visits. Prior to study entry, each patient provided a written informed consent and was free to withdraw from the survey at any time for any reason without consequences on the care provided. Because this study involved minimal risk and no personally identifiable information, ethics committee approval was not required.

Questionnaire

The 18 items in the questionnaire were developed by a scientific advisory board consisting of physicians, psychologists, and computer scientists. The questionnaire surveyed the frequency and nature of personal computer, website, email, and mobile phone use by MS patients, as well as demographic characteristics. Specifically, the questionnaire included 4 demographic items about age, sex, time since MS onset (in years), and residential area (by postal code) (items 1–4), 5 items about computer use (items 5–9), 5 items about Internet use (items 10–14), and 3 items about mobile phone use (items 15–17). We added 1 item about the respondent's general attitude toward using new media for communication with health care providers (item 18) (see Multimedia Appendix 1).

Statistical Analysis

Patients were assigned to the *rural area* group (population of place of residence <100,000) or to the *city* group (population of place of residence ≥100,000) based on postal code (questionnaire item 3) [35,36]. We determined disease duration (in years) by calculating the difference between MS diagnosis (questionnaire item 4) and the year of study entry. Use of each medium at least once a week indicated a sufficient frequency for future communication with health care providers. New media acceptance was reported for patients accepting at least one of three types of new media (website, email, and mobile Internet application or short message service) for communication with health care providers. To analyze the influence of age on the a priori use of each medium, we grouped respondents by a median split (younger half: 17-40 years, n = 283; older half: 41-73 years, n = 298).

All statistical comparisons were 2-tailed, and a *P* value of <.05 indicated statistical significance. We used IBM SPSS version 19.0 (IBM Corporation, Somers, NY, USA) for all statistical computations. Acceptance was evaluated by a logistic regression model for each new media type (website, email, mobile Internet application or short message service) using demographic variables and items of specific media type and reporting goodness of fit (Hosmer-Lemeshow test; chi-square) and effect size (odds ratio). Differences in categorical variables were assessed with a chi-square test for between-participants effects and with McNemar test for matched pairs.

Results

Characteristics of the Survey Population

In the statistical analysis, we included 586 of 591 MS patients who submitted answers to the last survey question (questionnaire item 18). The majority of study patients were female (408/586, 69.6%) with a mean age of 40.93 (SD 10.84) years (5 missing answers), comparing well with typical population statistics of patients with MD [37]. A total of 6 in 10 patients (336/586, 57.3%) lived in a city area with a population of 100,000 people or more. The disease duration averaged 8.60 (SD 6.52) years since onset (7 missing answers). In all analyses, we observed no impact of residential area (rural area, town, or city) on any item response. Furthermore, patients did not differ between states of the former German Democratic Republic (Saxony) and the Federal Republic of Germany (Baden-Württemberg).

Computer Use

The vast majority of MS patients (558/586, 95.2%) had access to a computer and 75.4% (442/586) of them personally owned one (Table 1); 70.8% (415/586) used it daily and 91.1% (534/586) used it at least weekly. Men reported higher rates of general use (*P* < .001), computer ownership (men: 151/178, 84.8%; women: 291/408, 71.3%; *P* < .001), and experience in the installation of new software (men: 144/178, 80.9%; women: 231/408, 56.6%; *P* < .001). Nevertheless, computer use was widespread among women. Most patients (509/586, 86.9%) reported acquainting themselves quickly with new software. Younger MS patients tended toward a greater (*P* < .001) and more diversified use of computers (emailing: *P* = .003; browsing the Internet: *P* = .007; chatting: *P* < .001; having already installed new computer programs: *P* < .001; and becoming quickly acquainted with new computer programs: *P* < .001). In all subgroups of MS patients, about 40% (62/178, 34.8%, to 170/408, 41.7%) reported using a computer to retrieve information about their disease. All results are summarized in Multimedia Appendix 2.

Table 1. Computer use by patients with multiple sclerosis (MS) (n = 586).

Computer use	%	n
Frequency of computer use		
Several times a day	46.6	273
Daily	24.2	142
Several times a week	15.0	88
Once a week	5.3	31
Rarely or never	8.9	52
Computer ownership or shared access	95.2	558
Own a computer	75.4	442
Type of regular computer use		
Browsing the Internet	81.9	480
Emailing	81.7	479
Word processing	62.6	367
Getting information about MS	39.6	232
Chatting	12.5	73
Quick familiarization with a new computer program		
Definitely applies to me	29.0	170
Mostly applies to me	33.6	197
Slightly applies to me	24.2	142
Does not apply to me	13.1	77
Installation of computer programs	64.4	375

Internet Use

About 94% of MS patients (551/586) reported they had Internet access (Table 2). Men browsed websites more frequently (*P* = .009), whereas the number of male and female nonusers was similar. The pattern was the same in the case of reading emails (*P* = .002) and sending emails (*P* = .03). Young MS patients more frequently than older patients performed tasks such as browsing (*P* < .001), chatting (*P* < .001), and emailing (reading: *P* = .02; sending: *P* = .004). MS-related information seeking was reported by between 35.2% (105/298) and 38.9% (110/283) among all subgroups of MS patients. A small number of patients had already communicated with their physician (22/586, 3.8%) or with other patients (29/586, 5.0%) via the Internet (Table 2).

Table 2. Internet use by patients with multiple sclerosis (MS) (n = 586).

Internet use	%	n
Frequency of browsing websites on the Internet		
Several times a day	34.6	203
Once a day	25.9	152
Several times a week	23.2	136
Once a week	6.1	36
Rarely or never	10.1	59
Internet access at home		
Broadband access	69.6	408
Low-speed access	20.1	118
No access	6.0	35
Access type unknown	4.3	25
Type of regular Internet use		
Browsing websites	81.4	477
Getting information about MS	37.2	218
Video chatting	11.1	65
Chatting	10.2	60
Communicating with other MS patients	5.0	29
Communicating with physician	3.8	22
Frequency of sending emails		
Several times a day	31.1	182
Once a day	15.4	90
Several times a week	23.0	135
Once a week	11.6	68
Rarely or never	18.9	111
Frequency of reading emails		
Several times a day	37.2	218
Once a day	23.2	136
Several times a week	17.4	102
Once a week	9.2	54
Rarely or never	13.0	76

Mobile Phone Use

Nearly all MS patients possessed a mobile phone (553/576, 96.0%) but older patients used it less frequently ($P < .001$) and less extensively (calling: $P = .002$; text messaging: $P < .001$; browsing websites: $P < .001$; reading or sending emails: $P = .001$; and scheduling: $P < .001$) (Table 3). Women showed a greater tendency for text messaging ($P < .001$) but men were more likely to operate smartphone abilities such as browsing websites via mobile Internet ($P = .01$) or scheduling appointments ($P = .03$).

Acceptance of Electronic Communication With Health Care Providers

When asked about the acceptability of using various modes of communication, including electronic communication devices, for receiving information and guidance from their physician for managing their MS, 20.5% (120/586) of MS patients accepted communication by mobile phone Internet application or short message service, 41.0% (240/586) by website, 54.3% (318/586) by email, and 67.8% (397/586) by at least one of these modes of communication (Table 4). The majority of patients (539/586, 92.0%) found existing traditional methods acceptable. Acceptance of email services for this purpose exceeded the acceptance of telephone conversations (263/586, 44.9%; $P = .007$). Women were more likely to accept communication by telephone ($P = .01$).

We included media-specific items and demographic variables in all cross-sectional logistic regression models resulting in admissible model fits (Hosmer-Lemeshow test for email: $\chi^2_8 = 7.0$, $P = .54$; for websites: $\chi^2_8 = 6.7$, $P = .58$; and for mobile phone features: $\chi^2_8 = 9.2$, $P = .32$). Neither sex nor age had an impact on the acceptance of any type of electronic communications media. We eliminated duration of disease from the analyses to avoid multicollinearity, since it was highly correlated with age.

In general, patients with regular a priori use of new electronic media were more likely to accept this form of communication with health care providers (Table 5). More specifically, reading emails at least once a week, browsing Internet websites at least several times a week, and using a mobile phone daily raised the level of acceptance significantly.

Furthermore, online health information seekers (odds ratio 2.4) and patients having already communicated with their physician through a website (odds ratio 3.3) showed a greater interest in website-based communication (Table 5). Scheduling appointments on a mobile phone (odds ratio 2.1) was the only specific task that raised the likelihood of accepting mobile phone features for communication with health care providers.

Table 3. Mobile phone use by patients with multiple sclerosis (n = 586)[a].

Mobile phone use	%	n
Mobile phone ownership	96.0	553
Missing answers (item 15)		10
Frequency of mobile phone use		
Several times a day	38.0	219
Once a day	17.5	101
Several times a week	21.5	124
Once a week	7.6	44
Rarely or never	14.2	82
Missing answers (item 16)		16
Type of regular mobile phone use		
Calling	89.1	513
Text messaging	63.9	368
Scheduling	21.7	125
Audio or video messaging	5.2	30
Reading or sending emails	4.9	28
Browsing websites	4.7	27
Missing answers (item 17)		16

[a] For all data percentages were based on valid answers.

Table 4. Acceptance of modes of communication with health care providers for being informed and instructed during multiple sclerosis therapy (MS) by patients with MS (n = 586).

Type of communication	%	n
By physician (in-person)	92.0	539
By email or mobile phone or website	67.8	397
By email	54.3	318
By telephone call	44.9	263
By website	41.0	240
Via mobile Internet application or short message service	20.5	120

Table 5. Acceptance of new media types for communication with health care providers by patients with multiple sclerosis (MS).

Characteristics	Odds ratio	95% confidence interval	*P* value[a]
Acceptance of emails (n = 581)			
Reading emails several times a day	16.3	4.8–55.1	<.001
Reading emails daily	13.1	4.2–41.0	<.001
Reading emails several times a week	13.3	4.2–41.8	<.001
Reading emails once a week	7.0	2.4–20.3	<.001
Reading emails rarely or never	1 (reference)		
Sending emails several times a day	1.0	0.4–2.6	.97
Sending emails daily	1.4	0.6–3.7	.45
Sending emails several times a week	1.4	0.6–3.3	.43
Sending emails once a week	0.8	0.4–1.8	.53
Sending emails rarely or never	1 (reference)		
Age	1.0	0.8–1.2	.88
Females	1.0	0.7–1.5	.95
Males	1 (reference)		
Acceptance of Internet websites (n = 581)			
Browsing websites several times a day	5.3	1.8–15.8	.003
Browsing websites daily	5.2	1.8–15.4	.003
Browsing websites several times a week	5.7	1.9–16.8	.002
Browsing websites once a week	3.1	0.9–10.7	.07
Browsing websites rarely or never	1 (reference)		
Using the Internet for browsing	1.3	0.7–2.3	.45
Using the Internet for chatting	1.4	0.8–2.7	.30
Using the Internet for video chatting	1.5	0.9–2.7	.13
Using the Internet for information about MS	2.4	1.7–3.5	<.001
Using the Internet for communication with physicians	3.3	1.1–10.1	.03
Using the Internet for communication with other MS patients	1.9	0.8–4.5	.15
Age	1.1	0.9–1.3	.49
Females	1.2	0.8–1.8	.39
Males	1 (reference)		
Acceptance of mobile Internet applications or short message service (n = 565)			
Using a mobile phone several times a day	9.3	2.0–41.7	.004
Using a mobile phone daily	5.9	1.3–27.3	.02
Using a mobile phone several times a week	2.9	0.6–13.7	.17
Using a mobile phone once a week	1.3	0.2–9.6	.81
Using a mobile phone rarely or never	1 (reference)		
Using a mobile phone for calling	6.9	0.9–52.4	.06
Using a mobile phone for text messaging	1.3	0.8–2.3	.32
Using a mobile phone for audio or video messaging	1.4	0.6–3.2	.45
Using a mobile phone for browsing websites	1.3	0.4–4.2	.62
Using a mobile phone for sending and reading emails	0.5	0.2–1.5	.22
Using a mobile phone for scheduling	2.1	1.3–3.6	.002

Characteristics	Odds ratio	95% confidence interval	P value[a]
Age	0.9	0.7–1.2	.59
Females	1.2	0.8–2.0	.37
Males	1 (reference)		

[a] $P < .05$ was considered significant.

Discussion

In this survey, we analyzed the a priori use of computers, the Internet, emails, and mobile phones by patients with MS, as well as their willingness to adopt them for communication with health care providers. Our results indicate that the use of new communication technologies (computers, websites, emails, and mobile phones) by MS patients is already widespread. In addition, the majority of patients reported relevant information and communication technology skills such as installing software. Sharing information and receiving guidance in MS management via email and via website was well accepted among MS patients, with lower levels of acceptability for communication via mobile phone features (mobile Internet applications or short message service). A priori use was the most important predictor of accepting new media for communication with health care providers. Although differences in the a priori use of electronic communication devices between men and women and between younger and older patients were significant, neither sex nor age had an impact on the acceptance of these tools for communicating with health care providers.

Communication Technology Use by MS Patients

The vast majority of MS patients had access to a computer or owned one themselves. They were regularly visiting websites, and reading and writing emails, and possessed a mobile phone, consistent with results of smaller MS-focused studies [26,31]. The number of Internet users was slightly higher than in more general reports from the Pew Internet & American Life Project [4] but corresponded to the number from Lejbkowicz et al [31]. In agreement with Green et al [38], our study found less-intensive and less-diversified use of computers and Internet services among older people. Our results indicated that women used computers, the Internet, and email services less often, but the prevalence of female nonusers was similar to that of male nonusers. Similarly, Kummervold et al [6] reported an increasing use of the Internet by females in Europe between 2005 and 2007. Mobile phones were the most commonly used devices. As anticipated, young male patients were more likely to use smartphone abilities such as mobile Internet or appointment scheduling.

In our study, 4 in 10 MS patients were already using the Internet regularly for health information seeking and other disease-related tasks. In comparison, 56% to 61% of Americans searched online for health information [3,4]. However, differences between study findings in the use of eHealth services by chronically ill patients may be due to how the studies define media use as described by Wagner et al [39], indicating that 11% accessed health information services at least monthly but 45.9% had ever used them in the previous year [2]. Several European studies noted an increasing rate of Internet use for health purposes, with the highest increase being in Germany (from 44.4% in 2005 to 56.6% in 2007) [6,8,9]. In contrast to Santana et al [8], we did not find any differences in the use of new media with respect to place of residence.

In contrast to Hay et al [11], who reported that 82% of MS patients retrieved medical information online before the first visit, our rates were lower. Lejbkowicz et al [31] observed that 63% of MS patients used the Internet for MS-related tasks, but they did not state how they obtained that information. Nonetheless, only a minority of MS patients used the opportunity to correspond with their physician or other patients via the Internet. Although other studies reported similar results, evidence for increasing interest in online health care services is growing [2,6,8,9,40].

Acceptance of Electronic Communication with Health Care Providers

In addition to the a priori use of electronic communication devices, we investigated the a priori attitude toward adopting these services for aiding in MS disease management. Most participants accepted more than one type of new media for communication with health care providers. Communication via email and website was considered as acceptable as conventional telephone calls. A notable proportion of patients perceived the integration of mobile phones and mobile Internet to aid their therapy as useful. Two-thirds of MS patients were willing to use at least one new media service for communication with health care providers. The main factor in accepting a service was a priori use. Patients who read emails at least once a week, browsed websites several times a week, and used a mobile phone daily were more likely to approve a type of communication. Likewise, current users of electronic communication devices for health care management showed a greater interest in Web-based communication.

Several studies support the open-mindedness of patients toward using email communication in disease management. Hassol et al [16] reported that patients who used electronic health records generally preferred emails to telephone calls. In a study focusing on older patients, Singh et al [40] found that 49.3% of them could imagine using emails for communication with their physician. In agreement with our findings, a priori use of email services was the best predictor for acceptance of email communication with health care providers [40]. Santana et al [9] reported that about 18% of Europeans usually contact their physician via Internet and 25% would like to make their appointments online. Moreover, 4 out of 10 people in Europe would select their physician with regard to eHealth services provided [9]. Similarly, 49% of Israeli MS patients welcomed the opportunity to communicate with their medical team online [31].

Limitations

Some limitations of our survey warrant consideration. Participation in the survey was nonrestricted and hence may have introduced selection bias by motivation or interest. However, this was minimized by the large number of participants. Moreover, relevant demographic data, such as income or ethnicity, that other studies had used were lacking in our survey. Therefore, we were unable to test assumptions about potential obstacles posed by new media such as the digital divide. As income is a sensitive issue, participants may not have been willing to specify their personal particulars. We conducted our survey using paper-and-pencil questionnaires to avoid selection bias. Differences in the methods of similar studies in defining and assessing media use are important when attempting to compare studies on the use of health care information and communication technology. The questionnaire provided only a single item for the acceptability of new communication methods with health care providers, which may have been insufficient to obtain nuance and detail about the patients' true attitudes. We surveyed attitudes toward electronic communication devices but lacked data on actual behavior or objective use.

Conclusions

In our survey, we obtained data on the a priori use of new media by chronically ill patients with MS, as well as their attitudes toward future use of new media for communication with health care providers and differentiating between common communication types.

To summarize, our results indicate that the majority of MS patients are willing to use new media for further eHealth implementations. Although the potential benefits (and risks) of using electronic communication devices in MS health care services remain to be established, our data suggest these tools can be integrated with electronic health records and patient relationship management systems in order to increase the range of potential users and capabilities.

We agree with Nijland et al [41] that delivering high-quality health care to patients in as many suitable ways as possible should make a significant impact on the design of eHealth applications. eHealth services such as patient relationship management systems and electronic health records that focus on only a single communication type will miss the chance to maximize their effectiveness for use in the health care process. According to our findings, there are two gaps. First, although 40% to 50% of patients would like to use eHealth services to aid their therapy, only 5% of patients currently use them. Second, 90% of patients are technically skilled enough to use information and communication technology but only 50% are willing to use health information and communication technology. To overcome these obstacles, the design and implementation of eHealth applications in the health care process have to be tailored to patients' individual needs. Further research should focus on educational content as well as on technical options to deliver the content to those who would benefit from it.

Acknowledgments

Anonymous data collection from MS patients was supported by Merck Serono, Germany, which provided financial compensation for Dr Gottwald, Stuttgart, and NeuroTransData GmbH, Neuburg.

Conflicts of Interest

None declared.

Multimedia Appendix 1

English version of the new media questionnaire.

[PDF File (Adobe PDF File), 51KB - jmir_v14i5e135_app1.pdf]

Multimedia Appendix 2

Results of the survey on the use and acceptance of electronic communication by patients with multiple sclerosis.

[PDF File (Adobe PDF File), 75KB - jmir_v14i5e135_app2.pdf]

Multimedia Appendix 3

English version of the new media questionnaire.

[PDF File (Adobe PDF File), 9KB - jmir_v14i5e135_app3.pdf]

References

1. Eysenbach G, Kohler C. What is the prevalence of health-related searches on the World Wide Web? Qualitative and quantitative analysis of search engine queries on the internet. AMIA Annu Symp Proc 2003:225-229. [Medline: 14728167]

2. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the Internet and e-mail for health care information: results from a national survey. JAMA 2003 May 14;289(18):2400-2406. [doi: 10.1001/jama.289.18.2400] [Medline: 12746364]
3. Ybarra ML, Suman M. Help seeking behavior and the Internet: a national survey. Int J Med Inform 2006 Jan;75(1):29-41. [doi: 10.1016/j.ijmedinf.2005.07.029] [Medline: 16129659]
4. Fox S, Jones S. The social life of health information. Washington, DC: Pew Internet & American Life Project; 2009 Jun. URL: http://www.pewinternet.org/~/media//Files/Reports/2009/PIP_Health_2009.pdf [accessed 2012-04-11] [WebCite Cache ID 66qPOVMrH]
5. Gesundheitsberichterstattung des Bundes. Berlin: Robert Koch-Institut; 2003. Nutzung von Informationsquellen über Gesundheit URL: https://www.gbe-bund.de/gbe10/ergebnisse.prc_fid_anzeige?p_fid=9778&p_fund_typ=TAB&p_sprachkz=D&p_uid=gast&p_aid=26085307&p_prot=1 [accessed 2012-04-11] [WebCite Cache ID 66qQEfwog]
6. Kummervold PE, Chronaki CE, Lausen B, Prokosch HU, Rasmussen J, Santana S, et al. eHealth trends in Europe 2005-2007: a population-based survey. J Med Internet Res 2008;10(4):e42 [FREE Full text] [doi: 10.2196/jmir.1023] [Medline: 19017584]
7. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]
8. Santana S, Lausen B, Bujnowska-Fedak M, Chronaki CE, Prokosch HU, Wynn R. Informed citizen and empowered citizen in health: results from an European survey. BMC Fam Pract 2011;12:20 [FREE Full text] [doi: 10.1186/1471-2296-12-20] [Medline: 21496309]
9. Santana S, Lausen B, Bujnowska-Fedak M, Chronaki C, Kummervold PE, Rasmussen J, et al. Online communication between doctors and patients in Europe: status and perspectives. J Med Internet Res 2010;12(2):e20 [FREE Full text] [doi: 10.2196/jmir.1281] [Medline: 20551011]
10. Kobelt G, Berg J, Lindgren P, Elias WG, Flachenecker P, Freidel M, et al. Costs and quality of life of multiple sclerosis in Germany. Eur J Health Econ 2006 Sep;7 Suppl 2:S34-S44. [doi: 10.1007/s10198-006-0384-8] [Medline: 17310337]
11. Hay MC, Strathmann C, Lieber E, Wick K, Giesser B. Why patients go online: multiple sclerosis, the internet, and physician-patient communication. Neurologist 2008 Nov;14(6):374-381. [doi: 10.1097/NRL.0b013e31817709bb] [Medline: 19008743]
12. Nielsen AS, Halamka JD, Kinkel RP. Internet portal use in an academic multiple sclerosis center. J Am Med Inform Assoc 2012;19(1):128-133. [doi: 10.1136/amiajnl-2011-000177] [Medline: 21571744]
13. Ziemssen T. Symptom management in patients with multiple sclerosis. J Neurol Sci 2011;311(suppl 1):S48-S52. [doi: 10.1016/S0022-510X(11)70009-0]
14. Harland J, Bath P. Assessing the quality of websites providing information on multiple sclerosis: evaluating tools and comparing sites. Health Informatics J 2007 Sep;13(3):207-221. [doi: 10.1177/1460458207079837] [Medline: 17711882]
15. Hoffmann T, Twork S, Pöhlau D, Kugler J. Patientenorientierung im Internet: qualitative Bewertung von Internetseiten für Multiple-Sklerose-Betroffene. Akt Neurol 2009;36(04):151-157. [doi: 10.1055/s-0028-1090238]
16. Hassol A, Walker JM, Kidder D, Rokita K, Young D, Pierdon S, et al. Patient experiences and attitudes about access to a patient electronic health care record and linked web messaging. J Am Med Inform Assoc 2004;11(6):505-513 [FREE Full text] [doi: 10.1197/jamia.M1593] [Medline: 15299001]
17. Lowe-Strong A, McCullagh PJ. Monitoring of symptoms and interventions associated with multiple sclerosis. Stud Health Technol Inform 2005;117:223-228. [Medline: 16282674]
18. Michel-Verkerke MB, Schuring RW, Spil TAM. Workflow management for multiple sclerosis patients: IT and organization. In: Proceedings of the 37th Annual Hawaii International Conference on Systems Sciences.: IEEE; 2004 Presented at: 37th Annual Hawaii International Conference on Systems Sciences; Jan 5-8, 2004; Big Island, Hawaii. [doi: 10.1109/HICSS.2004.1265369]
19. Schultheiss T, Kempcke R, Kratzsch F, Eulitz M, Pette M, Reichmann H, et al. [Multiple sclerosis management system 3D. Moving from documentation towards management of patients]. Nervenarzt 2012 Apr;83(4):450-457. [doi: 10.1007/s00115-011-3376-6] [Medline: 22038386]
20. Hatzakis MJ, Allen C, Haselkorn M, Anderson SM, Nichol P, Lai C, et al. Use of medical informatics for management of multiple sclerosis using a chronic-care model. J Rehabil Res Dev 2006;43(1):1-16. [Medline: 16847767]
21. Fernandez-Luque L, Elahi N, Grajales FJ. An analysis of personal medical information disclosed in YouTube videos created by patients with multiple sclerosis. Stud Health Technol Inform 2009;150:292-296. [Medline: 19745316]
22. Wicks P, Massagli M, Kulkarni A, Dastani H. Use of an online community to develop patient-reported outcome instruments: the Multiple Sclerosis Treatment Adherence Questionnaire (MS-TAQ). J Med Internet Res 2011;13(1):e12 [FREE Full text] [doi: 10.2196/jmir.1687] [Medline: 21266318]
23. Goverover Y, O'Brien AR, Moore NB, DeLuca J. Actual reality: a new approach to functional assessment in persons with multiple sclerosis. Arch Phys Med Rehabil 2010 Feb;91(2):252-260. [doi: 10.1016/j.apmr.2009.09.022] [Medline: 20159130]

24. Negreiros MA, Mattos P, Landeira-Fernandez J, Paes RA, Alvarenga RP. A brief neuropsychological screening test battery for cognitive dysfunction in Brazilian multiple sclerosis patients. Brain Inj 2008 May;22(5):419-426. [doi: 10.1080/02699050801998243] [Medline: 18415722]
25. Younes M, Hill J, Quinless J, Kilduff M, Peng B, Cook SD, et al. Internet-based cognitive testing in multiple sclerosis. Mult Scler 2007 Sep;13(8):1011-1019. [doi: 10.1177/1352458507077626] [Medline: 17623730]
26. Atreja A, Mehta N, Miller D, Moore S, Nichols K, Miller H, et al. One size does not fit all: using qualitative methods to inform the development of an Internet portal for multiple sclerosis patients. AMIA Annu Symp Proc 2005:16-20. [Medline: 16778993]
27. Matti AI, McCarl H, Klaer P, Keane MC, Chen CS. Multiple sclerosis: patients' information sources and needs on disease symptoms and management. Patient Prefer Adherence 2010;4:157-161 [FREE Full text] [Medline: 20622916]
28. Ball MJ, Lillis J. E-health: transforming the physician/patient relationship. Int J Med Inform 2001 Apr;61(1):1-10. [Medline: 11248599]
29. Wald HS, Dube CE, Anthony DC. Untangling the Web--the impact of Internet use on health care and the physician-patient relationship. Patient Educ Couns 2007 Nov;68(3):218-224. [doi: 10.1016/j.pec.2007.05.016] [Medline: 17920226]
30. Lode K, Larsen JP, Bru E, Klevan G, Myhr KM, Nyland H. Patient information and coping styles in multiple sclerosis. Mult Scler 2007 Jul;13(6):792-799. [doi: 10.1177/1352458506073482] [Medline: 17613608]
31. Lejbkowicz I, Paperna T, Stein N, Dishon S, Miller A. Internet usage by patients with multiple sclerosis: implications to participatory medicine and personalized healthcare. Mult Scler Int 2010;2010:640749 [FREE Full text] [doi: 10.1155/2010/640749] [Medline: 22096625]
32. Pucci E. Is the internet transforming the physician-consumer relationship? Preliminary data in a neurological setting. Eur J Neurol 2003 Mar;10(2):192. [Medline: 12603300]
33. Gerber BS, Eiser AR. The patient physician relationship in the Internet age: future prospects and the research agenda. J Med Internet Res 2001;3(2):E15 [FREE Full text] [doi: 10.2196/jmir.3.2.e15] [Medline: 11720957]
34. Polman CH, Reingold SC, Edan G, Filippi M, Hartung HP, Kappos L, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". Ann Neurol 2005 Dec;58(6):840-846. [doi: 10.1002/ana.20703] [Medline: 16283615]
35. Statistisches Landesamt Baden-Wuerttemberg, Germany. Statistische Berichte Baden-Wuerttemberg. 2011 Jul 20. Bevoelkerungsentwicklung in den Gemeinden Baden-Wuerttembergs 2010 URL: http://www.statistik-bw.de/Veroeffentl/Statistische_Berichte/3126_10001.pdf [accessed 2012-07-05] [WebCite Cache ID 68vaifj2U]
36. Statistisches Landesamt des Freistaates Sachsen, Germany. www.statistik.sachsen.de. 2012 Jan. Bevoelkerung des Freistaates Sachsen jeweils am Monatsende ausgewaehlter Berichtsmonate nach Gemeinden URL: http://www.statistik.sachsen.de/download/010_GB-Bev/Bev_Gemeinde.pdf [accessed 2012-07-05] [WebCite Cache ID 68vb49nJu]
37. Minden SL, Frankel D, Hadden L, Hoaglin DC. Access to health care for people with multiple sclerosis. Mult Scler 2007 May;13(4):547-558. [doi: 10.1177/1352458506071306] [Medline: 17463077]
38. Green BB, Anderson ML, Ralston JD, Catz S, Fishman PA, Cook AJ. Patient ability and willingness to participate in a web-based intervention to improve hypertension control. J Med Internet Res 2011;13(1):e1 [FREE Full text] [doi: 10.2196/jmir.1625] [Medline: 21371993]
39. Wagner TH, Baker LC, Bundorf MK, Singer S. Use of the Internet for health information by the chronically ill. Prev Chronic Dis 2004 Oct;1(4):A13 [FREE Full text] [Medline: 15670445]
40. Singh H, Fox SA, Petersen NJ, Shethia A, Street RL. Older patients' enthusiasm to use electronic mail to communicate with their physicians: cross-sectional survey. J Med Internet Res 2009;11(2):e18 [FREE Full text] [doi: 10.2196/jmir.1143] [Medline: 19632972]
41. Nijland N, van Gemert-Pijnen J, Boer H, Steehouder MF, Seydel ER. Evaluation of internet-based technology for supporting self-care: problems encountered by patients and caregivers when using self-care applications. J Med Internet Res 2008;10(2):e13 [FREE Full text] [doi: 10.2196/jmir.957] [Medline: 18487137]

Abbreviations

MS: multiple sclerosis

Edited by P Wicks; submitted 12.04.12; peer-reviewed by R Bove, T Holm; comments to author 07.05.12; revised version received 18.06.12; accepted 08.08.12; published 15.10.12.

Please cite as:
Haase R, Schultheiss T, Kempcke R, Thomas K, Ziemssen T
Use and Acceptance of Electronic Communication by Patients With Multiple Sclerosis: A Multicenter Questionnaire Study
J Med Internet Res 2012;14(5):e135
URL: http://www.jmir.org/2012/5/e135/
doi:10.2196/jmir.2133
PMID:23069209

Original Paper

Two h-Index Benchmarks for Evaluating the Publication Performance of Medical Informatics Researchers

Khaled El Emam[1,2], PhD; Luk Arbuckle[1], MSc; Elizabeth Jonker[1], BA; Kevin Anderson[1], BA

[1]Electronic Health Information Laboratory, Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada
[2]Department of Pediatrics, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

Corresponding Author:
Khaled El Emam, PhD
Electronic Health Information Laboratory
Children's Hospital of Eastern Ontario Research Institute
401 Smyth Road
Ottawa, ON, K1H 8L1
Canada
Phone: 1 613 737 7600 ext 4181
Fax: 1 613 731 1374
Email: kelemam@uottawa.ca

Abstract

Background: The *h*-index is a commonly used metric for evaluating the publication performance of researchers. However, in a multidisciplinary field such as medical informatics, interpreting the *h*-index is a challenge because researchers tend to have diverse home disciplines, ranging from clinical areas to computer science, basic science, and the social sciences, each with different publication performance profiles.

Objective: To construct a reference standard for interpreting the *h*-index of medical informatics researchers based on the performance of their peers.

Methods: Using a sample of authors with articles published over the 5-year period 2006–2011 in the 2 top journals in medical informatics (as determined by impact factor), we computed their *h*-index using the Scopus database. Percentiles were computed to create a 6-level benchmark, similar in scheme to one used by the US National Science Foundation, and a 10-level benchmark.

Results: The 2 benchmarks can be used to place medical informatics researchers in an ordered category based on the performance of their peers. A validation exercise mapped the benchmark levels to the ranks of medical informatics academic faculty in the United States. The 10-level benchmark tracked academic rank better (with no ties) and is therefore more suitable for practical use.

Conclusions: Our 10-level benchmark provides an objective basis to evaluate and compare the publication performance of medical informatics researchers with that of their peers using the *h*-index.

(J Med Internet Res 2012;14(5):e144) doi:10.2196/jmir.2177

KEYWORDS

h-Index; medical informatics; bibliometrics; evaluation; research output

Introduction

Publication metrics, such as the impact factor of journals and the number of citations to papers, are often used directly or indirectly to evaluate the performance of researchers for hiring, promotion, and funding decisions [1-7]. For example, the US National Institutes of Health has developed an electronic Scientific Portfolio Assistant linked to publication metrics [8,9] (such as impact factor and number of citations) and is used by National Institutes of Health staff to “make close-call funding decisions on individual grants” [10]. Similarly, some Wellcome Trust panels have used impact factor and applicant citation data to make grant funding decisions [7]. Publication metrics are also used to evaluate research institutions [11-13] and assess the impact of biomedical research funding policies and programs [9,10,14-16].

Direct comparisons of researchers from different backgrounds and disciplines on publication metrics can be quite misleading [17-21]. This can be a challenge for medical informatics in that it is generally considered a multidisciplinary field [22-26]. For example, one analysis of the US National Library of Medicine's Medical Subject Headings (MeSH) that were used for medical informatics articles identified clusters indexed by terms related to the science and art of medicine, molecular genetics, statistical analysis, immunology, and biochemical communications [25]. A citation analysis of medical informatics articles found that work in general medicine journals was often cited [22].

The comparability problem is demonstrated in Table 1, which shows the average number of citations per paper over a 10-year period for a variety of disciplines [18]. There is an almost 10-fold difference in the average number of citations per paper between a researcher in molecular biology and genetics, and a researcher in computer science. Consider a computer scientist who, with a mean of 5 citations to her papers, would be considered an above-average performer on that metric (for a computer scientist) but, when compared with a basic scientist with average performance she would be assessed quite poorly. Given that both a computer and a basic scientist can be medical informatics researchers and possibly affiliated with the same institution or department, there is a need for methods to evaluate and interpret their publication metrics that allow fair and meaningful comparisons with their medical informatics peers.

Table 1. Average number of citations per paper between 1995 and 2005 by discipline [18].

Discipline	Average citations per paper
Clinical medicine	10.58
Computer science	2.49
Economics and business	4.17
Engineering	3.17
Mathematics	2.66
Molecular biology and genetics	24.57
Neuroscience and behavior	16.41
Pharmacology and toxicology	9.4
Psychiatry and psychology	8.24
Social sciences, general	3.46

The h-Index

One of the more-commonly used metrics to evaluate the publication performance of researchers is the *h*-index [27]. This was first proposed and defined by Hirsch in 2005 as follows: "A scientist has an index *h* if *h* of his or her *Np* papers have at least *h* citations each and the other (*Np* – *h*) papers have ≤*h* citations each" [27]. Hirsh designed the *h*-index to avoid the problems of other common bibliometrics, such as the total number of papers, total number of citations, number of citations per paper, number of significant papers with >*y* citations (*y* is determined by the evaluator), and number of citations to each of the *q* most-cited papers (*q* is determined by the evaluator). The *h*-index measures the impact of an individual's output rather than the volume, controls for the effect of a small number of highly cited papers, rewards consistent output, and is less arbitrary than measures for number of significant papers or number of citations to the *q* most-cited papers [27]. Its ease of use is also a benefit, as it is a single number that is simple to calculate using readily available databases that provide citation counts. Another advantage is that the *h*-index has been shown to predict the impact a researcher will make in the future. In a 2007 study by Hirsh, the predictive power of *h* was compared with that of 3 other bibliometrics: total number of papers, total number of citations, and mean number of citations per paper [28]. It was found that authors' *h*-index scores after the first 12 years of publication were best able to predict performance in the subsequent 12-year period, as well as cumulative achievement over the entire 24-year period. A high correlation has also been found between an individual's *h*-index and his or her receipt of academic awards, appointments, and funding [17,29]. A person's *h*-index has also been found to be relatively unaffected by normal citation record errors—a 2010 review reported that *h*-scores before and after the correction of errors remained stable [29].

Many variations of and alternatives to the *h*-index have been proposed since 2005 [17,29-31], for example, to give more weight to highly cited papers [30], incorporating the variables of total number and age of citations [29] and allowing comparability across disciplines [31].

However, many of the subsequent variations proposed have been shown to be highly correlated with the *h*-index, and hence do not provide much additional information [32], and each variation increases the complexity and computational difficulty of the metric. As noted in a recent review, "many *h*-index variations, although being designed to overcome some of its supposed limitations, do indeed correlate quite heavily. This fact has made some researchers think that there is probably no need to introduce more *h*-index variations if it is not possible to prove that they are not redundant in real examples" [17]. Aided by the inclusion of automated *h*-index calculators in popular interdisciplinary databases, use of the *h*-index continues to grow [17].

A more-detailed critical review of the *h*-index and its measurement is provided in Multimedia Appendix 1.

Uses and Interpretation of the h-Index

In the basic, natural, and applied sciences, there has been a trend toward objective performance evaluations of researchers for hiring, promotion, and funding decisions using bibliometrics, including the *h*-index [1-6]. In the health sciences, individuals, departments, and institutions have been compared using their *h*-index scores [21,33-39]. The *h*-index has also been used in medical informatics to evaluate the quality of panel sessions at the annual American Medical Informatics Association symposium [40] and to evaluate the national influence of medical informatics research [41].

Traditionally, subjective peer evaluations have been used as the main method to evaluate researcher performance. There is evidence that the *h*-index scores correlate well with peer assessments [42-45]. However, a case can be made for using the *h*-index to inform the peer-review decision-making process, which can arguably enhance interreviewer reliability (see [46]).

Proper interpretation of publication metrics requires a robust reference standard [47], and there is none for the *h*-index in the area of medical informatics. Given the relative advantages of the *h*-index as an objective measure of scientific output, such a standard is needed for the *h*-index to be used effectively. A defensible standard to interpret the *h*-index can help accelerate its adoption in medical informatics and allow objective, repeatable, and fair evaluations and comparisons of researchers.

In the past, different types of reference standards for the *h*-index have been constructed in other disciplines. Examples include using the mean for multiple scientific disciplines [18], by rank of radiology academics [33], by rank of neurosurgery academics [21], by comparison with chairs of medical departments in US medical schools [34], as a median for academic anesthesiologists [38,39], by rank of academic anesthesiologists [37], and by rank for academic urologists [48].

Because citation and publication distributions are known to be heavily skewed [49,50], reference standards based on percentiles have been recommended [51]. In this study we developed 2 percentile-based benchmarks for interpreting the value of the *h*-index for researchers who publish in medical informatics journals and validated the benchmarks using an independent measure of performance.

Methods

Our objective was to develop appropriate *h*-index benchmarks for researchers who publish in medical informatics journals.

Requirements for Benchmarks

We considered the following as requirements to maximize the utility of the benchmarks. We therefore used them to guide the methodological decisions made during their construction:

R1. The benchmarks should allow for the evaluation of researchers' performance at multiple stages in their careers. This means that the benchmarks should have sufficient granularity and variation to reflect the performance of early career investigators as well as more-established researchers.

R2. The benchmarks need to be contemporary, reflecting the performance of researchers at this point of time rather than serving as a historical analytical tool.

R3. The benchmarks should reflect the performance of researchers who publish influential work rather than work that does not get cited often.

We describe below how we made the tradeoffs necessary to meet these requirements.

Author Sampling Frame

We intended the benchmarks to apply to individuals who are considered *medical informatics researchers*. One approach to identifying medical informatics researchers is to use a subset of prominent individuals in the field, such as American College of Medical Informatics (ACMI) fellows. This approach has an important disadvantage in that ACMI fellows are not necessarily representative of the rest of the medical informatics researchers and would therefore not meet requirement R1 above because (1) they have higher *h*-index values than other researchers, and (2) they constitute a very small fraction of all medical informatics researchers.

While constructing benchmarks based only on ACMI fellows would meet requirement R3, this group has higher *h*-index values and would not be a good representation of all medical informatics researchers. To confirm this, we compared the *h*-index values for fellows with those who were not fellows in a simple random sample of 430 authors from all authors in the top (by impact factor) 2 medical informatics journals as classified by *Journal Citation Reports*: the *Journal of Medical Internet Research* (JMIR) and the *Journal of the American Medical Informatics Association* (JAMIA), with 5-year impact factors of 4.98 and 3.94, respectively (according to the Web of Knowledge). Fellows had a mean *h*-index value of 16.5 versus 8.8 for the nonfellows. This difference was statistically significant ($P < .05$ for a 2-tailed t test, and $P < .05$ for a Mann-Whitney nonparametric test).

A very small proportion of authors in medical informatics journals are ACMI fellows. This is because there are only 338 fellows, they do not all publish in medical informatics journals, and many of the fellows are now retired or deceased and are no longer actively publishing. Table 2 shows the percentage of authors in some of the *Journal Citation Reports* medical informatics journals who were ACMI fellows during the period 2006-2011. The first number in the table is the maximum possible and assumes that all ACMI fellows publish in that journal. The second number is based on an exact match of the names of fellows and authors in the journal, and the third number is based on an approximate match of the names using the Jaro-Winkler string comparison distance [52] (with a cut-off of 0.9). The exact match rate is likely to have a higher false-negative rate and the latter, with the approximate matching, a higher false-positive rate. Therefore, the correct match rate is expected to be within these two values. Across all journals, the fellows account for 0.5%-0.68% (using exact and approximate matching, respectively) of all authors.

Table 2. Percentage of American College of Medical Informatics fellows who published in some of the *Journal Citation Reports* medical informatics journals over the period 2006–2011.

Journal name	Maximum match (%)	Exact match (%)	Approximate match (%)
Journal of the American Medical Informatics Association (JAMIA)	12.8	4.63	5.62
Journal of Medical Internet Research (JMIR)	23.5	0.9	1.32
IEEE Engineering in Medicine and Biology	37.3	0	0.22
Artificial Intelligence in Medicine	28.8	0.94	1.45
BMC Medical Informatics and Decision Making	21.5	1.4	2.1
Computers, Informatics, Nursing	34.5	1.43	2.14
Computer Methods and Programs in Biomedicine	11.5	0.24	0.37
IEEE Transactions on Information Technology in Biomedicine	12.7	0.07	0.22
International Journal of Medical Informatics (IJMI)	15.5	2.65	3.52
International Journal of Technology Assessment in Health Care	21.5	0.06	0.25
Journal of Biomedical Informatics (JBI)	16.9	3.4	4.35
Journal of Medical Systems	14.8	0.13	0.17
Medical & Biological Engineering & Computing	11.3	0	0.1
Medical Decision Making	21.4	0.63	1
Methods of Information in Medicine	18.7	0.33	2.77
Statistics in Medicine	8.1	0.17	0.38
Statistical Methods in Medical Research	45.3	0.13	0.4

Another approach to constructing a sampling frame is to identify all authors who publish in medical informatics journals over a specified time period and consider them to be medical informatics researchers. Various approaches have been used in the literature to identify a core set of medical informatics journals, which we review below.

A bottom-up method uses index terms in article databases to identify journals [22,23,53,54]. For example, some studies used MeSH terms for medical informatics concepts. However, a recent analysis found that the journals that published the majority of papers classified in this way had central concepts outside of medical informatics and were "not typically identified as medical informatics-specific journals," such as physics, imaging, and engineering journals [23]. Therefore, this approach would not have strong face validity and is unlikely to be intuitively convincing.

A variant of that approach is to identify the journals with the relatively most-cited articles that are indexed or classified under medical informatics [22]. However, many of the journals with the most-cited articles were general medicine or general science journals, since these journals tend to have quite high average citations per article. A survey of ACMI fellows found that general medicine and general science journals ranked lower in terms of readership than did journals more typically associated with medical informatics [55]. Again, our benchmark would not pass the face validity test if it were based on publications that have a low readership among the most experienced members of the community.

Some authors subjectively decide on an initial set of medical informatics journals to study [24-26,55,56] or ask ACMI fellows to rank or rate journals [55,57]. Others use existing classifications of journals, such as the *Journal Citation Reports* [58]. Sometimes multiple approaches are used [57]. The journals in which prominent members of the community publish have also been used as the core set of medical informatics journals, such as the most-cited ACMI fellows [58].

Our approach to identifying source journals for selecting medical informatics researchers was informed by the methods used in the literature. We selected the list of journals in the *Journal Citation Reports* in the medical informatics category. This is consistent with previous approaches [57,58]. We identified the top-2 ranked journals by impact factor at the time of writing: JMIR and JAMIA. If we had considered other definitions of "core medical informatics journals" [24,57], these 2 journals would still have had the highest impact factors among journals in those sets.

We considered all authors who published more than 1 article over the 2006–2011 period in any of the 17 journals in Table 2. Approximately 70%–77% (using exact and approximate matching, respectively) of these JMIR and JAMIA authors had also published at least one paper in one of the other 15 journals. While the choice of JMIR and JAMIA authors seemingly limited our analysis to those who published the articles that were most cited, in fact there is still significant community overlap with other journals.

By defining the sampling frame to consist of all authors in JMIR and JAMIA, plus the overlap in authorship with other journals, we met requirement R1. Requirement R3 was also met because these 2 journals have the highest impact factors.

Sampling Period

We considered the *h*-index of authors who had published in the 5-year period 2006-2011 in these 2 journals: JMIR 2006;8(1) to 2011;13(2) and JAMIA 2006;13(1) to 2011;18(5). We chose the 5-year period because it is quite common for studies evaluating scholars, institutions, and programs to examine the previous 5 years' data on publications [59-68], and previous studies assessing the structure of the medical informatics literature and community have often used 5-year intervals [26,58]. In addition, a longer period would likely include more researchers who may no longer be as active in the field, hence reducing the benchmarks' representativeness of current researchers, and would therefore not meet requirement R2 above.

Author Order

In addition to constructing a benchmark based on all authors within the sampling frame and sampling period, we could construct benchmarks based on first authors only. However, there is a lack of consistency across disciplines in how authors are ordered. For example, in some cases authors are ordered alphabetically or by the extent of contribution. This makes it difficult to determine when a first-author benchmark should be used. Furthermore, there is evidence of a strong correlation between the *h*-index values based on being a first author and those ignoring author order [69]. We therefore constructed benchmarks ignoring author order.

Benchmark Levels

In general, a reference standard for publication metrics using percentiles has been recommended [70], and specifically one based on deciles [71]. We refer to a decile benchmark as PR10. A PR10 benchmark provides a 10-level scale based on deciles, where a level 10 means that a researcher is in the top 10% among his or her peers in terms of his or her *h*-index value, and a level 1 means that a researcher is in the bottom 10% among his or her peers. We deemed benchmarks with fewer levels, such as 5 levels based on quintiles, to be too coarse, as they would provide inadequate granularity to assess researchers' publication output as they move through their career stages and would therefore not meet requirement R1.

Another evaluation scheme used in the Science and Engineering Indicators by the US National Science Foundation has only 6 percentile levels, which we refer to as PR6 [72]. PR6 focuses on authors with *h*-index values that are higher than the median and on benchmarking more-established researchers.

Calculation of h-Index

For our computation of the *h*-index we used Scopus. We manually checked the identity of all authors in Scopus by confirming their affiliation and comparing their listed publications with personal or academic websites. Scopus also uses specific algorithms [73] to combine all references from each author under a unique ID, which means that the time needed to manually match references to author names is reduced. Furthermore, Scopus has been shown to have a more-accurate automated calculator for the *h*-index (vs Web of Science]) [74], to include more peer-reviewed conference proceedings than Web of Science [75] and to avoid the problems of duplicate records and false-positives associated with Google Scholar [74-76].

The *h*-index not only is computed from articles published in the 2 top medical informatics journals (JMIR and JAMIA) from which we sampled authors, but also covers all publications by these researchers in all journals and conferences that are indexed, going back to when indexing started. In the case of Scopus, indexing started from 1996. For example, if selected medical informatics researchers also published papers in general medicine journals in 2000, their *h*-index would include their publications in the general medicine journals from that year.

Sample Size Calculation

We manually computed the *h*-index for medical informatics researchers using Scopus; therefore, rather than computing it for all authors, we decided to estimate it from a random sample. To ensure we had a representative sample of the underlying population, we decided to use a strict quantitative definition of representativeness. This definition also provides insight into the confidence intervals we can expect, based on the uncertainty we chose in sampling, therefore estimating the true unknown distribution of the *h*-index. We will use the PR10 benchmark to describe the method.

Following Sobel and Huyett [77], we selected a random sample of authors based on a nonparametric definition of representativeness. Namely, our sample of authors should be simultaneously representative of the true unknown cumulative distribution F of the *h*-index for deciles. That is, we divide the *h*-index into 10 pairwise disjoint subsets that we denote $C_1,...,C_{10}$. These subsets are unknown but have probability under F greater than zero, and in the case of equiprobable deciles are given by $F(C_i) = 0.10$, or $i = 1,...,10$ (where F as used here is a probability measure).

For an observed cumulative sample distribution F^*_n based on n observations, we say that a sample is representative relative to the fixed disjoint subsets $C_1,...,C_{10}$ to within the common allowance $ß^*$ if we have $|F^*_n(C_i) - F(C_i)| \leq ß^*$ simultaneously for $i = 1,...,10$. The degree of representativeness of F under this setup is subsequently defined as $d^*_g = 1 - 10ß^*$ for deciles. In particular, we consider a common allowance $ß^* = 0.05/2 = 0.025$, as it is a standard threshold used in 1-sided definitions, resulting in a degree of representativeness of $d^*_g = 0.75$ for deciles.

Next we choose a probability P^* that the sample will at least have the degree of representativeness of F that we selected. That is, $P(|F^*_n(C_i) - F(C_i)| \leq ß^*$ for all $i = 1,...,10) \geq P^*$, where we seek the sample size n needed to satisfy this equation. Note that larger sample sizes may not satisfy the inequality, and it is therefore necessary to treat this as an approximation only. For the sample sizes we consider, however, the duplicates in the original author list should decrease the author count such that the sample is inflated enough to meet the probability threshold we desire.

Determining the sample size n involves using a multinomial distribution for an infinite population and a hypergeometric distribution for a finite population. Sobel and Huyett [77] provide tables that give the sample size required under a variety of circumstances. Note that for a fixed number of pairwise disjoint subsets C_i and fixed allowance $ß^*$, the greater the probability P^*, the greater the required sample size. Let N be the population size, n_∞ the sample size for an infinite population, and n_N the sample size for a finite population. Then, given n_∞ from the tables, the required sample size can be adjusted using the approximation $n_\infty \approx n_N((N - 1) / (N - n_N))$.

We identified N = 3220 total authors (regardless of position in the author list) and chose a probability of simultaneous representativeness $P^* = .75$ for deciles ($ß^* = 0.025$, $d^*_g = 0.75$), resulting in a minimum sample size of $n_{3220} = 430$ (where $n_\infty = 500$ for an infinite population).

Sample of Authors

We extracted article names, author names, and journal edition for JMIR and JAMIA over the 5 years of journal issues. We drew a simple random sample of 430 from the 3220 authors regardless of position in the author list, excluding correspondences or letters, editorials, errata, highlights, and articles without a designated author.

Results

The distribution of *h*-index values from our sample is shown in Figure 1. We fitted a kernel density estimate with a normal kernel function and a bandwidth that minimizes the approximate mean integrated square error between the density estimator and the true density. As can be seen, the distributions have a long tail. This means that the maximum *h*-index value will be significantly larger than the 90th percentile value (level 10 in PR10) and the 99th percentile in PR6. For example, the highest value for an author in our sample was 74, but the point estimate of the 90th percentile for our benchmarks was approximately 23. Therefore, authors with values significantly larger than the 90th percentile were all in the top 10% of medical informatics researchers. Since percentiles are a ranking, the actual *h*-index value for a level-10 author can be much higher than the 90th percentile value.

Percentile estimates of the *h*-index of authors who published in JAMIA or JMIR over the 5 years we examined are given in Table 3. Confidence intervals were calculated using the Woodruff method, which inverts the confidence intervals for the estimated distribution function (first proposed Woodruff [78] and further justified by Francisco and Fuller [79], Dorfman and Valliant [80], Sitter and Wu [81] and Chatterjee [82]).

We will consider an example to illustrate how to interpret Table 3. A medical informatics researcher with an *h*-index of 21 could be said to be in the 90th percentile of his or her peers in medical informatics, since that value falls right within the confidence interval. Any *h*-index value as high as 25 would still be in the 90th percentile. A researcher with a value in the range 17–20 is above the 80th percentile (since 17 is larger than the upper confidence limit for the 80th percentile), but not at the 90th percentile. To move beyond the 90th percentile, that researcher would need an *h*-index value of 26 or higher.

Table 4 provides the *h*-index benchmark values that would indicate statistically significant values for each of the levels. These can be used to directly determine the level of researchers based on their *h*-index values.

Table 3. *h*-Index percentile estimates for authors published in the *Journal of the American Medical Informatics Association* (JAMIA) or *Journal of Medical Internet Research* (JMIR) over the 5-year period 2006–2011.

PR6[a]			PR10[b]		
Percentile	Estimate	95% CI[c]	Percentile	Estimate	95% CI
<50%			<10%		
50%	4.9	4.2–5.6	10%	0.6	0.5–0.6
75%	12.6	10.9–14.3	20%	1.4	1.1–1.6
90%	22.9	20.2–25.6	30%	2.3	1.9–2.7
95%	28.9	26–31.7	40%	3.4	2.9–3.9
99%	48.5	32.2–64.9	50%	4.9	4.2–5.6
			60%	7.5	6.3–8.7
			70%	10.8	9.3–12.3
			80%	14.8	13.1–16.5
			90%	22.9	20.2–25.6

[a] 6 Percentile-level benchmark.

[b] Decile benchmark.

[c] Confidence interval.

Table 4. *h*-Index benchmarks for authors published in the *Journal of the American Medical Informatics Association* (JAMIA) or *Journal of Medical Internet Research* (JMIR) over the 5-year period 2006–2011.

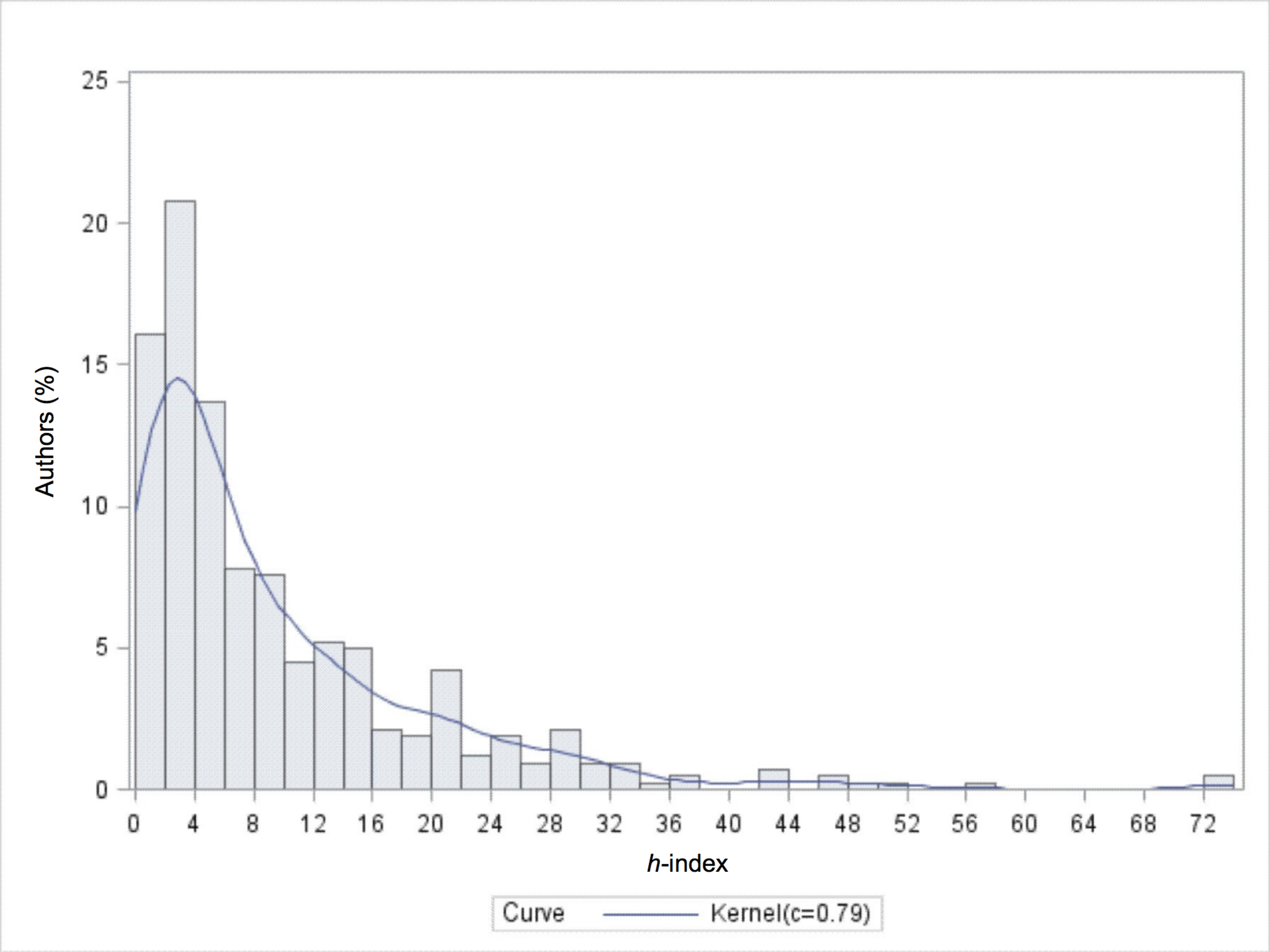
Level	Benchmark	
	PR6[a]	PR10[b]
1	0–5	0
2	6–14	1
3	15–25	2
4	26–31	3
5	32–64	4–5
6	≥65	6–8
7		9–12
8		13–16
9		17–25
10		≥26

[a] 6 Percentile-level benchmark.

[b] Decile benchmark.

To explain how Table 4 was constructed, we take as an example the 50th percentile. Here the upper confidence limit is 5.6. Keeping in mind that *h*-index values can only be integers, any *h*-index value that is 5 or less will be in the lower 50% of all authors. Similarly, if we take the 75th percentile in PR6, any *h*-index value that is 14 or less will be in the bottom 75% of all authors. Consequently, any value that is higher than 5 and equal to or less than 14 will be in the percentile range greater than 50% and less than or equal to 75% (in the third quartile). This is the 6–14 range shown in Table 4.

Figure 1. Distribution of *h*-index values from a sample of 430 authors.



Discussion

Summary

The objective of this study was to construct empirical benchmarks for interpreting the *h*-index for medical informatics researchers. We had specified three requirements to ensure the utility of the benchmarks to a large segment of the community: (1) they allow the tracking of career progress for researchers, (2) the benchmarks should be contemporary, reflecting current work in the field, and (3) the benchmarks should be based on researchers doing work that is cited often. The values we computed represent all publications by authors going back to 1996.

The benchmarks allow for the evaluation of researchers relative to their peers. These peers come from a mix of other disciplines, but they do represent the contemporary medical informatics researcher community.

More precisely, we have provided 2 empirical benchmarks that are slightly different. The first, PR6, uses a National Science Foundation scheme that has 6 percentile levels, and the second, PR10, is a broader 10-level benchmark based on deciles. The PR10 benchmark allows for the evaluation of performance of researchers at early and late stages of their careers. The PR6 benchmark is focused on the top half of performers.

Validation

To validate the benchmarks, we examined the relationship between the *h*-index and some independent measure or ranking of researcher performance. One option was to rank scientists based on the number of recognitions and awards they have received (eg, the Nobel Prize) and the number of degrees they have received, and then determine whether our benchmarks reflect that ordering [83]. A proxy for achievements of researchers is their academic faculty rank.

We examined whether our benchmarks track the mean *h*-index value of medical informatics faculty in the United States. Several studies have used the mean (or median) *h*-index of faculty ranks to characterize performance levels within a discipline [21,33,34,37-39,48].

We identified medical informatics departments in the United States using the list of medical informatics departments funded by the National Library of Medicine under the University-based Biomedical Informatics Research Training Programs [84], augmented with medical informatics departments listed in the Open Directory Project [85]. For each department we manually identified all faculty at any of the three ranks listed on their websites: assistant professor, associate professor, and full professor, for a total of 463 individuals. We then selected using simple random sampling 50 from each rank and computed their mean *h*-index and the confidence interval for the mean. The results are shown in Figure 2. There is greater variation in performance as the rank increases.

Given that the confidence intervals for the three ranks do not overlap, the differences in the mean *h*-index are statistically significant. Furthermore, our PR10 benchmark levels track the mean values by faculty rank well as seen in Table 5. This provides validation that the PR10 benchmark can be a useful tool for assessing the scientific performance of medical informatics faculty. For example, full professors in level 10 (top decile) on our PR10 benchmark would be above the average for US medical informatics faculty. On the other hand, a researcher with an *h*-index of 4 would be on level 5 of the PR10 benchmark and would therefore be below the mean for an assistant professor. However, such a level could be considered an acceptable target for someone completing a postdoctoral fellowship, for instance.

These results also indicate that the PR6 benchmark does not track medical informatics faculty rank very well, since both associate and full professors would be within the same level. This is due to PR6 making finer distinctions at the top end of the distribution and coarser distinctions otherwise. This results in multiple faculty ranks grouped into the same performance level. Therefore, one can argue that for practical purposes the PR10 benchmark is a more-useful tool for assessing and tracking performance of medical informatics academic faculty.

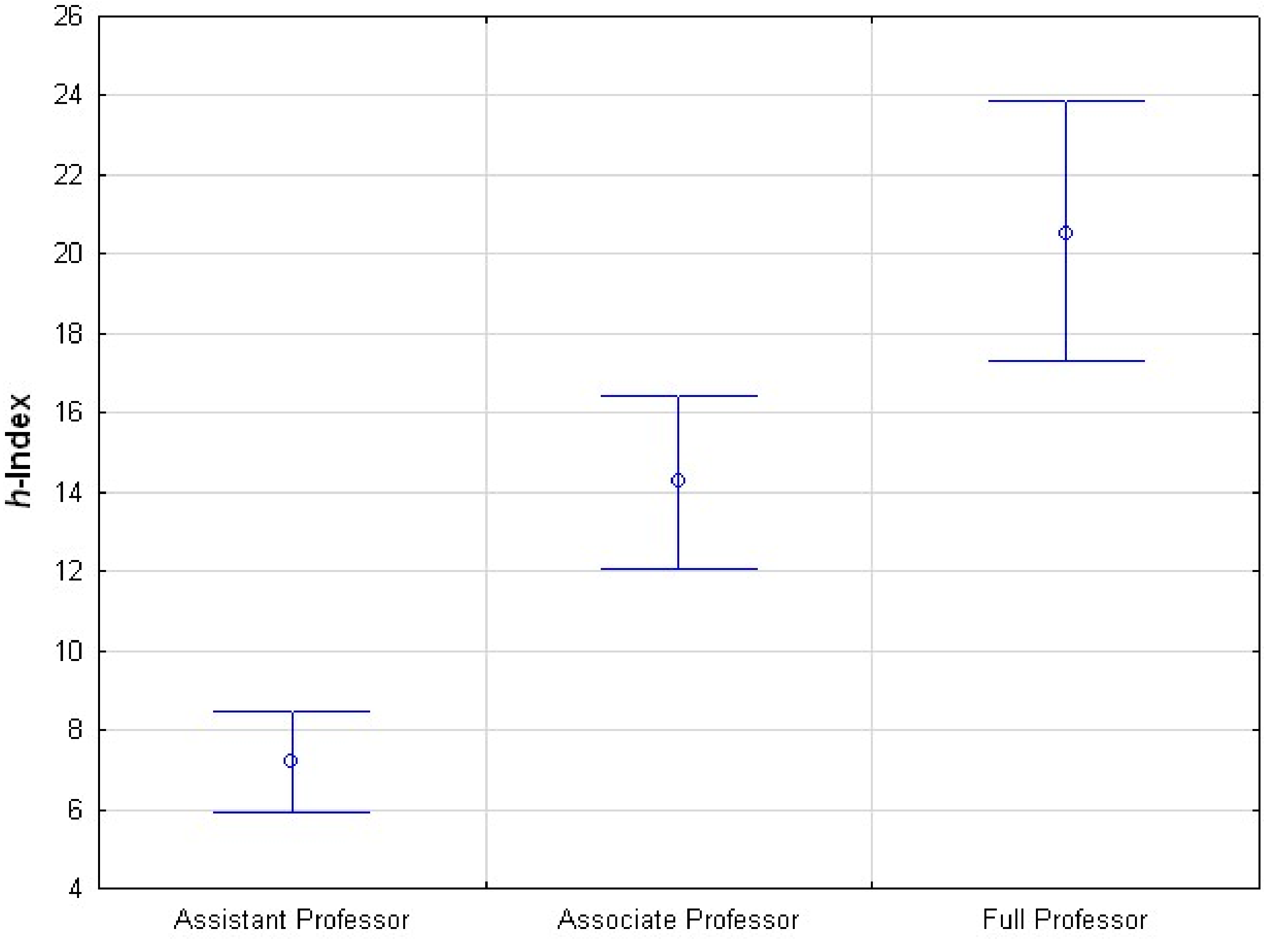
Table 5. Benchmark levels based on the mean *h*-index for the three US medical informatics academic faculty ranks.

Rank of faculty in US medical informatics departments	PR6 level[a]	PR10 level[b]
Assistant professor	2	6
Associate professor	3	8
Full professor	3	9

[a] 6 Percentile-level benchmark.

[b] Decile benchmark.

Figure 2. Mean *h*-index and 95% confidence interval for the three faculty ranks in US medical informatics departments.



Interpretation and Use of the Benchmarks

Although medical informatics researchers come from multiple disciplines, the PR10 benchmark applies to the subset of individuals from these disparate disciplines who do medical informatics research. If the vast majority of medical informatics researchers were basic scientists, for example, then the benchmark would tilt more toward the performance of basic scientists. Similarly, if the vast majority of medical informatics researchers were computer scientists, then the benchmark would tilt toward the publication performance of that community.

The benchmark would be meaningful only for researchers whose area of research is clearly medical informatics. For example, a clinical researcher may have a high ranking on PR10, but this would not be relevant unless that individual has medical informatics as a primary area of work.

We suggest three scenarios where the PR10 benchmark can be useful. Researchers at early stages of their careers would be expected to be at the lower levels on the benchmark and to progress to higher levels over time. Therefore, the benchmark can be used as a yardstick to track performance over time. For research leaders, administrators, or funding agencies evaluating researchers at the same career stage competing for positions, funding, or promotions, the PR10 benchmark can be used to assess their relative standing in terms of scientific output. For instance, researchers with scores of 6 and 7 are in the same decile and may be considered equal on the *h*-index metric. Finally, the PR10 benchmark can be used to set objective gates (see Table 5), such as for hiring or promotion decisions.

In practice, the PR10 benchmark could replace or augment metrics such as the number of published papers or number of citations when assessing performance. It should not be used as the sole method for evaluating the publication performance of medical informatics researchers, but can serve as another useful input for such an evaluation. Furthermore, the PR10 benchmark would need to be updated on a regular basis to ensure that it reflects contemporary performance levels in the field.

Limitations

An underlying assumption of our method is that medical informatics researchers will at some point publish in medical informatics journals (they only need to publish once in a top medical informatics journal to be in our sampling frame). For example, an author who has published medical informatics papers only in general medicine or general science journals, or who has published in conferences only but never in a top medical informatics journal, would not be in our sampling frame. Such an individual, however, would also likely not be considered to have medical informatics as a primary area of his or her research.

Our results are limited by the journals we selected from which to sample researchers. It is possible that a different set of journals would have produced different values for the benchmarks because they would have included a different group of researchers. However, we have argued that our choice of journals balances representativeness of the community and covers authors who publish influential work in the field.

While we used Scopus to compute our benchmark, one can argue that the use of another tool, such as Web of Science, may have produced different results. For example, Scopus indexes publications only since 1996. This would not account for citations to earlier research articles. On the other hand, medical informatics is a recent discipline, with JAMIA starting publication in 1994 and JMIR in 1999. Furthermore, there is evidence that Web of Science and Scopus produce very similar citation counts [86], which would also mean very similar *h*-index values.

Acknowledgments

We wish to thank Brian Dewar for his assistance with data collection. We also would like to thank the reviewers for their thoughtful feedback on an earlier version of this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

A critical review of the h-index and its measurement.

[PDF File (Adobe PDF File), 70KB - jmir_v14i5e144_app1.pdf]

References

1. Abbott A, Cyranoski D, Jones N, Maher B, Schiermeier Q, Van Noorden R. Metrics: Do metrics matter? Nature 2010 Jun 17;465(7300):860-862. [doi: 10.1038/465860a] [Medline: 20559361]
2. Académie des Sciences. Annex 2: Evaluation of individual researchers in theoretical and experimental sciences. In: On the Proper Use of Bibliometrics to Evaluate Individual Researchers. Paris: Académie des Sciences, Institut de France; 2011:31-36.
3. Académie des Sciences. Annex 3: Bibliometric practices by discipline. In: On the Proper Use of Bibliometrics to Evaluate Individual Researchers. Paris: Académie des Sciences, Institut de France; 2011:37-47.
4. European Research Council. 2011. Supporting Top Researchers From Anywhere in the World URL: http://erc.europa.eu/ [accessed 2012-10-07] [WebCite Cache ID 6BFC6fUjn]
5. European Commission. Assessing Europe's University-Based Research. Luxembourg: European Commission; 2010.
6. Patel VM, Ashrafian H, Ahmed K, Arora S, Jiwan S, Nicholson JK, et al. How has healthcare research performance been assessed?: a systematic review. J R Soc Med 2011 Jun;104(6):251-261. [doi: 10.1258/jrsm.2011.110005] [Medline: 21659400]
7. Lewison G, Cottrell R, Dixon D. Bibliometric indicators to assist the peer review process in grant decisions. Res Eval 1999;8(1):47-52. [doi: 10.3152/147154499781777621]
8. Fitzpatrick S. Thomson Reuters. 2009 Aug 3. National Institutes of Health Chooses Web Of Science® to Power Electronic Scientific Portfolio Assistant (e-SPA) URL: http://thomsonreuters.com/content/press_room/science/468364 [accessed 2012-10-07] [WebCite Cache ID 6BFD5G6yD]
9. Bertuzzi S. Behind Medicine: Science and Policy. A View From the US National Institutes of Health on How Return on Investment Analysis Can Inform Science Policy. Brussels: European Parliament; 2010.
10. Wadman M. Study says middle sized labs do best. Nature 2010 Nov 18;468(7322):356-357. [doi: 10.1038/468356a] [Medline: 21085145]
11. Hendrix D. An analysis of bibliometric indicators, National Institutes of Health funding, and faculty size at Association of American Medical Colleges medical schools, 1997-2007. J Med Libr Assoc 2008 Oct;96(4):324-334. [Medline: 18979684]
12. National Research Council (US) Committee on an Assessment of Research Doctorate Programs; Ostriker J, Holland P, Kuh C, Voytuk J, editors. In: Ostriker JP, Kuh CV, Voytuk JA, editors. A Data-Based Assessment of Research-Doctorate Programs in the United States (National Research Council). Washington, DC: National Academies Press; 2011.
13. Oppenheim C. The correlation between citation counts and the 1992 research assessment exercise ratings for British research in genetics, anatomy and archaeology. J Doc 1997;53(5):477-487. [doi: 10.1108/EUM0000000007207]
14. Boyack KW, Jordan P. Metrics associated with NIH funding: a high-level view. J Am Med Inform Assoc 2011;18(4):423-431 [FREE Full text] [doi: 10.1136/amiajnl-2011-000213] [Medline: 21527408]
15. Lyubarova R, Itagaki BK, Itagaki MW. The impact of National Institutes of Health funding on U.S. cardiovascular disease research. PLoS ONE 2009;4(7):e6425 [FREE Full text] [doi: 10.1371/journal.pone.0006425] [Medline: 19641617]
16. US National Science Foundation. Research Performance Progress Report (RPPR) URL: http://www.nsf.gov/bfa/dias/policy/rppr/ [accessed 2012-09-28] [WebCite Cache ID 6B1O4gYWb]

17. Alonso S, Cabrerizo FJ, Herrera-Viedma E, Herrera F. h-Index: a review focused in its variants, computation and standardization for different scientific fields. J Informetr 2009;3:273-289. [doi: 10.1016/j.joi.2009.04.001]
18. Iglesias JE, Pecharromán C. Scaling the h-index for different scientific ISI fields. Scientometrics 2007;73:303-320. [doi: 10.1007/s11192-007-1805-x]
19. Radicchi F, Fortunato S, Castellano C. Universality of citation distributions: toward an objective measure of scientific impact. Proc Natl Acad Sci U S A 2008 Nov 11;105(45):17268-17272 [FREE Full text] [doi: 10.1073/pnas.0806977105] [Medline: 18978030]
20. Podlubny I. Comparison of scientific impact expressed by the number of citations in different fields of science. Scientometrics 2005;64:95-99. [doi: 10.1007/s11192-005-0240-0]
21. Lee J, Kraus KL, Couldwell WT. Use of the h index in neurosurgery. Clinical article. J Neurosurg 2009 Aug;111(2):387-392. [doi: 10.3171/2008.10.JNS08978] [Medline: 19392590]
22. Vishwanatham R. Citation analysis in journal rankings: medical informatics in the library and information science literature. Bull Med Libr Assoc 1998 Oct;86(4):518-522. [Medline: 9803294]
23. Deshazo JP, Lavallie DL, Wolf FM. Publication trends in the medical informatics literature: 20 years of "Medical Informatics" in MeSH. BMC Med Inform Decis Mak 2009;9:7 [FREE Full text] [doi: 10.1186/1472-6947-9-7] [Medline: 19159472]
24. Morris TA, McCain KW. The structure of medical informatics journal literature. J Am Med Inform Assoc 1998;5(5):448-466 [FREE Full text] [Medline: 9760393]
25. Morris TA. Structural relationships within medical informatics. Proc AMIA Symp 2000:590-594. [Medline: 11079952]
26. Malin B, Carley K. A longitudinal social network analysis of the editorial boards of medical informatics and bioinformatics journals. J Am Med Inform Assoc 2007;14(3):340-348 [FREE Full text] [doi: 10.1197/jamia.M2228] [Medline: 17329730]
27. Hirsch JE. An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S A 2005 Nov 15;102(46):16569-16572 [FREE Full text] [doi: 10.1073/pnas.0507655102] [Medline: 16275915]
28. Hirsch JE. Does the H index have predictive power? Proc Natl Acad Sci U S A 2007 Dec 4;104(49):19193-19198 [FREE Full text] [doi: 10.1073/pnas.0707962104] [Medline: 18040045]
29. Norris M, Oppenheim C. The h-index: a broad review of a new bibliometric indicator. J Doc 2010;66(5):681-705. [doi: 10.1108/00220411011066790]
30. Egghe L. An improvement of the h-index: the g-index. ISSI Newsl 2006;2(1):8-9.
31. Antonakis J, Lalive R. Quantifying scholarly impact: IQp Versus the Hirsch h. J Am Soc Inf Sci Technol 2008;59(6):956-969. [doi: 10.1002/asi.v59:6]
32. Bornmann L, Mutz R, Hug SE, Daniel HD. A multilevel meta-analysis of studies reporting correlations between the h index and 37 different h index variants. J Informetr 2011;5(3):346-359. [doi: 10.1016/j.joi.2011.01.006]
33. Rad AE, Brinjikji W, Cloft HJ, Kallmes DF. The H-index in academic radiology. Acad Radiol 2010 Jul;17(7):817-821. [doi: 10.1016/j.acra.2010.03.011] [Medline: 20471868]
34. Hedley-Whyte J, Milamed DR, Hoaglin DC. Chairpersons of pathology in the United States: limited benchmarks for publications. Am J Clin Pathol 2010 Aug;134(2):185-192 [FREE Full text] [doi: 10.1309/AJCPMVAYD5V8DLZN] [Medline: 20660319]
35. Sypsa V, Hatzakis A. Assessing the impact of biomedical research in academic institutions of disparate sizes. BMC Med Res Methodol 2009;9:33 [FREE Full text] [doi: 10.1186/1471-2288-9-33] [Medline: 19480665]
36. Ponce FA, Lozano AM. Academic impact and rankings of American and Canadian neurosurgical departments as assessed using the h index. J Neurosurg 2010 Sep;113(3):447-457. [doi: 10.3171/2010.3.JNS1032] [Medline: 20380531]
37. Bould MD, Boet S, Sharma B, Shin E, Barrowman NJ, Grantcharov T. h-Indices in a university department of anaesthesia: an evaluation of their feasibility, reliability, and validity as an assessment of academic performance. Br J Anaesth 2011 Mar;106(3):325-330 [FREE Full text] [doi: 10.1093/bja/aeq403] [Medline: 21258072]
38. Pagel PS, Hudetz JA. An analysis of scholarly productivity in United States academic anaesthesiologists by citation bibliometrics. Anaesthesia 2011 Oct;66(10):873-878. [doi: 10.1111/j.1365-2044.2011.06860.x] [Medline: 21864299]
39. Pagel PS, Hudetz JA. H-index is a sensitive indicator of academic activity in highly productive anaesthesiologists: results of a bibliometric analysis. Acta Anaesthesiol Scand 2011 Oct;55(9):1085-1089. [doi: 10.1111/j.1399-6576.2011.02508.x] [Medline: 22092205]
40. Kashyap R, Pickering B, Assudani J, Singh B, Ahmed A, Chute C, et al. H-index of panel session participants from AMIA 2010 annual meeting. 2011 Presented at: AMIA 2011 Annual Symposium; Oct 22-26, 2011; Washinton, DC. USA.
41. Shahbodaghi A, Sajjadi S. A scientometric investigation of the publication trends of Iranian medical informatics articles based on ISI Citation Databases. J Paramed Sci 2010;1(4):2-11.
42. Bornmann L. Daniel H-D. Does the h-index for ranking of scientists really work? Scientometrics 2005;65(3):391-392. [doi: 10.1007/s11192-005-0281-4]
43. Bornmann L, Wallon G, Ledin A. Is the h index related to (standard) bibliometric measures and to the assessments by peers? An investigation of the h index by using molecular life sciences data. Res Eval 2008;17(2):149-156. [doi: 10.3152/095820208X319166]

44. van Raan AFJ. arxiv. 2006. Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups URL: http://arxiv.org/ftp/physics/papers/0511/0511206.pdf [accessed 2012-10-08] [WebCite Cache ID 6BGH6KOJW]
45. Lovegrove BG, Johnson SD. Assessment of research performance in biology: how well do peer review and bibliometry correlate? BioScience 2008 Feb 1;58(2):160-164.
46. Bornmann L. Scientific peer review. Annu Rev Inf Sci Technol 2011;45:199-245.
47. Bornmann L, Mutz R, Neuhaus C. Daniel H-D. Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results. Ethics Sci Environ Polit 2008;8:93-102. [doi: 10.3354/esep00084]
48. Benway BM, Kalidas P, Cabello JM, Bhayani SB. Does citation analysis reveal association between h-index and academic rank in urology? Urology 2009 Jul;74(1):30-33. [doi: 10.1016/j.urology.2008.10.045] [Medline: 19567283]
49. Seglen PO. The skewness of science. J Am Soc Inf Sci 1992;43(9):628-638. [doi: 10.1002/(SICI)1097-4571(199210)43:9<628]
50. Seglen PO. Why the impact factor of journals should not be used for evaluating research. BMJ 1997 Feb 15;314(7079):498-502. [Medline: 9056804]
51. Leydesdorff L, Bornmann L, Mutz R, Opthof T. Turning the tables on citation analysis one more time: principles for comparing sets of documents. J Am Soc Inf Sci Technol 2011;62(7):1370-1381. [doi: 10.1002/Asi.21534]
52. Winkler W. String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage. In: Proceedings of the Survey Research Methods Section. Alexandria, VA: American Statistical Association; 1990.
53. Otero P, Pedernera F, Montenegro S, Borbolla D, Garcia Marti S, Luna D, et al. Evolution of medical informatics in bibliographic databases. Stud Health Technol Inform 2004;107(Pt 1):301-305. [Medline: 15360823]
54. Sittig DF. Identifying a core set of medical informatics serials: an analysis using the MEDLINE database. Bull Med Libr Assoc 1996 Apr;84(2):200-204. [Medline: 8826624]
55. Greenes RA, Siegel ER. Characterization of an emerging field: approaches to defining the literature and disciplinary boundaries of medical informatics. Proc Annu Symp Comput Appl Med Care 1987 Nov 4:411-415.
56. Raghupathi W, Nerur S. Research themes and trends in health information systems. Methods Inf Med 2008;47(5):435-442. [Medline: 18852917]
57. Sittig DF, Kaalaas-Sittig J. A quantitative ranking of the Biomedical Informatics serials. Methods Inf Med 1995 Sep;34(4):397-340. [Medline: 7476471]
58. Andrews J. An author co-citation analysis of medical informatics. J Med Libr Assoc 2003 Jan;91(1):47-56. [Medline: 12568157]
59. Glass RL. An assessment of systems and software engineering scholars and institutions. J Syst Softw 1994;27(1):63-67.
60. Glass RL. An assessment of systems and software engineering scholars and institutions (1993-1997). J Syst Softw 1998;43(1):59-64. [doi: 10.1016/S0164-1212(98)10037-7]
61. Wong WE, Tse TH, Glass RL, Basili VR, Chen TY. An assessment of systems and software engineering scholars and institutions (2003-2007 and 2004-2008). J Syst Softw 2011;84(1):162-168. [doi: 10.1016/j.jss.2010.09.036]
62. Ren J, Taylor RN. Automatic and versatile publications ranking for research institutions and scholars. Commun ACM 2007;50(6):81-85. [doi: 10.1145/1247001.1247010]
63. Geist R, Chetuparambil M, Hedetniemi S, Turner AJ. Computing research programs in the U.S. Commun ACM 1996;39(12):96-99. [doi: 10.1145/240483.240505]
64. Glass RL, Ramesh V, Vessey I. An analysis of research in computing disciplines. Communications of the ACM 2004;47(6):89-94. [doi: 10.1145/990680.990686]
65. Bornmann L. Daniel H-D. Selecting scientific excellence through committee peer review: a citation analysis of publications previously published to approval or rejection of post-doctoral research fellowship applicants. Scientometrics 2006;68:427-440. [doi: 10.1007/s11192-006-0121-1]
66. Endler NS. Where the “stars” are: the 25 most cited psychologists in Canada (1972-1976). Can Psychol Rev 1979;20(1):12-21. [doi: 10.1037/h0081489]
67. Jokic M. Scientometric evalution of the projects in biology funded by the Ministry of Science and Technology, Republic of Croatia, in the 1991-1996 period. Period Biol 2000;102:129-142.
68. Tutarel O. Geographical distribution of publications in the field of medical education. BMC Med Educ 2002 Apr 30;2:3 [FREE Full text] [Medline: 12031092]
69. Opthof T, Wilde AA. The Hirsch-index: a simple, new tool for the assessment of scientific output of individual scientists: The case of Dutch professors in clinical cardiology. Neth Heart J 2009 Apr;17(4):145-154. [Medline: 19421360]
70. Garfield E. Citation Indexing: Its Theory and Applications in Science, Technology, and Humanities. New York, NY: Wiley; 1979.
71. Schubert A, Braun T. Cross-field normalizations of scientometric indicators. Scientometrics 1996;36(3):311-324. [doi: 10.1007/BF02129597]
72. National Science Foundation. 2012 Jan. Science and Engineering Indicators 2012 URL: http://www.nsf.gov/statistics/seind12/appendix.htm [accessed 2012-10-08] [WebCite Cache ID 6BGaKExEu]

73. Scopus. 2009 Mar. Frequently Asked Questions: Author Identifier URL: http://www.info.sciverse.com/documents/files/scopus-training/resourcelibrary/pdf/FAQ_Author_Identifier_09.pdf [accessed 2012-10-08] [WebCite Cache ID 6BGaL0yc6]
74. Jacso P. Testing the Calculation of a Realistic h-index in Google Scholar, Scopus,Web of Science for F. W. Lancaster. Libr Trends 2008;56(4):784-815. [doi: 10.1353/lib.0.0011]
75. Meho LI, Yang K. Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. J Am Soc Inf Sci Technol 2007;58(13):2105-2125. [doi: 10.1002/asi.v58:13]
76. Jacso P. The pros and cons of computing the h-index using Google Scholar. Online Inf Rev 2008;32(3):437-452.
77. Sobel M, Huyett MJ. Nonparametric definition of the representativeness of a sample: with tables. Bell Syst Tech J 1958;37.
78. Woodruff RS. Confidence intervals for medians and other position measures. J Am Stat Assoc 1952;47(260):635-646.
79. Francisco CA, Fuller WA. Quantile estimation with a complex survey design. Ann Stat 1991;19(1):454-469.
80. Dorfman A, Valliant R. Quantile Variance Estimators in Complex Surveys. Alexandria, VA: American Statistical Association; 1993.
81. Sitter RR, Wu C. A note on Woodruff confidence intervals for quantiles. Stat Probab Lett 2001;52(4):353-358.
82. Chatterjee A. Asymptotic properties of sample quantiles from a finite population. Ann Inst Stat Math 2008;63:157-179.
83. Kosmulski M. Calibration against a reference set: a quantitative approach to assessment of the methods of assessment of scientific output. J Informetr 2012;6(3):451-456. [doi: 10.1016/j.joi.2012.03.004]
84. US National Library of Medicine. 2012 Apr 13. NLM's University-Based Biomedical Informatics Research Training Programs URL: http://www.nlm.nih.gov/ep/GrantTrainInstitute.html [accessed 2012-10-08] [WebCite Cache ID 6BGHoDbwm]
85. Wikipedia. 2012 Aug 31. Open Directory Project URL: http://en.wikipedia.org/wiki/Open_Directory_Project [accessed 2012-10-08] [WebCite Cache ID 6BGHpoStc]
86. Bornmann L, Marx W, Schier H, Rahm E, Thor A, Daniel HD. Convergent validity of bibliometric Google Scholar data in the field of chemistry: citation counts for papers that were accepted by Angewandte Chemie International Edition or rejected but published elsewhere, using Google Scholar, Science Citation Index, Scopus, and Chemical Abstracts. J Informetr 2009;3(1):27-35. [doi: 10.1016/j.joi.2008.11.001]

Abbreviations

ACMI: American College of Medical Informatics
JAMIA: Journal of the American Medical Informatics Association
JMIR: Journal of Medical Internet Research
MeSH: Medical Subject Headings
PR6: 6 percentile-level benchmark
PR10: decile benchmark

Edited by G Eysenbach; submitted 17.05.12; peer-reviewed by R Scherer, L Bornmann; comments to author 13.06.12; revised version received 18.08.12; accepted 23.09.12; published 18.10.12.

Please cite as:
El Emam K, Arbuckle L, Jonker E, Anderson K
Two h-Index Benchmarks for Evaluating the Publication Performance of Medical Informatics Researchers
J Med Internet Res 2012;14(5):e144
URL: http://www.jmir.org/2012/5/e144/
doi:10.2196/jmir.2177
PMID:23079075

Original Paper

Issues in mHealth: Findings From Key Informant Interviews

Robyn Whittaker[1], MB ChB, MPH, PhD

National Institute for Health Innovation, University of Auckland, Auckland, New Zealand

Corresponding Author:
Robyn Whittaker, MB ChB, MPH, PhD
National Institute for Health Innovation
University of Auckland
Private Bag 92019
Auckland, 1142
New Zealand
Phone: 64 99234766
Fax: 64 93731710
Email: r.whittaker@nihi.auckland.ac.nz

Abstract

Background: mHealth is enjoying considerable interest and private investment in the United States. A small but growing body of evidence indicates some promise in supporting healthy behavior change and self-management of long-term conditions. The unique benefits mobile phones bring to health initiatives, such as direct access to health information regardless of time or location, may create specific issues for the implementation of such initiatives. Other issues may be shared with general health information technology developments.

Objective: To determine the important issues facing the implementation of mHealth from the perspective of those within the US health system and those working in mHealth in the United States.

Methods: Semistructured interviews were conducted with 27 key informants from across the health and mHealth sectors in the United States. Interviewees were approached directly following an environmental scan of mHealth in the United States or recommendation by those working in mHealth.

Results: The most common issues were privacy and data security, funding, a lack of good examples of the efficacy and cost effectiveness of mHealth in practice, and the need for more high-quality research. The issues are outlined and categorized according to the environment within which they predominantly occur: policy and regulatory environments; the wireless industry; the health system; existing mHealth practice; and research.

Conclusions: Many of these issues could be addressed by making the most of the current US health reform environment, developing a strategic and coordinated approach, and seeking to improve mHealth practice.

(J Med Internet Res 2012;14(5):e129) doi:10.2196/jmir.1989

KEYWORDS

Cellular phone; mobile health; mobile phone

Introduction

mHealth has been most succinctly defined as health-related services delivered via mobile communications devices [1]. While this definition is broad in scope, some feel it may not be particularly helpful to define mHealth as a separate entity because it may lead to logistic or regulatory limits being imposed that could stifle innovative integration into the system. Holt has used the term unplatforms to recognize the fact that people already use a multitude of devices and channels that will continue to evolve and develop [2]. Definitions aside, there appears to be value in delivering health-related information and interventions, and improving access to health services via devices that are personal, intelligent, connected, and always with people [3]. Such delivery channels may enable real-time health advice, prompts, monitoring, feedback, personalized support, and interventions that were not easily provided prior to the almost ubiquitous uptake of mobile phones.

A small but growing body of evidence supports the use of mobile phones in health interventions. These include the use of mobile phones for smoking cessation [4-6], other behavior change programs such as weight management [7-9], and self-management of long-term conditions such as diabetes [10-14]. Text messaging is being used to provide reminders for health care appointments [15,16] and to improve the efficiency of health systems [17]. Text messages are also being used to improve medication adherence [18-22] and to provide health information [23,24]. However, there is relatively little research published on the effectiveness of downloadable applications (apps) or software for mobile phones with computer operating systems (smartphones) [25,26].

In the United States approximately 83% of the population uses mobile phones and 73% of them use text messaging [27]. African Americans and English-speaking Hispanic Americans are more likely than white Americans to use text messaging (76% and 83% vs 70%, respectively), to access the Internet via mobile (56% and 51% vs 39%), to download an app (36% and 36% vs 28%), and to engage in social networking via mobile (39% and 35% vs 25%) [27]. In the past few years, smartphones have become increasingly popular, to the extent that over 35% of US mobile phone users now have a smartphone [27]. In another survey, 17% of cell phone users had used their phones to look for health information online, and 9% had a smartphone app to help them track or monitor their health. However, of those who had any apps on their phone (35% of US adults), only one-quarter said they actually used them [28].

The unique feature of all mobile phones—portable persistent connectivity—brings opportunities for health services but also creates some unique issues facing the implementation of mHealth initiatives. While mHealth is in an early stage of expansion with much hype, private investment (estimated to be more than US $500 million in 2011 [29]), and federal government interest [30], it may be timely to consider a strategic approach to these issues.

This paper summarizes some of the key issues facing the implementation of mHealth in the Unites States based on findings from interviews with people across the health and mHealth sectors. Opportunities to address these issues are then outlined from the author's perspective.

Methods

Initial discussions were held by the author, as a 2010/11 Commonwealth Fund Harkness Fellow in Healthcare Policy and Practice, with people in the federal health services and the private mHealth sector. These discussions informed the development of important themes to be covered in the interviews. The themes were the priority and value of mHealth, potential benefits to underserved populations, issues and barriers to mHealth implementation, and issues around mHealth research. These initial discussions, along with an environmental scan, also provided a master list of possible interviewees. This was added to by asking each interviewee for recommendations of other potential key informants. The final master list consisted of approximately 76 individuals. It was decided to initially focus on representatives from several areas: relevant Department of Health and Human Services agencies; integrated health systems; people involved in delivering mHealth services; and academics working in mHealth. Potential interviewees were approached by the researcher directly by email or via email introductions from others. A small number did not respond at all (n = 7), but none of those who responded refused an interview. Some potential interviewees who agreed were never able to complete an interview (n = 6), some encounters became informal discussions instead of interviews (n = 5), and more potential interviewees were never contacted due to time constraints and reaching data saturation (n = 31).

Semistructured interviews were conducted by phone or in person between September 2010 and July 2011. Some themes and questions were prespecified (as above), but other topics were allowed to arise during the interviews. The interviews were recorded and transcribed, and a general thematic analysis was undertaken by the author. This analysis involved identifying all discussions on the prespecified themes and other themes raised by the interviewees. The analysis was followed by coding and categorizing discussions into subthemes. Responses to some prespecified questions were analyzed quantitatively where possible. A report on the findings of the interviews detailed by theme and subthemes was provided to interviewees for comment, and some amendments were made based on their feedback. The author developed the categorization of issues as presented in this paper, and the discussion on opportunities to address these issues.

Results

A total of 27 key informants participated in semistructured interviews. The interviewees were people interested in mHealth from various federal agencies (n = 10), those from integrated health systems (n = 3), academics working in this area (n = 6), and people working in mHealth companies (n = 4), wireless networks (n = 2), and organizations that sponsor mHealth initiatives or research (n = 2).

The Key Issues in mHealth Implementation

Key issues arising from the informant interviews have been divided into five areas covering the wider health and wireless environment within which mHealth is implemented: the policy and regulatory environment, the wireless network environment, the health system environment, mHealth in current practice, and mHealth research (see Table 1). Not all of these issues are specific to mHealth, and some are most likely specific to the United States.

Table 1. Issues facing the implementation of mHealth raised by key informants.

Area	Issues
Policy and regulatory	Privacy and data security FDA[a] regulation of mHealth initiatives as medical devices Medical practice across states or countries and with respect to clinical practice roles (eg, prescribing regulations) Bandwidth or spectrum availability
Wireless networks	Compatibility across multiple networks Compatibility across multiple platforms and proprietary systems Cost to the public or end user Coverage in remote areas
Health system	Lack of examples of sustainable business models Lack of reimbursement Lack of understanding of value mHealth may provide Clinical roles accountability and integration into clinical practice Integration into electronic health records and health information systems Competing health information technology priorities and broader opportunity cost
mHealth practice	Lack of knowledge of how to do it well Wrong focus on the technology or on advantaged populations (those who don't need it) Governance in mHealth Publicly available applications not evaluated and without basis in theory or evidence Stand-alone or siloed initiatives due to existing platforms or proprietary systems
Research	Need for more high-quality research Need to demonstrate efficacy and cost effectiveness Mismatch in pace and flexibility between research and technology development Measurement of reach or access for the underserved

[a] US Food and Drug Administration.

Policy and Regulatory Environment

The issues of privacy and health data security were discussed with all informants. Opinions varied, with some informants considering privacy of personal health information via mobile to be a major issue requiring high-level guidance and widespread discussion before the field can move ahead. A smaller number of informants focused on the potential for straightforward, technologically based security solutions to solve many of the privacy and security issues. The role of the Health Insurance Portability and Accountability Act's (HIPAA) Privacy, Security and Breach Notification rules was mentioned by most informants (the Act that aims to ensure health plans, health care clearinghouses, and health care providers and their business associates take "reasonable measures" to prevent any uses or disclosures of protected health information that are not permitted or consented to by the individuals). There are some unique aspects of mHealth relating to the Act that have been discussed elsewhere [31]. A further point raised by a small number of informants was the unique position of the wireless networks' involvement in handling data. Networks have indicated that they are working on technological security solutions.

Another issue raised by the majority of informants was the regulation of mHealth tools and programs as medical devices. Some accepted that regulation is appropriate (and some mHealth developers have proactively sought and received regulatory approval), although some informants expressed concerns, particularly around definitions of what should and what should not require regulation. The US Food and Drug Administration's draft guidance on this topic was made public after most of these interviews had been completed. This guidance indicates that general health and wellness applications are unlikely to be regulated; nor are applications "not intended for curing, treating, seeking treatment for mitigating, or diagnosing a specific disease, disorder, patient state, or any specific, identifiable health condition" [32]. As the field moves toward more comprehensive and integrated solutions that include mHealth, exactly where the line will be drawn may be less than obvious [33]. Some also expressed a concern that the ongoing iterations and improvements common in this type of agile technology development should not necessitate frequent updates to regulatory approval.

Wireless Network Environment

Issues raised in this category mainly came from interviewees involved in implementing mHealth initiatives in a variety of contexts. They commented on the large number of wireless networks in the United States, which can make establishing relationships and interfaces for comprehensive implementation difficult. One example of how this can successfully work was raised by several informants. CTIA-The Wireless Association, a nonprofit member organization that represents the industry, was involved in establishing industry-wide support for the national text4baby service, thereby ensuring that health information text messages for pregnant women and new mothers are available and free of charge regardless of network.

Informants involved in implementing mHealth initiatives discussed the fact that individuals are charged for receiving text messages as an issue. The cost of mHealth services to the public is of particular concern to those working with socioeconomically disadvantaged populations. The cost of mHealth was linked to the wider issue of the use of mHealth initiatives to reach underserved populations, which was uniformly recognized as a potential benefit due to the high levels of mobile phone ownership and use of text messaging in ethnic minority and low socioeconomic populations in the United States [27,34]. However there is, as yet, little published evidence that mHealth has actually improved access to health information or services by those who were not otherwise receiving the service. Some informants also mentioned that a focus on effectively using the ubiquitous aspects of mobile technologies (such as text messaging) for wider reach into populations that need it most contradicts a more popular emphasis on high-tech developments. A small number of informants felt there may be other issues around the use of mobile technologies by disadvantaged populations (for example, around language, culture, or network coverage) that can also affect the ability of mHealth initiatives to reach those in need [35].

Some informants mentioned a barrier to the development of more widely available and more comprehensive solutions arising from proprietary systems and multiple platforms. This is a phenomenon previously seen in other areas of information technology development [36].

Health System Environment

The predominant fee-for-service reimbursement structure of the US health care system is seen by almost all informants as less than ideal for the implementation of mHealth. Indeed, many mHealth initiatives aim to reduce in-person reimbursable visits for hospitals, clinics, and providers [6,19,25]. This is obviously a wider issue involving other forms of digital health, telehealth services, and preventive services.

The issue of funding was often linked in informant discussions to a stated lack of understanding of the potential of mHealth to contribute to the health system by funders, managers, and decision makers. This understanding was said to be slow to evolve, and some felt this to be due to a lack of good cost-effectiveness evidence to inform decisions.

Some informants did raise the current lack of demonstrably sustainable mHealth business models in the system. Suggested possible funders of mHealth initiatives were employers, health insurers, individuals, and the federal government. Specific examples discussed were recent private investments in mHealth start-up companies and the national-level public–private partnership behind the text4baby program. However, several interviewees mentioned that more urgent priorities within the health system are monopolizing focus and resources, particularly around the implementation of electronic health records (EHR) systems. Integration issues with EHR systems were also seen as a barrier for those working in the field.

A small number of informants discussed the need for a strategic framework and governance of the implementation of mHealth. This was particularly in the context of issues such as interoperability and standardization, but also with respect to changes in the way health care is delivered. For example, if mHealth becomes a catalyst in moving the locus of control or responsibility for health (and health information) toward patients, then existing governance systems may need to change accordingly. In a similar vein, clinicians' concerns about mHealth were mentioned by a small number of informants. This was not around adoption of mobile technology per se (doctors are disproportionately high adopters [37]), but more around their roles and accountabilities in this new model of delivery. The examples discussed included responding to direct messages from patients (in various formats) and to real-time, continuous, remote monitoring data, where there are no existing protocols for interpreting this intensity of data. This was expressed as a desire of clinicians to see mHealth implementation being well thought through and any potential adverse consequences addressed in advance.

Current mHealth Practice

Several informants raised the concern that many mHealth applications available in practice may not be effective, engaging, usable, or meeting the needs of users. Few applications have been evaluated, and those that have often involve complex interventions where the components or mechanisms have not been examined. Many felt that not a lot is known as yet about what aspects of mHealth work, for whom, and why. Few published health interventions delivered via mobile technologies discuss a theoretical basis or evaluate theoretical components hypothesized to be important in the intervention [38]. It was stated that there is much hype and lots of players all “doing their own thing.” Some informants felt that some mHealth developers may have a bias toward developing programs for people like themselves using the technologies they like, rather than starting with the problem and working with end users to develop the most useful and usable solution. Some pointed to statistics in the media showing that many smartphone applications are downloaded but not used. More recently, reviews have found poor quality in terms of accuracy, usability, consistency with national practice guidelines, and effective practices [26,27,39,40].

mHealth Research

When key informants were asked about mHealth research, many rated the state of the evidence as early or weak and identified various areas as being in need of more high-quality research. Most felt that there is sufficient proof that mHealth is worth pursuing, although more solid evidence is required in terms of cost effectiveness and in determining what works. Some felt quite strongly that there should be no need to wait for randomized controlled trials to provide evidence on the effectiveness of every aspect of mHealth before any progress can be made. Two examples given were (1) where there is no access or very limited access to a health service and we can extend access to that service via mHealth with very little possibility of harm, and (2) where mobile delivery can be seen as the natural extension of what we already do, such as in health communication using the current preferred methods of communicating with the population.

Many talked of the obvious mismatch in pace and flexibility between traditional health research methods and rapid technological development. This mismatch was compared with the way research and development are often undertaken concurrently and iteratively in the academic engineering and computer science fields and in commercial development. Since the time of the interviews, the issue of alternative methodologies, appropriate comparators, and standard measures in mHealth research has been under review by research and funding agencies in the United States [41,42].

Opportunities to Address the Issues

Some interviewees did discuss their thoughts on how these issues should be addressed, although these discussions were not included in every interview. During the analysis of the discussions, the author developed a list of potential opportunities to address the issues in each category (Table 2). These opportunities are summarized in three main areas below.

Table 2. Opportunities to address issues in mHealth implementation.

Area	Opportunities
Policy and regulatory	Federal-level guidance Coordination of wireless industry and health sector
Wireless networks	Shared standards for interoperability Consideration of open architecture or standardized interfaces Industry coordination and collaboration for mHealth
Health system	Use of opportunities of current health reform investment (eg, CMMI[a] demonstration projects; competition in electronic health records industry; beacon community projects and evaluation)
mHealth practice	Collaboration with end users to develop solutions to their problems Foundation of theory and evidence of what works Collaborations and shared learning for iteration and improvement, integration of public data, and integration into health systems Consideration of open source & other methods to reduce barriers to more comprehensive integrated initiatives
Research	Consideration of alternative research methods to increase pace and retain rigor, including careful consideration of comparators Inclusion of measures of increased access Publication of formative research & evaluations of existing interventions

[a] Center for Medicare & Medicaid Innovation.

Health Reform Opportunities

The US health reform environment includes a focus on investment in health information technology and consideration of alternative models of payment and health care provision. This environment creates opportunities to demonstrate the value of mHealth and to ensure that future capabilities are integrated into health information technology systems. For example, the Center for Medicare & Medicaid Innovation “has the resources and flexibility to rapidly test innovative care and payment models and encourage widespread adoption of practices that deliver better health care at lower cost” [43]. Their process is to solicit ideas for new payment and service delivery models, select the most promising models, test and evaluate them, and then spread the successful models. mHealth seems an appropriate means to support their stated priority areas, including bundled payments for patient care, seamless coordinated care, and community and population health models to keep people healthy.

A total of 17 beacon communities, identified as those that have already adopted EHR systems and health information exchange, have been awarded extra funding via the Health Information Technology for Economic and Clinical Health (HITECH) Act to demonstrate the transformative ability of health information technology with respect to quality, cost efficiency, and population health [44]. These communities already include engaged and connected health services, clinicians, and community health workers who have agreed to implement combinations of innovative interventions, to measure their performance, and to share their learnings [45]. These enthusiasts could be encouraged to develop and evaluate mHealth initiatives that make sense for their local health improvement goals, thus helping to demonstrate the efficacy of such initiatives in practice.

Hundreds of EHR system vendors have had an even greater number of products certified by the Office of the National Coordinator for Health Information Technology [46]. This competitive environment could be used to push for the integration of mobile functionality as a point of difference between systems, ensuring that future capability is built into systems being adopted now, instead of as later add-ons.

Federal-Level Guidance

A degree of national-level guidance may be warranted, particularly in terms of coordinating a strategic approach to some of the wider issues. This guidance could include coordinating efforts across the wireless and mHealth-related industries and the health sector around topics such as developing common standards and interfaces to allow interoperability; considering privacy and data security solutions; and discussing the movement toward an open architecture and the ability to innovate and integrate across platforms and across networks. Estrin and Sim have stated that mHealth should learn from the Internet's development and from the previous siloed approaches to health information systems, and should work together toward an ecosystem for innovation [36]. mHealth need not be confined by geographic borders and, if the ecosystem is global, there may need to be international agreement on approaches to these issues. Elsewhere, the adoption of international standards has been facilitated by organizations such as the International Telecommunication Union.

Improving mHealth Practice and Research

The need to improve the development of mHealth initiatives came through quite strongly in these interviews. We have the opportunity early in the growth of this field to consider a structured approach to development that is guided by relevant theory and current evidence, and that may lead us to a logical framework or model of how, why, and what works in mHealth. Riley et al suggest that current health behavior theories may not be adequate for interactive and adaptive mobile technology-enabled interventions [38]. Perhaps more research is required on new theories that may advance our understanding of how mHealth initiatives can be effective.

The development process, as in many other fields, needs to start with the problem and work with the end users (clinicians, health service providers, patients, and the general public) to develop the most appropriate and useful solution to that problem. Real-world implementation of initiatives should be considered from the outset, so that practical issues such as intellectual property, scaling up, and integration into practice are addressed. With a more open philosophy, it may be possible to iterate, adapt, and improve on what others have done. This is not always easy, but the end goal should be integration with other health information technology systems, particularly with access to personal health information and publicly available information, in order to develop seamless services that are centered on individuals rather than providers or locations.

Evaluations of effectiveness and usability are required and should be made publicly available. Where evaluation is planned during the development stage, data collection can be built in as an integral part of the program. The ideal of randomized controlled trials will still be necessary in some contexts. In these cases, careful consideration should be given to the appropriate comparator to ensure the right question is being answered. For example, what is usual care for this target audience? Can we measure an improvement in access as an outcome? Other research methods will be more appropriate in other circumstances, such as adaptive trials to allow the intervention to develop and improve as part of the research; observational trials and qualitative research methods to detect unintended consequences and changes to workflow; and qualitative studies to test acceptability. Evaluating effectiveness and usability is also possible while implementing a system, for example, with novel designs such as the stepped wedge cluster randomized trial, and particularly where there is little likelihood of harm.

The other great opportunity for mHealth in practice and research is data collection in environmental and population health surveillance, as early warning systems of public health issues and as emergency information systems in natural disasters or pandemics. Many such systems are being developed and will surely advance the body of knowledge around using mHealth for public health.

Discussion

This paper outlines key issues in the implementation of mHealth in the United States as raised by an environmental scan and key informants from the health and mHealth sectors. The issues have been categorized according to policy environment, wireless environment, health system, mHealth practice, and research, although many issues could be seen to cut across several categories. This categorization allowed a matching of issues with current opportunities to address them, which were then aggregated into three main action areas: using the opportunities provided by the current health care reform investment and processes; establishing some degree of federal-level guidance and coordination across the industry; and improving mHealth practice with good intervention development principles and sound research methodology.

Despite the large number of issues raised, most informants were optimistic about our ability to address these issues in the near future. These optimists included those who see the transformative potential of mHealth to change the way we deliver health-related services, as well as those who see the use of mobile communications devices as just the natural evolution of what we currently do in health communication and health care.

There are limitations to this study. Although no further new themes were being added, the list of possible issues presented here still may not be exhaustive due to the selection of interviewees. The qualitative analysis of the transcribed interviews was undertaken by the author only. Reflexivity may be an issue, as the author is a researcher in the field of mHealth and therefore not a completely independent observer. However, prior to this period the author had not worked in the United States or with any of the interviewees. It is possible that the initial environmental scan and the author's own expertise in this area may even have brought some degree of validity filtering to the discussions and the analysis. The categorization of issues and the identified opportunities for addressing them are the opinions of the author only.

Conclusions

To many people, the use of mobile phones in health is obvious—if their hairdresser is sending them text message reminders, why isn't their doctor? To date, we have failed to develop mHealth initiatives that are so easy to use and so obviously useful that large numbers of people want to regularly use them over time. As one informant noted, once that is achieved many of the other issues will be resolved.

The overall opportunity of mHealth will come from accessing large, widespread populations directly with individually tailored programs. Even where these programs have only a small effect, with such wide reach there is the opportunity for population-level changes—moving the bell-shaped curve of health a small amount with huge public health impact [47].

Acknowledgments

RW was sponsored by The Commonwealth Fund as a 2010/11 Harkness Fellow in Healthcare Policy and Practice. The author would like to thank all the interviewees for their time and support for this project.

Conflicts of Interest

None declared.

References

1. Vital Wave Consulting. mHealth for Development: The Opportunity of Mobile Technology for Healthcare in the Developing World. Washington, DC: UN Foundation-Vodafone Foundation Partnership; 2009. URL: http://unpan1.un.org/intradoc/groups/public/documents/unpan/unpan037268.pdf [accessed 2012-09-09] [WebCite Cache ID 6AYdjAazg]
2. Holt M. The Health Care Blog. Diversinet: A Backbone for Unplatforms? 2010 Aug 17 URL: http://thehealthcareblog.com/blog/tag/unplatforms/ [accessed 2011-10-31] [WebCite Cache ID 62rJQhMQt]
3. Fogg BJ, Adler R. In: Fogg BJ, Adler R, editors. Texting 4 Health: A Simple, Powerful Way to Improve Lives. Stanford, CA: Captology Media Stanford University; 2009.
4. Rodgers A, Corbett T, Bramley D, Riddell T, Wills M, Lin RB, et al. Do u smoke after txt? Results of a randomised trial of smoking cessation using mobile phone text messaging. Tob Control 2005 Aug;14(4):255-261 [FREE Full text] [doi: 10.1136/tc.2005.011577] [Medline: 16046689]
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(4):CD006611. [doi: 10.1002/14651858.CD006611.pub2] [Medline: 19821377]
6. Free C, Knight R, Robertson S, Whittaker R, Edwards P, Zhou W, et al. Smoking cessation support delivered via mobile phone text messaging (txt2stop): a single-blind, randomised trial. Lancet 2011 Jul 2;378(9785):49-55. [doi: 10.1016/S0140-6736(11)60701-0] [Medline: 21722952]
7. Fjeldsoe BS, Miller YD, Marshall AL. MobileMums: a randomized controlled trial of an SMS-based physical activity intervention. Ann Behav Med 2010 May;39(2):101-111 [FREE Full text] [doi: 10.1007/s12160-010-9170-z] [Medline: 20174902]
8. Hurling R, Catt M, Boni MD, Fairley BW, Hurst T, Murray P, et al. Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. J Med Internet Res 2007;9(2):e7 [FREE Full text] [doi: 10.2196/jmir.9.2.e7] [Medline: 17478409]
9. Patrick K, Raab F, Adams MA, Dillon L, Zabinski M, Rock CL, et al. A text message-based intervention for weight loss: randomized controlled trial. J Med Internet Res 2009;11(1):e1 [FREE Full text] [doi: 10.2196/jmir.1100] [Medline: 19141433]
10. Franklin VL, Waller A, Pagliari C, Greene SA. A randomized controlled trial of Sweet Talk, a text-messaging system to support young people with diabetes. Diabet Med 2006 Dec;23(12):1332-1338. [doi: 10.1111/j.1464-5491.2006.01989.x] [Medline: 17116184]
11. Yoon KH, Kim HS. A short message service by cellular phone in type 2 diabetic patients for 12 months. Diabetes Res Clin Pract 2008 Feb;79(2):256-261. [doi: 10.1016/j.diabres.2007.09.007] [Medline: 17988756]
12. Kim HS. A randomized controlled trial of a nurse short-message service by cellular phone for people with diabetes. Int J Nurs Stud 2007 Jul;44(5):687-692. [doi: 10.1016/j.ijnurstu.2006.01.011] [Medline: 16618486]
13. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. Epidemiol Rev 2010 Apr;32(1):56-69. [doi: 10.1093/epirev/mxq004] [Medline: 20354039]
14. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health 2009 Apr;15(3):231-240. [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]
15. Leong KC, Chen WS, Leong KW, Mastura I, Mimi O, Sheikh MA, et al. The use of text messaging to improve attendance in primary care: a randomized controlled trial. Fam Pract 2006 Dec;23(6):699-705 [FREE Full text] [doi: 10.1093/fampra/cml044] [Medline: 16916871]
16. Downer SR, Meara JG, Da Costa AC, Sethuraman K. SMS text messaging improves outpatient attendance. Aust Health Rev 2006 Aug;30(3):389-396. [Medline: 16879098]
17. Mahmud N, Rodriguez J, Nesbit J. A text message-based intervention to bridge the healthcare communication gap in the rural developing world. Technol Health Care 2010;18(2):137-144. [doi: 10.3233/THC-2010-0576] [Medline: 20495253]
18. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010 Nov 27;376(9755):1838-1845. [doi: 10.1016/S0140-6736(10)61997-6] [Medline: 21071074]

19. Strandbygaard U, Thomsen SF, Backer V. A daily SMS reminder increases adherence to asthma treatment: a three-month follow-up study. Respir Med 2010 Feb;104(2):166-171. [doi: 10.1016/j.rmed.2009.10.003] [Medline: 19854632]
20. Cocosila M, Archer N, Haynes RB, Yuan Y. Can wireless text messaging improve adherence to preventive activities? Results of a randomised controlled trial. Int J Med Inform 2009 Apr;78(4):230-238. [doi: 10.1016/j.ijmedinf.2008.07.011] [Medline: 18778967]
21. Miloh T, Annunziato R, Arnon R, Warshaw J, Parkar S, Suchy FJ, et al. Improved adherence and outcomes for pediatric liver transplant recipients by using text messaging. Pediatrics 2009 Nov;124(5):e844-e850 [FREE Full text] [doi: 10.1542/peds.2009-0415] [Medline: 19822583]
22. Pena-Robichaux V, Kvedar JC, Watson AJ. Text messages as a reminder aid and educational tool in adults and adolescents with atopic dermatitis: a pilot study. Dermatol Res Pract 2010;2010 [FREE Full text] [doi: 10.1155/2010/894258] [Medline: 20885940]
23. Danis CM, Ellis JB, Kellogg WA, van Beijma H, Hoefman B, Daniels SD, et al. Mobile phones for health education in the developing world: SMS as a user interface. 2010 Presented at: ACM DEV '10: First Annual Symposium on Computing for Development; Dec 17-18, 2010; London, UK URL: http://www.mobileactive.org/files/file_uploads/IBM-TTC.pdf
24. Lim MS, Hocking JS, Aitken CK, Fairley CK, Jordan L, Lewis JA, et al. Impact of text and email messaging on the sexual health of young people: a randomised controlled trial. J Epidemiol Community Health 2012 Jan;66(1):69-74. [doi: 10.1136/jech.2009.100396] [Medline: 21415232]
25. Quinn CC, Clough SS, Minor JM, Lender D, Okafor MC, Gruber-Baldini A. WellDoc mobile diabetes management randomized controlled trial: change in clinical and behavioral outcomes and patient and physician satisfaction. Diabetes Technol Ther 2008 Jun;10(3):160-168. [doi: 10.1089/dia.2008.0283] [Medline: 18473689]
26. Abroms LC, Padmanabhan N, Thaweethai L, Phillips T. iPhone apps for smoking cessation: a content analysis. Am J Prev Med 2011 Mar;40(3):279-285. [doi: 10.1016/j.amepre.2010.10.032] [Medline: 21335258]
27. Smith A. Americans and their cell phones. Washington, DC: Pew Internet & American Life Project; 2011 Aug 15. URL: http://pewinternet.org/~/media//Files/Reports/2011/Cell%20Phones%202011.pdf [accessed 2012-09-05] [WebCite Cache ID 6AST1zoyI]
28. Fox S. Mobile health 2010. Washington, DC: Pew Internet & American Life Project; 2010 Oct 19. URL: http://www.pewinternet.org/~/media//Files/Reports/2010/PIP_Mobile_Health_2010.pdf [accessed 2012-09-05] [WebCite Cache ID 6AST312Y8]
29. Dolan B. Mobihealthnews. 2012 Jan 24. 2011: More than $500M in mobile health investiments URL: http://mobihealthnews.com/16065/2011-more-than-500m-in-mobile-health-investments/ [accessed 2012-09-10] [WebCite Cache ID 6AYefFiVv]
30. US Department of Health and Human Services Text4Health Taskforce. US Department of Health and Human Services. 2011. Health Text Messaging Recommendations to the Secretary URL: http://www.hhs.gov/open/initiatives/mhealth/recommendations.html [accessed 2011-11-03] [WebCite Cache ID 62vrAFuJr]
31. Greene AH. Chester Street Publishing. 2011 Jun 16. Mobihealthnews: When HIPAA Applies to Mobile Applications URL: http://mobihealthnews.com/11261/when-hipaa-applies-to-mobile-applications/ [accessed 2011-11-03] [WebCite Cache ID 62vrLMLYg]
32. US Department of Health and Human Service, Food and Drug Administration. 2011 Jul 20. Draft Guidance for Industry and Food and Drug Administration Staff: Mobile Medical Applications. July URL: http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm263280.htm [accessed 2011-10-31] [WebCite Cache ID 62rJ6x3KQ]
33. Thompson BM. Chester Street Publishing, Inc. 2010 Jun. FDA Regulation of Mobile Health: Mobihealthnews 2010 Report URL: http://mobihealthnews.com/wp-content/pdf/FDA_Regulation_of_Mobile_Health.pdf [accessed 2011-11-03] [WebCite Cache ID 62vrynVkN]
34. Health Research Institute. PriceWaterhouseCoopers. 2010. Healthcare Unwired: New Business Models Delivering Care Anywhere URL: http://www.pwc.com/us/en/health-industries/publications/healthcare-unwired.jhtml [accessed 2012-09-09] [WebCite Cache ID 6AYeOOaRK]
35. Tirado M. Role of mobile health in the care of culturally and linguistically diverse US populations. Perspect Health Inf Manag 2011;8:1e. [Medline: 21307988]
36. Estrin D, Sim I. Health care delivery. Open mHealth architecture: an engine for health care innovation. Science 2010 Nov 5;330(6005):759-760. [doi: 10.1126/science.1196187] [Medline: 21051617]
37. Jackson & Coker Research Associates. 2011. Special Report: Apps, Doctors, and Digital Devices. Industry Report 2011;4(7) URL: http://www.jacksoncoker.com/physician-career-resources/newsletters/monthlymain/des/Apps.aspx [accessed 2011-10-31] [WebCite Cache ID 62rIq2YTl]
38. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar 1;1(1):53-71. [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
39. Gan KO, Allman-Farinelli M. A scientific audit of smartphone applications for the management of obesity. Aust N Z J Public Health 2011 Jun;35(3):293-294. [doi: 10.1111/j.1753-6405.2011.00707.x] [Medline: 21627732]
40. Consumer Health Information Corporation. 2010. Motivating Patients to Use Smartphone Health Apps URL: http://www.consumer-health.com/press/2008/NewsReleaseSmartPhoneApps.php [accessed 2011-11-06] [WebCite Cache ID 630BZrGMk]

41. Shar A. Robert Wood Johnson Foundation. What Does mEvidence Need to Look Like? 2011 Aug 19 URL: http://rwjfblogs.typepad.com/pioneer/2011/08/moving-mhealth-forward-through-evidence.html [accessed 2011-11-06] [WebCite Cache ID 630BwRqZR]
42. Anonymous. US Department of Health and Human Services, Office of Behavioral and Social Sciences Research, National Institutes of Health. 2011 Aug 16. mHealth: Workshop on "mHealth Evidence" URL: http://obssr.od.nih.gov/scientific_areas/methodology/mhealth/mhealth-workshop.aspx [accessed 2011-11-06] [WebCite Cache ID 630C1k1ja]
43. Centers for Medicare & Medicaid Services, Center for Medicare & Medicaid Innovation. 2012 Aug. Welcome to the CMS Innovation Center URL: http://innovations.cms.gov/ [accessed 2012-09-09] [WebCite Cache ID 6AYeGjINi]
44. US Department of Health and Human Services, The Office of the National Coordinator for Health Information Technology. 2011 Jan 27. HITECH Programs URL: http://healthit.hhs.gov/portal/server.pt/community/healthit_hhs_gov__hitech_programs/1487 [accessed 2011-11-06] [WebCite Cache ID 630EW8Cxv]
45. McKethan A, Brammer C, Fatemi P, Kim M, Kirtane J, Kunzman J, et al. An early status report on the Beacon Communities' plans for transformation via health information technology. Health Aff (Millwood) 2011 Apr;30(4):782-788. [doi: 10.1377/hlthaff.2011.0166] [Medline: 21471501]
46. US Department of Health and Human Services, The Office of the National Coordinator for Health Information Technology. 2010 Dec 23. Certified Health IT Product List URL: http://onc-chpl.force.com/ehrcert [accessed 2011-11-06] [WebCite Cache ID 630EumK9n]
47. Glasgow RE, McKay HG, Piette JD, Reynolds KD. The RE-AIM framework for evaluating interventions: what can it tell us about approaches to chronic illness management? Patient Educ Couns 2001 Aug;44(2):119-127. [Medline: 11479052]

Abbreviations

EHR: electronic health records
HIPAA: Health Insurance Portability and Accountability Act
HITECH: Health Information Technology for Economic and Clinical Health

Edited by T Houston; submitted 09.11.11; peer-reviewed by R Lefebvre, B Holtz; comments to author 26.12.11; revised version received 31.01.12; accepted 05.07.12; published 02.10.12.

Please cite as:
Whittaker R
Issues in mHealth: Findings From Key Informant Interviews
J Med Internet Res 2012;14(5):e129
URL: http://www.jmir.org/2012/5/e129/
doi:10.2196/jmir.1989
PMID:23032424

Original Paper

Development and Testing of a Multimedia Internet-Based System for Fidelity and Monitoring of Multidimensional Treatment Foster Care

Edward G Feil[1], PhD; Peter G Sprengelmeyer[2], PhD; Betsy Davis[1], PhD; Patricia Chamberlain[3], PhD

[1]Oregon Research Institute, Eugene, OR, United States
[2]OSLC Community Programs, Eugene, OR, United States
[3]Oregon Social Learning Center, Eugene, OR, United States

Corresponding Author:
Edward G Feil, PhD
Oregon Research Institute
1715 Franklin Blvd
Eugene, OR, 97403
United States
Phone: 1 541 484 2123 ext 2259
Fax: 1 484 1108
Email: edf@ori.org

Abstract

Background: The fields of mental health, child welfare, and juvenile justice are jointly faced with the challenge of reducing the prevalence of antisocial behavior among adolescents. In the last 20 years, conduct disorders have moved from being considered intractable difficulties to having complex but available solutions. The treatments for even long-standing offending behavior among adolescents are now well documented and supported by a growing and compelling body of evidence. These empirically validated interventions are being widely disseminated, but the replication of the results from clinical trials in community settings has yet to be documented. The treatments, which produced impressive effects in a research context, are difficult to replicate without intensive monitoring of fidelity by the developers. Such monitoring is a barrier toward adoption; as the distance between the adopter and developer increases, so does cost. At the same time, states, communities, and agencies are under increasing pressure to implement those intervention services that have been shown to be most effective. The use of the Internet offers a potential solution in that existing reporting and data collection by clinicians can be subject to remote supervision. Such a system would have the potential to provide dissemination teams with more direct access to higher-quality data and would make adopters more likely to be able to implement services at the highest possible conformity to research protocols.

Objective: To create and test such an innovative system for use with the Multidimensional Treatment Foster Care (MTFC) program, which is an in-home treatment (alternative to a residential- or group-home setting) for antisocial youths. This research could advance the knowledge base about developing innovative infrastructures in community settings to disseminate empirically validated treatments.

Methods: The fidelity system was used and reviewed by parent and professional users: 20 foster parent users of the Parent Daily Report function, 9 professional MTFC program supervisors, and 4 MTFC consultants. All participants rated the system’s ease of use, quality of the website, and observational videos recorded at agency meetings. In addition, foster parents entered data on child behavior.

Results: All professionals and foster parents rated the system as very easy to use. We found particularly high levels of use by parents. Professionals rated the computer-collected videos of clinical meetings as being of high quality and easily codeable.

Conclusions: The project developed a user-friendly and secure Web-based system using state-of-the-art computer-based protocols for recording questionnaire and observational data generated by community-based MTFC staff and foster parents, with positive satisfaction and utilization results.

(J Med Internet Res 2012;14(5):e139) doi:10.2196/jmir.2034

KEYWORDS

Mental health; fidelity; clinical supervision; multidimensional treatment foster care; multimedia

Introduction

In the recent past, adolescent delinquent behavior and conduct disorder generally were considered to be intractable problems for which the only interventions were removal from the community and individual-based sanctions (eg, detention and community service work). Since then, several treatment and prevention approaches have been developed (eg, Multisystemic Therapy, Functional Family Therapy, and Multidimensional Treatment Foster Care [MTFC] [1]), each of which have proven to be extremely effective in reducing the number and seriousness of offenses for individual youths, in linking the problematic youths into their families and communities, and in preventing placements for youths in more restrictive settings (eg, boot camps and secure locked facilities). Such restrictive placements are both more costly and correlated with a much higher rate of entry into adult corrections than are community placements [2].

Empirically Validated Treatments

Along with the efforts to examine capacity expansion and dissemination, there has been a consistent increase in the call for communities to adopt empirically validated practices from federal programs, state policies, and local advocacy. Some of the necessary infrastructure for dissemination has been or is being developed. The programs identified as validated have been manualized, a cadre of trainers has been established, and the listed programs shown to be effective have been widely publicized. However, an ongoing problem has been that without extensive direct support from the program developers, the community implementations of these practices remain substantially less effective than they were during the efficacy trials [3]. There is likely a variety of reasons for this decrement in the effectiveness as these interventions are disseminated into communities: for instance, lack of fidelity to core intervention practices, attempting to expand the services beyond the population shown to be positively affected, and limits of the original efficacy studies (eg, restricted clinical population, guaranteed funding for services, and highly trained interventionists) have all been suggested as increasing the treatment effects in efficacy trails and limiting the effectiveness of the interventions when applied in the community (real world).

Multidimensional Treatment Foster Care

MTFC [1] is the only empirically validated intervention that has been demonstrated to be effective with youth who have been removed from their homes because of their extensive antisocial behavioral histories. MTFC has been supported with the results of an independent cost-benefit analysis by the Washington State Public Policy group [4] and evidence from several randomized trials (eg, [5-7]). In addition, the MTFC model has been selected as 1 of 10 evidence-based national model programs (The Blueprints Programs [8]) in the United States by the Office of Juvenile Justice and Delinquency Prevention and as 1 of 9 national Exemplary and Promising Safe, Disciplined, and Drug-Free Schools model programs. MTFC was also highlighted in two US Surgeon General reports [9,10] and was selected by the Center for Substance Abuse Prevention and Office of Juvenile Justice and Delinquency Prevention as an Exemplary I program for Strengthening America’s Families [11]. Early study participants have now been followed up as much as 10 years after initial MTFC participation, and treatment effects appear to be sustained for some time following the intervention [12].

The basic MTFC model involves placing an adolescent in a well-trained and supervised foster home where MTFC parents have undergone 20 hours of preservice orientation and training after having been certified by the state to be foster parents. Experienced case managers with small caseloads (8-10 MTFC families each) maintain daily contact with MTFC parents to provide ongoing consultation, support, and crisis intervention. Case managers coordinate all aspects of the youth’s placement, including ensuring that interventions in the family and individual therapy and with the skills trainer are targeting key behaviors. Basic components of MTFC are (1) daily telephone contact (during the workweek) with MTFC parents using the Parent Daily Report (PDR) measure [13]), (2) participation of MTFC parents in a weekly group supervision and support meeting led by the case manager, (3) implementation by MTFC parents of an individualized daily point-and-level program for the adolescent placed with them, (4) individual therapy for the participating adolescent, (5) family therapy (for aftercare resource) focusing on parent management strategies, (6) close monitoring of school attendance, performance, and homework completion, (7) case management to coordinate the interventions in the MTFC, family, peer, and school settings, (8) round-the-clock on-call program staff availability to MTFC and biological parents, (9) psychiatric consultation as needed, and (10) weekly clinical team meetings where the multiple intervention components are coordinated.

Barriers to Adoption and Dissemination

Like many of the empirically validated treatments, MTFC requires a shift in the manner in which traditional mental health, juvenile justice, and child welfare services are conducted. Staffing, funding, and reporting are often quite different within MTFC from how these functions operate in the broader community service organizations that may choose to replicate the MTFC model. For example, several elements of the MTFC intervention program (eg, PDR data collection or weekly staff meetings) do not fall under typical Medicaid reimbursement codes. Most service agencies also do not have established infrastructures to recruit, train, and certify foster care providers. Therefore, the practical implementation of the MTFC model can meet with several practical barriers early in the implementation process.

Because of these barriers, early efforts to disseminate MTFC were difficult and not satisfactory. An organization, separate from the direct research and treatment endeavors, was created (in 2001) with the sole purpose of disseminating the MTFC program: TFC Consultants, Inc. TFC Consultants helped develop structured program manuals, wrote training materials, scripted training procedures, and recruited clinical and foster parent trainers. There are currently 30 organizations in the United States, Canada, the United Kingdom, Sweden, and Norway working at various stages of program development. It has been clear during this process that fidelity standards and processes were also needed to track the process of program implementation in addition to the functional elements of the program. For example, an Internet-based PDR data-gathering tool was developed to allow MTFC site consultants immediate access to behavioral reports of program youths. Reporting instruments were also developed for the weekly telephone consultations with the implementation sites, for quarterly reporting of fidelity, and most recently for site certification (by a separate research organization, the Center for Research to Practice). While MTFC has tracked fidelity in this way, other empirically validated programs either offer no tracking of program fidelity after the initial training or offer continual quality improvement support without any method to measure or reduce the costs of this support. The position of the MTFC developers has been that there must be a balance between not requiring service organizations to bear ongoing consultation expenses and needing to maintain fidelity for those running MTFC programs. For MTFC, the developers hold that the optimal balance involves giving implementation sites clear feedback about their position relative to fidelity standards during training, and then supporting a certification process where organizations can demonstrate that the program is meeting fidelity standards and is no longer in need of external support.

We present our findings on the development and the feasibility evaluation of an Internet-based treatment fidelity monitoring system (ITFMS) funded by a Phase I Small Business Technology Transfer grant from the US National Institute of Mental Health (#R42MH075174). The ITFMS has three primary components: (1) an agency-tailored secure MTFC website with multiple levels of access restrictions for data storage, and display of reports (both graphical and tabular) and observations for dynamic feedback, (2) computer-mediated forms for clinical (eg, daily reporting of child behavior) and supervisory activities (eg, program tasks such as therapy appointments and clinical meetings) tailored to the respondent with Internet-based and automated phone-based response formats, and (3) computer-mediated video observation of clinical meetings.

Methods

Development of User-Interface Programming

Using our experience with previous Internet-based interventions for parenting [14-17], type 2 diabetes [18,19], and smoking cessation [20], we developed a graphic user interface that is appealing and easy to navigate. The programming included the use of the PHP scripting language for the generation of dynamic content that is read from a MySQL database. The combination of Hypertext Markup Language (HTML) and cascading style sheets was used to structure and lay out the elements of a page. The development phase of the Phase I project involved designing and programming the video capture software, delineating a protocol for securely transferring the video-recorded observations, developing the database, and programming a secure website for data entry and display. The software (eg, automated video recording) was menu driven using the mouse and icons to minimize the need for keyboarding skills. The video recording software was developed using Flex (Adobe Systems Inc, San Jose, CA, USA) to create a Flash-based object to securely stream it to the server for review.

We captured the videos of clinician–foster parent interactions in clinical meeting settings to a digital format using a preloaded video capture Flash-based object. This produced high-quality, low-data-rate movies that take up minimal bandwidth. We captured the video at a resolution of 464 × 346 ppi, 15 fps, about 40–45 KB/s. As the MTFC intervention makes extensive use of video observation (each week, a clinical team supervision and a foster parent support group meeting, each lasting 2 hours, are recorded and transferred to the site consultant for review), the central task of developing the fidelity management system for the computer was to use videos in less performance-intensive elements (ie, reducing video resolution to improve performance in Flash using Adobe Flex).

Encryption and Internet Security Development

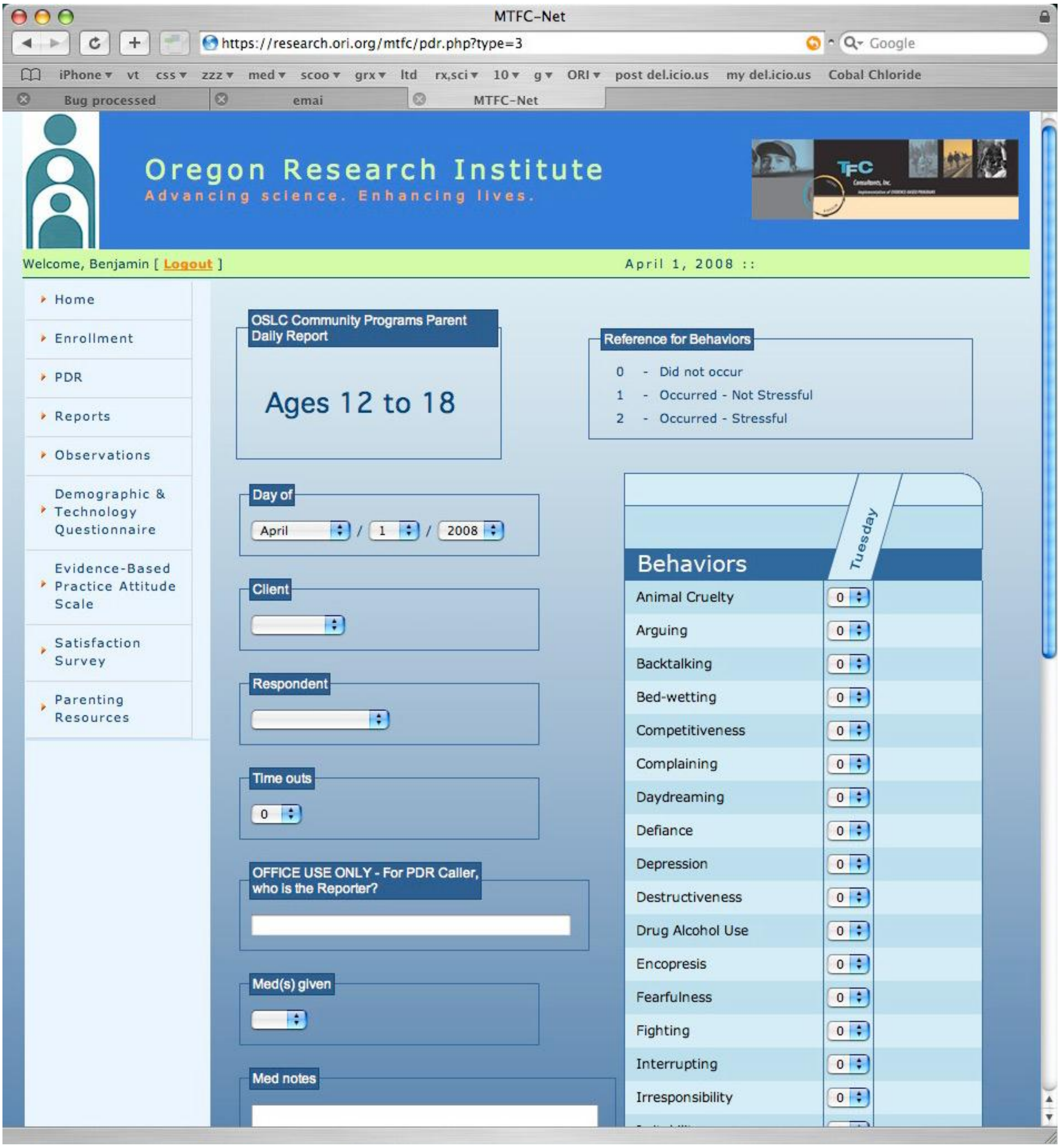
We used state-of-the-art security protocols in all data collection and monitoring activities, as is used for e-commerce, using secure sockets layer data encryption technology. The website was established within a suite of servers housed in an existing secure network and with independent firewall security. The website is accessible via the Internet with username and password protection. Once admitted to the site, each user has a hierarchically assigned set of permissions that allow the user to view or edit, or both, only that information that is necessary for his or her job functions. So, an individual foster parent may be allowed to enter PDR data, but would be able to do so only for the youth in his or her home (this can be reassigned to permit respite) and would not be able to alter or delete any data. Alternatively, a program supervisor has access to the data for the youth in his or her program. A consultant has access to data across programs assigned to him or her. PDR data are reviewed at weekly meetings between the foster parents and program supervisor to monitor accuracy.

Structure of the ITFMS

The ITFMS system contains 9 left-hand, navigational menu options, 6 of which were developed. First, the *Home* option provides a welcome to the system and a brief description of what the system will provide the user. Second, the *Enrollment* option provides 3 submenus for (1) Enroll Agency, which serves as the initial system setup, including information such as agency name, Web and street address, MTFC certification contact person information, the date of MTFC program initiation at the agency, and a place for specific MTFC role positions to be identified, (2) Enroll Client, which contains a place for first name only, age group related to MTFC programs (3–6 years, 7–11 years, and 12–18 years), and the month and year the client began the program, and (3) User, which includes a place for name, address, phone number, email address, and a username and password to get into the ITFMS system for PDRs (these pieces of information were designed to allow for the entry of pieces of data necessary for successful completion of the MTFC certification process). Third, the *PDR* option provides 3 submenus reflecting MTFC age groups (3–6 years, 7–11 years, and 12–18 years), to which a foster parent has access based on the developmental level of the youth in his or her home. The list of available options is created at each log-in. Beside each item listed (eg, animal cruelty) is a pull-down menu with options 0, 1, and 2 (0 = behavior did not occur; 1 = behavior did occur and was not stressful; and 2 = behavior did occur and was stressful) from which to select an answer (see Figure 1). Fourth, the *Reports* option provides a client pull-down menu that allows agencies and consultants to select the foster adolescent for whom they wish to view PDR and Behavioral Points summaries, as well as a date window from which a specific week can be selected through the use of a calendar image. Through this report system, a week's worth of PDR and Behavioral Points data, entered through any system (Internet, phone, or agency caller), can be displayed by day of the week, indicating for each day the number of Time Outs Given, Behavioral Points Earned for good behavior, Behavioral Points Lost for inappropriate behavior, Total PDR Behaviors Reported as occurring and stressful, Total Intensity of these behaviors, the Interviewer and Respondent for each day, as well as whether any medications were given. Fifth, the *Observation* option provides 2 submenu options for recording or playing. When each option is selected, a video screen appears, with the option to record video or play video. Additionally, on the play video page, a bland tabular window is presented with a pull-down menu that allows users to select the type of videos they wish to view, whether clinical or foster parent meetings. When type of video is selected, the rows in the tabular fields are filled in with a list of sessions recorded and, for each session, a display name for that video file. When selected, the video file is retrieved from the database and loaded into the video screen for reviewing. This retrieval system allows multiple users (eg, clinicians and consultants) to simultaneously stream recorded video from multiple locations for direct supervision (see Figure 2).

Within the PDR system, foster parents are allowed to decide whether they wish to report their PDR data using the Internet-based system, using the phone-based interface, or having an agency caller contact them. Having options provides parents with convenience in terms of their time and comfort with technology. MTFC practice is to have an individual caller contact each foster parent daily and ask a series of standardized questions about the youth's behavior during the preceding 24 hours. With the Web-based system, the respondent simply enters the PDR information directly, using a series of drop-down menus (as described above). With the phone-based system, the respondent calls into the system, enters a pass code (that authorizes access to the system and links the caller to a specific youth), and then is prompted with a standard series of behaviors to report using the phone's keypad relative to whether the behavior occurred. All 3 formats collect the same data, and data are stored in the same format within the ITFMS, but the use of the human caller creates scheduling limitations and an additional level of potential data-entry errors. In addition, the systems are set to allow multiple systems to be used, such that a foster parent who does not enter data by a given time might be provided with an automated call. The figures show pages from the ITFMS system, one reflecting the video recording of an agency clinical staff meeting (Figure 2), as viewed by consultants, and the other reflecting the PDR page wherein foster parents provide their daily behavior ratings and the number of behavior points given or taken away (Figure 1).

Figure 1. Parent Daily Report form in the Internet-based treatment fidelity monitoring system.



OSLC Community Programs Parent Daily Report

Ages 12 to 18

Reference for Behaviors

0 - Did not occur
1 - Occurred - Not Stressful
2 - Occurred - Stressful

Day of: April / 1 / 2008

Client:

Respondent:

Time outs: 0

OFFICE USE ONLY - For PDR Caller, who is the Reporter?

Med(s) given:

Med notes:

Behaviors	Tuesday
Animal Cruelty	0
Arguing	0
Backtalking	0
Bed-wetting	0
Competitiveness	0
Complaining	0
Daydreaming	0
Defiance	0
Depression	0
Destructiveness	0
Drug Alcohol Use	0
Encopresis	0
Fearfulness	0
Fighting	0
Interrupting	0
Irresponsibility	0

Figure 2. Video reviewing screen showing a sample video of a clinical meeting.

Session	Display Name
00	09/19/2008, 3:31:17 pm
00	10/24/2008, 2:06:40 pm
00	11/14/2008, 2:44:22 pm
00	11/17/2008, 12:25:55 am
00	11/20/2008, 04:31:30 am
00	11/20/2008, 09:42:49 am
00	11/20/2008, 10:02:29 am
00	11/20/2008, 1:12:27 pm

Measures

We assessed the ITFMS with measures administered to parent and professional users. Computer-administered measures included demographics, child behavior, and satisfaction. All agency supervisors and foster parents rated the ease of use of the system on a 7-point Likert-type scale from 1 = very difficult to 7 = very easy. MTFC consultants rated video clips relative to audio and video quality on 5-point scales anchored by 1 = impossible, 3 = codeable, and 5 = easy to code. User tracking measures of log-ins and time on website were collected automatically by the server. Data were collected from 20 foster parent PDR users, 9 professional MTFC program supervisors, and 4 MTFC consultants, who rated the quality of observational videos recorded from agency meetings. The foster parents completed PDR questionnaires on the behavior of 20 foster children in their care. As noted above, encryption and permissions allowed foster parents to view only their own data. Staff were allowed to view PDR data only on children on their caseload.

Results

Participants

Foster parents were on average 33.2 years old, had 3.6 people in their household, and had a median family income range of US $40,000–$45,000; 14 (70%) had completed high school, with 6 (30%) reporting attending college; and 15 (75%) of the parents reported working outside of the home, with 4 (20%) indicating that were not working, by their own choice. All of the participating parents were white, with 1 parent (5%) reporting their ethnicity as Hispanic/Latino; 18 (90%) of the parents reported being married or living together with another as a couple. Of the participating parents, all reported having a working telephone, and 18 (90%) indicated they had a computer at home. Parents reported high levels of computer use, with 16 (80%) indicating more than 10 hours per week, 12 (60%) indicating more than 20 hours per week, and only 4 (20%) reporting 5 or less hours per week. Of the parent PDR users, only 4 (20%) had computer access through their work, and so computer use was predominantly at home.

Professional participants, both agency and consultant, were on average 36.3 years old and had a median family income range of US $50,000–$60,000; 11 (85%) had completed a postgraduate degree. A total of 10 (77%) of the professionals reported they were white, with the remainder declining to state their race; 1 (11%) of the professionals reported being Hispanic/Latino. In addition, 10 (91%) of the professionals reported being married or living together with another as a couple. All professionals reported owning a phone and having a computer at home. Much like the parents, professionals reported high levels of computer use, with 9 (69%) reporting more than 20 hours per week and only 1 (8%) reporting 5 or less hours per week; 11 (85%) of the professionals reported having access to a computer at work.

Ease of Use

All agency supervisors and foster parents rated the ease of use of the system as very easy to use. When asked if they would recommend the system to other staff or parents, on a 7-point

Likert-type scale from strongly recommend to strongly not recommend, all responses were strongly recommend (score of 7) or recommend (6). All individuals further reported a timesaving benefit on the PDR behavioral reports, either for entering data (parents) or for viewing summaries (consultants), relative to the typical PDR method (ie, PDR caller). Comments were overwhelmingly positive. Example comments included "I definitely enjoy this better than a caller who calls me at home! Thanks!" (parent); "The system seems very well designed and is easy to look at and find what you are looking for!" (parent); "It would be interesting to compare the data that we get from a live person asking questions" (staff); "The whole process is very simple to use" (staff); "Very good support, thanks!!;" and "I like doing PDR on my own time schedule."

ITFMS Utilization

For both the PDR (parents) and Report (supervisor) functions, the Phase I trial was designed to last 1 week. Thus, submission of 7 PDRs via the Web-based or automated telephone PDR system indicated completion of the trial. All 20 parents completed the trial, some providing data for multiple children in the home. Reports were submitted for 28 foster children, and even though the trial was intended for week, a high number of parents continued on past the 7-day trial. Some parents enjoyed using the Web-based entry because it allowed them to give their daily reports at times convenient to them, rather than having to schedule a call with staff within their busy schedules. As many states require foster parents to demonstrate sources of income outside of child welfare system support payments, this flexibility is increasingly important. A total of 12 parents completed more than 14 PDRs, with 6 families completing at least 21. Utilization was particularly high among 2 parents, who completed 145 and 176 PDRs, respectively, on their children. A total of 667 PDRs were completed during the trial. In terms of supervisor program utilization, all supervisors reported using the Web-based PDR system to view weekly behavior summaries. In addition, 4 (36%) reported using Web-based viewing of meeting observations and 7 (64%) reported using the Web-based enrollment and administrative features.

Satisfaction

Of the 9 supervisors, 8 (89%) found the overall system very easy to use, with all 9 indicating the system was either fairly easy or very easy to use. Supervisors also rated the Web-based PDR system between fairly easy and very easy to use. Similarly, when using the Web-based viewing of observations feature, all 9 supervisors found it very easy or fairly easy to use. Of those who used the Web-based enrollment and administrative features, 6 (67%) found it very easy or easy to use, with only 1 individual indicating a neutral rating. Of the professionals, 7 (78%) rated the system as much easier than previous tracking methods they had used, with an additional 1 (11%) indicating it was easier and 1 (11%) indicating they were neutral about this method. All professionals indicated they would recommend this system to other programs or foster parents.

Foster parents also reported high levels of satisfaction. Using a 7-point scale from very easy to very difficult, 18 (93%) described the systems as easy or very easy to use. Regarding the system features, 19 (95%) of parents used the Web-based PDR feature and 2 (10%) reported using the automated telephone PDR system (the greater than 100% sum reflects parents who used both methods across the 7 days). All users reported the Web-based PDR was fairly easy, easy, or very easy to use. Only 1 user rated the automated telephone PDR system as fairly difficult. All parents rated the PDR system as somewhat easier, easier or much easier than the previously used tracking methods, and all parents indicated they would either recommend or strongly recommend this system to other foster parents.

Quality of Video Observation

Recording and viewing the videos of the clinical and foster parent meetings largely involved balancing factors such as frame size, image resolution, frame speed, and audio quality with the intent of making a viewable image that does not make file transfer prohibitive. As the meetings were scheduled for 2 hours and there were generally quite a few people involved, we attempted to push the limits of file size given the current technology and the information technology resources available at average community mental health centers. The research team worked with numerous microsocial coding schemes [21,22] and determined that the video recordings of clinical team and foster parent support group meetings that were made through the system were suitable for fidelity coding for ratings and fidelity checks. While the frame size and image quality would not be suitable for microsocial coding of facial expression and affect, it was reported to be more than suitable for MTFC fidelity coding. The inclusion of a broad-range array microphone during testing appeared to enhance the ability of the system to capture audio content from the numerous individuals in the confined meeting rooms.

With respect to the videos of clinical staff meetings that were recorded for observation, 4 MTFC consultants rated 3 video clips relative to audio and video quality on 5-point scales anchored by 1 = impossible, 3 = codeable, and 5 = easy to code. The video content was rated an average of 4.75 and the audio content an average 4.42. Consultants also indicated a dichotomous (yes/no) response as to whether each segment was codeable, with 100% reporting yes.

Discussion

We developed a user-friendly and secure Web-based system using state-of-the-art computer-based protocols for recording questionnaire and observational data generated by community-based MTFC staff and foster parents. We are greatly encouraged by the positive satisfaction and utilization results, and we believe these results provide a good evidence base for further evaluation in a randomized controlled trial in future research. Consistent with other studies, our study found that reducing cost and increasing convenience for the user were given as reasons for delivery over the Internet [23]. We believe that these positive results are due to following development guidelines described by other researchers. We followed the methodology used for the development of ClinicalTrials.com as delineated by McCray and colleagues [24]: (1) internal testing by the development team, (2) review by collaborators and consultants, (3) user testing with participants, and (4) measurement of participant Web accessibility. As well, we adapted a development methodology outlined by Ritterband and colleagues [25]: (1) identify the problem area, (2) find an empirically supported intervention, (3) operationalize the treatment, (4) consider the legal issues, (5) design engaging components, (6) tailor them as appropriate, (7) incorporate feedback, (8) develop a Web-based program, and (9) evaluate the program. For future development work, we have high expectations, with Web technology being advanced by the advent of video recording and HTML5 and using these guidelines for development.

For many Internet-based mental health projects, most ethical considerations are limited to issues of security and confidentiality. In this study we followed all secure sockets layer security and institutional review board standards, yet the overall ethical considerations of whether to use technology in this application is still in question. Internet delivery overcomes isolation due to time, mobility, and geographical constraints, but it may not be a substitute for face-to-face contact [23]. For the foster parents' report of child behavior, there are many anecdotal reports of assistance provided through person-to-person telephone communication using a live PDR caller. In future studies with a larger sample, we will investigate whether the lack of a live caller reduces the opportunity for person-to-person communication.

While the system was created for the specific requirements of MTFC, the usefulness of the system with other validated programs is clear (eg, the ability to provide quick access to securely transmitted clinical session information can help with the remote supervision of therapists from a variety of treatment models). In addition, while the scope of this pilot project called for a demonstration of the feasibility of the system, several foster parents did not want to return to being called for behavioral reports on their foster children, and they continued to enter the information on this Web-mediated system. Increasing the options for foster parents to complete these program requirements appears to have resulted in greater compliance and more consistent completion of the PDR for some foster care providers, a critical aspect for fidelity. These increases in behavioral data will allow program supervisors to more closely match interventions to the current presentation of the involved youth, and to match program reactions to the needs and stress experienced in the foster home, thereby increasing foster parent retention while reducing the response cost for complying with the program requirements. Given that fidelity of these empirically validated treatments has been shown to be related to the quality of their outcomes, it is critically important to develop innovative service delivery systems such as that developed herein that increase fidelity in the dissemination of these proven treatment approaches.

Acknowledgments

Research supported by US National Institute of Mental Health grant #R42MH075174.

Conflicts of Interest

None declared.

References

1. Chamberlain P. Treating Chronic Juvenile Offenders: Advances Made Through the Oregon Multidimensional Treatment Foster Care Model (Law and Public Policy: Psychology and the Social Sciences). Washington, DC: American Psychological Association (APA); 2003.
2. Mihalic S, Irwin K, Fagan A, Ballard D, Elliott D. Successful program implementation: lessons from blueprints. Juv Justice Bull 2004;Jul:1-11 [FREE Full text]
3. Henggeler SW. Decreasing effect sizes for effectiveness studies- implications for the transport of evidence-based treatments: Comment on Curtis, Ronan, and Borduin (2004). J Fam Psychol 2004 Sep;18(3):420-423. [doi: 10.1037/0893-3200.18.3.420] [Medline: 15382966]
4. Aos S, Phipps P, Barnoski R, Lieb R. The Comparative Costs and Benefits of Programs to Reduce Crime: A Review of National Research Findings With Implications for Washington State, version 3.0. Olympia, WA: Washington State Institute for Public Policy; 1999. URL: http://www.wsipp.wa.gov/rptfiles/CB3.pdf [accessed 2012-09-28] [WebCite Cache ID 6B1R09tpq]
5. Chamberlain P, Reid JB. Using a specialized foster care community treatment model for children and adolescents leaving the state mental hospital. Journal of Community Psychology 1991;19(3):266-276. [doi: 10.1002/1520-6629(199107)19:3%3C266::AID-JCOP2290190310%3E3.0.CO;2-5]
6. Chamberlain P, Reid JB. Comparison of two community alternatives to incarceration for chronic juvenile offenders. Journal of Consulting and Clinical Psychology 1998 Aug;66(4):624-633. [doi: 10.1037/0022-006X.66.4.624] [Medline: 9735578]
7. Eddy JM, Chamberlain P. Family management and deviant peer association as mediators of the impact of treatment condition on youth antisocial behavior. Journal of Consulting and Clinical Psychology 2000 Oct;68(5):857-863. [doi: 10.1037/0022-006X.68.5.857] [Medline: 11068971]
8. Elliot DS. Blueprints for Violence Prevention. Boulder, CO: Center for the Study and Prevention of Violence, University of Colorado Boulder; 1998. URL: http://www.colorado.edu/cspv/publications/#blueprints_publications [accessed 2012-09-28] [WebCite Cache ID 6B1Ru0Yjb]

9. US Department of Health and Human Services. Mental Health: A Report of the Surgeon General. Rockville, MD: US Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institutes of Health, National Institute of Mental Health; 1999. Children and mental health URL: http://137.187.25.243/library/mentalhealth/pdfs/c3.pdf [accessed 2012-09-28] [WebCite Cache ID 6B1SR12p1]
10. US Department of Health and Human Services. Youth Violence: A Report of the Surgeon General. Rockville, MD: US Government Printing Office; 2000. URL: http://www.ncbi.nlm.nih.gov/books/NBK44294/ [accessed 2012-09-28]
11. Chamberlain P. Treatment foster care. Juv Justice Bull 1998 Dec;Dec:1-11 [FREE Full text]
12. Eddy JM, Whaley RB, Chamberlain P. The prevention of violent behavior by chronic and serious male juvenile offenders: A 2-year follow-up of a randomized clinical trial. J Emot Behav Disord 2004;12(1):2-8.
13. Chamberlain P, Reid JB. Parent observation and report of child symptoms. Behav Assess 1987;9:97-109.
14. Baggett KM, Davis B, Feil EG, Sheeber LB, Landry SH, Carta JJ, et al. Technologies for expanding the reach of evidence-based interventions: Preliminary results for promoting social-emotional development in early childhood. Topics Early Child Spec Educ 2010 Feb 1;29(4):226-238. [doi: 10.1177/0271121409354782] [Medline: 20454545]
15. Feil EG, Baggett KM, Davis B, Sheeber L, Landry S, Carta JJ, et al. Expanding the reach of preventive interventions: development of an Internet-based training for parents of infants. Child Maltreat 2008 Nov;13(4):334-346. [doi: 10.1177/1077559508322446] [Medline: 18843143]
16. Feil EG, Severson H, Taylor T, Bowles S, Webster-Stratton C. Computer based Incredible Years Parent Training for parents of Head Start children: challenges and outcomes. 2005 May Presented at: Society for Prevention Research; May 2005; Washington, DC.
17. Taylor TK, Webster-Stratton C, Feil EG, Broadbent B, Widdop CS, Severson HH. Computer-based intervention with coaching: an example using the Incredible Years program. Cogn Behav Ther 2008;37(4):233-246. [doi: 10.1080/16506070802364511] [Medline: 18803072]
18. Barrera M, Glasgow RE, McKay HG, Boles SM, Feil EG. Do Internet-based support interventions change perceptions of social support?: An experimental trial of approaches for supporting diabetes self-management. Am J Community Psychol 2002 Oct;30(5):637-654. [Medline: 12188054]
19. Feil EG, Glasgow RE, Boles S, McKay HG. Who participates in Internet-based self-management programs? A study among novice computer users in a primary care setting. Diabetes Educ 2000;26(5):806-811. [Medline: 11140008]
20. Feil EG, Noell J, Lichtenstein E, Boles SM, McKay HG. Evaluation of an Internet-based smoking cessation program: lessons learned from a pilot study. Nicotine Tob Res 2003 Apr;5(2):189-194. [Medline: 12745491]
21. Davis B, Sheeber LB, Hops H. Coercive family processesadolescent depression. In: Reid JB, Patterson GR, Snyder J, editors. Reid JB, Patterson GR, Snyder J, editors. Antisocial Behavior in Children and Adolescents: A Developmental Analysis and the Oregon Model for Intervention. Washington, DC: American Psychological Association; 2002:173-192.
22. Reid JB, Patterson GR, Baldwin DV, Dishion TJ. Observations in the assessment of childhood disorders. In: Rutter M, Tuma AH, Lann IS, editors. Rutter M, Tuma AH, Lann IS, editors. Assessment and Diagnosis in Child Psychopathology. New York, NY: Guilford; 1988:156-195.
23. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
24. McCray AT, Dorfman E, Ripple A, Ide NC, Jha M, Katz DG, et al. Usability issues in developing a Web-based consumer health site. Proc AMIA Symp 2000:556-560. [Medline: 11079945]
25. Ritterband LM, Gonder-Frederick LA, Cox DJ, Clifton AD, West RW, Borowitz SM. Internet interventions: in review, in use, and into the future. Prof Psychol Res Pr 2003;34(5):527-534. [doi: 10.1037/0735-7028.34.5.527]

Abbreviations

HTML: Hypertext Markup Language
ITFMS: Internet-based treatment fidelity monitoring system
MTFC: Multidimensional Treatment Foster Care
PDR: Parent Daily Report

Edited by G Eysenbach; submitted 20.12.11; peer-reviewed by B Hedman Ahlström, R Morais Lirio; comments to author 29.03.12; revised version received 02.06.12; accepted 18.06.12; published 16.10.12.

Please cite as:
Feil EG, Sprengelmeyer PG, Davis B, Chamberlain P
Development and Testing of a Multimedia Internet-Based System for Fidelity and Monitoring of Multidimensional Treatment Foster Care
J Med Internet Res 2012;14(5):e139
URL: http://www.jmir.org/2012/5/e139/
doi:10.2196/jmir.2034
PMID:23073495

Original Paper

Communicating the Experience of Chronic Pain and Illness Through Blogging

Pamela Katz Ressler[1*], RN, MS, HN-BC; Ylisabyth S Bradshaw[1,2*], DO, MS; Lisa Gualtieri[2*], PhD, ScM; Kenneth Kwan Ho Chui[3*], PhD, MS/MPH

[1]Pain Research, Education and Policy Program, Department of Public Health and Community Medicine, Tufts University School of Medicine, Boston, MA, United States

[2]Health Communication Program, Department of Public Health and Community Medicine, Tufts University School of Medicine, Boston, MA, United States

[3]Department of Public Health and Community Medicine, Tufts University School of Medicine, Boston, MA, United States

[*]all authors contributed equally

Corresponding Author:
Pamela Katz Ressler, RN, MS, HN-BC
Pain Research, Education and Policy Program
Department of Public Health and Community Medicine
Tufts University School of Medicine
136 Harrison Ave.
Boston, MA, 02111
United States
Phone: 1 978 369 5243
Fax: 1 978 369 5853
Email: pamela.ressler@tufts.edu

Abstract

Background: Although more individuals are sharing their experiences with chronic pain or illness through blogging (writing an Internet web log), research on the psychosocial effects and motivating factors for initiating and maintaining a blog is lacking.

Objective: The objective was to examine via online questionnaire the perceived psychosocial and health benefits of blogging among patients who use this media to communicate their experience of chronic pain or illness.

Methods: A 34-item online questionnaire was created, tested, and promoted through online health/disease forums. The survey employed convenience sampling and was open from May 5 to July 2, 2011. Respondents provided information regarding demographics, health condition, initiation and upkeep of blogs, and dynamics of online communication. Qualitative data regarding respondents' blogging experiences, expectations for blogging, and the perceived effects from blogging on the blogger's health, interpersonal relationships, and quality of life were collected in the form of written narrative.

Results: Out of 372 respondents who started the survey, 230 completed the entire questionnaire. Demographic data showed survey respondents to be predominantly female (81.8%) and highly educated (97.2% > high school education and 39.6% with graduate school or professional degrees). A wide spectrum of chronic pain and illness diagnoses and comorbidities were represented. Respondents reported that initiating and maintaining an illness blog resulted in increased connection with others, decreased isolation, and provided an opportunity to tell their illness story. Blogging promoted accountability (to self and others) and created opportunities for making meaning and gaining insights from the experience of illness, which nurtured a sense of purpose and furthered their understanding of their illness.

Conclusions: Results suggest that blogging about chronic pain and illness may decrease a sense of isolation through the establishment of online connections with others and increases a sense of purpose to help others in similar situations. Further study involving a larger sample size, a wider range of education levels, and respondents with different types and magnitudes of illnesses will be needed to better elucidate the mechanism of the observed associations in this understudied area.

(J Med Internet Res 2012;14(5):e143) doi:10.2196/jmir.2002

KEYWORDS

Blogging; narrative medicine; disease management; Internet; pain; chronic illness; survey; psychosocial support systems; holistic health; selfcare

Introduction

Chronic pain and illness are universal disorders. Globally, over 1.5 billion individuals suffer from chronic pain [1]. In the United States, 1 out of every 2 adults, or 133 million individuals, are living with at least one chronic disease [2], and 1 in 10 Americans live with severe chronic pain. Although we have made great strides in pain and disease management, chronic pain and illness remain costly health burdens and challenges for clinicians, patients, families, and society. In the United States alone, 25 million individuals report major limitations in daily living due to chronic disease [2]. Recent data suggest that 116 million US adults suffer from chronic pain creating an economic burden of $560 to $635 billion annually [1]. Advances in medical treatment allow many individuals to survive for a significant time with diseases that previously would have quickly resulted in death. Patients and families face challenges managing the physical as well as the social and emotional consequences of chronic pain and illness for extended lengths of time, often years or decades. To address this new paradigm of medical care for chronic conditions, it will become increasingly necessary for our health care system to identify and adopt effective interventions that could provide support to patients and families throughout their journey of diagnosis, treatment, and management.

Disciplines across the health care professions, including medicine, nursing, psychology, social work, and sociology, are actively engaged in understanding the psychosocial and emotional consequences of chronic pain and illness [3]. Methods of creating reflection, connection, and finding meaning and understanding through use of illness narratives are advocated for patients and family members. Extensive medical and social science research has examined the use of patient narratives as methods of telling the story of illness and health [4-8]. Review of the literature on expressive writing suggests there are psychological and physical benefits associated with patients engaging in the process of expressive writing [9-15]. In her book, *Narrative Medicine: Honoring the Stories of Illness*, Charon [16] describes the experience of patients and families as they enter the divide between the sick and the well, a separation that often feels huge and unbridgeable to those navigating a new way of being:

> *“These divides between the sick and the well are unspeakably wide. Leveraged open by shame, rage, loss, and fear, these chasms can be unbridgeable. And yet, to get better, the patient needs to feel included among those who are not ill. The sick person needs to continue to be, somehow, the self he or she was before illness struck. For the sick patient to accept the care of well strangers, those strangers have to form a link, a passage between the sick and the healthy who tender care.”*

Chronic pain and illness may have an isolating effect on individuals, changing their perceived roles in society and challenging their ability to find meaning in life and their illness. Creating connections between the world of the sick and world of the well can be important in the positive psychosocial functioning of individuals.

Individuals with chronic pain or illness may have various purposes in articulating illness narratives. Perceived benefits may be to acknowledge and validate the experience of pain and illness, express beliefs and values, create a temporal nature of the experience of disease, and make visible the often-invisible nature of pain. Communicating the experience of pain and illness may facilitate sharing between teller and listener, allow for reflection on the experience, and acknowledge the emotional and suffering component of pain. Bridging this chasm, as Charon suggests, may be an important part of finding a sense of wellness within illness.

Traditionally, communicating the experience of pain or illness was oral or written on paper. As Pennebaker [11] and others have suggested, expressing emotional or traumatic experiences with the written word through the process of expressive writing, may have therapeutic benefit for individuals. Although expressive writing in medicine has been researched over the past decades, the digital communication platform of the Internet has expanded the ability to express emotional experiences of pain or illness to a larger, real-time audience and for those experiences to not only to be expressed, but to be shared and commented on by others. Use of Internet-based tools such as blogs (web-based journals with user-generated content) allow this expression to take place.

The Pew Internet study found that 78% of adults in the United States are online [17]. A digital divide of Internet usage exists between individuals with and individuals without chronic disease. Internet users who report no chronic disease are more likely (81%) than those with one or more chronic illnesses (61%) to be online [18]. However, once online, having a chronic illness increases the likelihood of interacting through social media, such as blogs, to share knowledge with peers. When other demographic factors are controlled, having a chronic illness significantly increases the probability of contributing to a blog or online discussion [18]. For those with chronic disease who consume health information online, 37% responded that they read other’s commentaries or medical/health experiences on online newsgroups, websites, or blogs, and 6% have posted health comments to a blog. As a form of online community, blogs function as platforms for patients with chronic pain and illness to share information, practical advice, and emotional support [18]. However, research on their attitudes and perceived benefits of blogging has been lacking. By using tools such as blogs, patients and families may find a new way to bridge this divide between the sick and the well. Illness blogs may allow a patient to articulate his or her narrative with the advantage of online interaction with others—family, friends, and other patients with similar health concerns. Illness blogs have similarities with traditional forms of illness narrative; however, features of blogs, in particular the ease of sharing or commenting, are dissimilar to illness narratives. Narrative writing has a long history of being self-initiated and directed, some of its recent attention has been its effectiveness as a directed activity, guided in part by health care professionals as a therapeutic intervention. No research to date has indicated whether illness narratives through blogging are potentially adaptable as such a therapeutic process.

The goal of this study was to explore the use of patient illness blogs as a means of communicating the experience of chronic illness and pain and to articulate their unique set of benefits and barriers. We surveyed current illness bloggers to understand their self-perceived psychosocial and health effects associated with the blogging activity. We hypothesized that communicating the experience of illness through blogging may provide perceived positive psychosocial benefits to some patients with chronic pain or illness.

Methods

As a primary step in the formulation of this project, a research theory in accordance with the discipline of nursing (the training of the first author) was identified. Nursing research has focused on the healing process within the context of the individual, environment, and community [19]. Nurse theorist Margaret Newman's *Health as an Expanding Consciousness* was chosen as a framework for this study based on its assumption that health is not simply the absence of disease [20]. Further recognizing that health is an awareness or consciousness of the evolving interaction between the individual and his or her environment, and is possible regardless of presence or absence of disease. The theory states, "every person in every situation, no matter how disordered and hopeless it may seem, is part of the universal process of expanding consciousness—a process of becoming more of oneself, of finding a greater meaning in life, and of reaching new dimensions of connectedness with other people and the world" [20]. Newman's theory defines this paradigm shift as moving away from only treatment of symptoms and moving toward recognition of pattern and connection by assisting individuals in viewing disease and disruption as part of a self-organizing dynamic process of health [21]. Theoretical constructs of finding greater meaning in life, reaching new dimensions of connectedness, and expanding views of illness and health were incorporated in the planning and development of a survey tool.

We developed a 34-item online questionnaire based on our theoretical constructs to examine the initiation, motivation, continuation, reinforcement, support, and process of blogging. We also queried our respondents on the impact of chronic illness and pain on their life and activities of daily living in the context of illness blogging and, if this changed over time, how it was manifested in their blogging. We asked how the illness or chronic pain impacted their life and queried them regarding their efforts at assisting other individuals as well as developing and maintaining support networks. Questions explored psychosocial themes of coping, isolation, connection, and support. Furthermore, we asked for respondents' interactions with health care providers about their blogging and their recommendations to others about blogging. Multiple choice, matrix, and open-ended essay or comment questions were incorporated on one scrollable screen with the goal of optimizing the depth of responses. All respondents answered the same questionnaire; no adaptive design or timing was employed. Question order was not randomized because some questions followed directly from others. Respondents were also allowed to skip to the next question without providing an answer. The open survey was formatted with a web-based survey design tool, SurveyMonkey [22]. SurveyMonkey uses secure sockets layer (SSL) data encryption and it recognizes and prevents repeated access based on Internet protocol (IP) addresses. A draft questionnaire was reviewed and pretested by a sample of three bloggers who provided qualitative assessments, allowing for small refinements to the survey, and the resulting questionnaire was used (Appendix 1).

After receiving approval from the Tufts Medical Center Institutional Review Board (IRB), the recruitment script and URL survey link was circulated among online health forums and online illness support groups, developing a compounding "snowball" convenience sample from Boston-centric academic and social media contacts. The authors also circulated the request for participation on the social media sites Twitter, Facebook, WEGO Health [23] (forums included chronic pain, fibromyalgia, and health bloggers subgroups of WEGO Health), and LinkedIn (American Pain Society discussion list). The disseminated request was in the following format (or abbreviated to fit into social media formats): "Do you blog about your illness or do you know someone who does? Help researchers at Tufts University School of Medicine better understand patient blogging by participating in a short online survey. Please forward the survey link https://www.surveymonkey.com/s/TuftsPatientBloggingSurvey to your friends and colleagues." Although not promoted in the requests, a drawing for a prize of two $25 online shopping gift certificates was promoted on the consent page, which also included information about the purpose of the study, approval from the IRB, and the investigators' contact information. The survey remained open for participation from May 5 to July 2, 2011, resulting in 230 completed surveys by the end of the survey timeframe.

Quantitative responses to the questionnaires were tallied and tabulated; text responses were analyzed qualitatively in an iterative process, using grounded theory principles to develop primary and subthemes [24]. The first two authors (PKR and YSB) separately reviewed the narrative responses and agreed on primary and secondary themes within each question. Categorizations were reviewed, differences were discussed, and then they were re-sorted independently until agreement was reached. Quotations were selected to represent the core themes, excluding direct quotes containing specific details or identifying references. We adhered to the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) in reporting the results of this online questionnaire [25]. No weighting scheme was used.

Results

Response Statistics

Out of the 526 submissions, 379 agreed to consent, 1 disagreed, and 146 did not answer the consent question. Out of the 379 consented submissions, 149 did not answer any questions but applied for the prize with an email address, and hence were excluded from the analysis. The effective sample size is 230 and the completion rate is 60.7% (230/379). Due to the open sample frame, no response rate was computed.

Characteristics and Blogging Behaviors

As part of the survey design, respondents were permitted to skip any question within the survey, so the number of responses varied by question. Table 1 shows the demographics of the participants.

Respondents identified the disease or conditions prompting the blog. The varied responses had common features of chronicity and potential for associated suffering and pain. Diseases represented included cancer (non-specified, ovarian, breast, and leukemia), fibromyalgia, non-specified chronic pain conditions, diabetes, Addison’s disease, bipolar disorder, celiac disease, trigeminal neuralgia, Parkinson’s disease, systemic lupus erythematosus (SLE), rheumatoid arthritis, Sjögren’s syndrome, schizophrenia, cystic fibrosis, cerebral palsy, as well as other chronic diseases and syndromes. Several respondents reported comorbidities, such as fibromyalgia and depression, bipolar disease and chronic pain, cerebral palsy and trigeminal neuralgia, breast cancer and chronic depression, and diabetes and clinical depression.

Table 2 shows blogging-specific behaviors of respondents. The survey asked questions regarding the initiation and maintenance of the participants’ blogs. Most bloggers (60.0%) had initiated their blogs on their own, and the majority of the blogs (87.8%) were public and searchable on the Internet. Although 64.1% of the bloggers used their own names in blogging, 35.8% chose to protect their identity by use of a pseudonym/pen name or by blogging anonymously. Although 89.7% of the respondents shared their blogs with friends and family members, less than half (42.1%) shared their blogs with their health care providers.

Theme analysis

Open-ended questions (with unlimited space in which to answer) explored respondents’ attitudes, motivations, and explanations. These narrative responses were qualitatively analyzed and grouped into primary themes according to the major ideas and concerns expressed.

The 57.9% of respondents who had not shared their blogs with their health care providers were asked why they made this choice. Their narrative answers provided information that was sorted into four primary themes: (1) negative concern (ie, respondent expressed concern that a negative outcome would result if the blog was shared), (2) perceived lack of provider interest and/or time to read the blog, (3) interest or intention to share the blog with their health care provider, and (4) desire to write freely about the illness experience.

In primary theme 1, negative concern, five subthemes were identified: (1) fear of provider judgment of patient feelings—concern of negative or pejorative judgment by the health care provider for the feelings expressed in the blog, (2) fear of provider judgment of patient behavior—concern that the provider would not approve of patient’s behaviors and/or lifestyle choices, (3) confidentiality/negative impact, (4) worry over self-editing, and (5) awkwardness in sharing negative opinions of health care provider and care (Table 3).

Table 1. Demographics of the survey participants (N = 230).

Demographic	n	%
Gender		
Female	172	81.5
Male	39	18.5
Age group		
18-25	19	9.0
26-39	71	33.8
40-55	89	42.4
56-65	25	11.9
66-75	6	2.9
Education		
Some high school	3	1.4
Graduated high school or GED	3	1.4
Some college	36	17.0
Graduated college	86	40.6
Graduate or professional degree	84	39.6

Table 2. Blogging-related behaviors of the survey participants (N = 230).

Blogging behavior	n	%
Who suggested blogging? (multiple choices allowed)		
No one	138	60.0
Family member	33	14.3
Friend	50	21.7
Work colleague	10	4.3
Support group	8	3.5
Health care provider	6	2.6
When did you start blogging?		
Before 2007	40	17.5
2007	17	7.5
2008	26	11.4
2009	48	21.1
2010	65	28.5
2011	32	14.0
Blog is...		
Open to public	194	87.8
Private, only individuals with password can access	27	12.2
Blog is written under...		
The author’s own name	143	64.1
A pen name or pseudonym	67	30.0
An anonymous identity	13	5.8
Blogging frequency		
More than once a day	9	4.3
Once a day	19	9.0
A few times a week	91	38.6
Once a week	41	19.5
Every few weeks	35	16.7
Once a month or less	25	11.9
Blogging frequency over time has...		
Increased	75	33.5
Stayed the same	68	30.4
Decreased	81	36.2
Have viewers ever left written comments?		
Yes	206	91.6
No	13	5.8
Not sure	6	2.7
Blogging platform/hosting used		
Blogger (by Google)	103	54.8
WordPress	71	37.8
CaringBridge	10	5.3
Others	4	2.1
Do you read other people’s health/illness blogs?		

Blogging behavior	n	%
Yes	214	95.1
No	11	4.9
Have you contributed comments on other's blogs?		
Yes	202	89.0
No	25	11.0
Shared own blog with...		
Family members or friends	192	89.7
Health care provider(s)	90	42.1
Participated in any online health/illness or patient communities?		
Yes	151	72.9
No	56	27.1
Has reading blog and blogging changed one's sense of connection with others?		
Yes	163	76.5
No	22	10.3
Not sure	28	13.1
Other social media used (multiple choices allowed)		
Facebook	195	84.8
Twitter	162	70.4
YouTube	108	47.0
Foursquare	26	11.3
Others (LinkedIn, Tumblr, WEGO Health, etc)	43	18.7

A minority of the bloggers (42.1%) had shared their blogs with their health care providers. The health care providers' responses to the illness blogs varied from positive to negative. Patient's perceptions of health care provider's reactions upon sharing illness blogs were grouped into three primary themes: (1) positive response, (2) negative response, and (3) neutral or indeterminate response (Table 4).

Motivation to start an illness blog clustered around three main themes: (1) reflection, (2) communication, and (3) connection with others. Additional subthemes surrounding the primary theme, connection with others, were identified as: (1) loneliness/isolation and (2) sharing knowledge/education. A number of respondents mentioned being able to "talk openly" and not being "so alone" with their disease as a positive impact of blogging about their illness. The words "share" and "sharing" were used often in the respondents' answers (Table 5).

When questioned whether blogging had made a difference in dealing with the challenges of chronic pain or illness, respondents indicated positive changes. Responses were grouped and the following themes identified: (1) reframing/ability to gain broader perspective on illness, (2) identifying patterns of illness, (3) providing an expressive outlet, (4) support/feedback, (5) accountability, and (6) helping to cope with illness. Respondents mentioned an ability to frame the impact of their illness differently and notice other aspects of their lives. Blogging provided an outlet of expression for some respondents, as well as a way to identify patterns of illness. Accountability was also a theme that was repeatedly mentioned, both accountability to one's self—with the implied commitment to blog—and accountability to others. Respondents reported that their overall ability to cope with chronic illness and pain was impacted positively by blogging (Table 6). Support and feedback from others who posted comments were also important factors for many respondents.

Blog-posting frequency and blog content evolved over time for many respondents. The initial use of a blog as a communication device for updating friends and family about a medical condition often evolved to helping educate others with the same disease or transitioning into a method of advocacy, mentorship, support, connection, and resolution with self and others. Identified themes regarding blog-posting frequency were (1) sense of evolving purpose (with subthemes of mentoring and advocacy), (2) making meaning of one's experience by understanding one's illness more completely, and (3) resolution, indicating an endpoint or shift in one's narrative (Table 7).

Table 3. Respondents' reasons for not sharing their blogs with their health care providers.

Themes identified	Quotes
Theme 1.1: Negative concern	
A. Fear of provider judgment of patient feelings	"I don't trust doctors to respect my feelings and opinions."
	"I would feel awkward doing so."
B. Fear of provider judgment of patient behavior	"Would be reluctant as they would probably disagree with my lifestyle choices."
C. Privacy concern/negative impact	"Concern for personal privacy."
	"I wonder if that would ever have a negative impact on my insurance rates or insurability."
D. Worry over self-editing	"I might edit myself if I knew they were reading."
E. Awkwardness in sharing negative opinions of health care provider and care	"They think they know everything and they don't and some of the blog is about them."
	"I did not share my blog with my doctor's because I sometimes wrote about frustrations with the practice in general...and felt it might have been awkward to share."
Theme 1.2: Lack of interest or time by health care provider to read patient blogs	"Doctors are not even close to interested in what I have to say, much less what I write!"
	"They've never seemed interested when I casually mentioned blogs, so I never mentioned that I write one myself."
	"No, my doctor just provides medication and doesn't discuss much with me."
Theme 1.3: Intention to share	"I plan to the next visit with her. This was one idea we came up with to overcome my fears."
	"I never specifically talked about my blog because it never came up in conversation. I would if it came up or she asked about it."
	"It hasn't really come up yet, but I suspect it will, the next time I visit."
Theme 1.4: Desire to write freely about illness experience	"I want to be able to say whatever I want even about them."
	"The content on my blog is personal and very opinionated."
	"I wanted the freedom to post about my medical experiences without worrying about what my doctor would think."

Sense of connection with others is an integral part of health and well-being for humans. Health as Expanding Consciousness theory posits that health is a dynamic quality in our evolving sense of interaction and connection with self and others, and a process of discovering meaning and understanding. Chronic illness and pain can separate an individual from their usual forms of interaction and meaning (ie, negative impact on quality of life measurements, such as work, family relationships, independence, and mobility) and new forms of interaction and meaning need to be cultivated in order to support positive psychosocial health. The primary themes of (1) helping others, (2) decreasing sense of isolation, (3) sharing of experiences, and (4) finding a sense of community and authentic voice within an individual's co-existence with a chronic illness were articulated through many of the responses to the question: "Has writing or reading patient blogs changed your sense of connection with others?" (Table 8).

Although most of the respondents did recommend blogging for someone with a chronic illness, some did express that blogging necessitated a time commitment and may not be appropriate for everyone. Of those who did recommend blogging, their reasons for recommendation were grouped into themes of (1) gaining understanding/perspective of one's chronic illness pattern, (2) emotional release, (3) decreased isolation, and (4) communication with others. A fifth theme of "would recommend with caveat" is also included (Table 9).

Privacy and disclosure, mental health issues, emotional vulnerability (discomfort/stress and anger were identified as subthemes), stage of disease (newly diagnosed), and stigma surrounding certain illnesses were the main themes of concern expressed when asked the question: "Are there circumstances when blogging should not be recommended?" The majority of those answering this question responded that there were circumstances when they would not recommend blogging to an individual (Table 10).

Table 4. Responses from health care providers upon knowing about the blog.

Themes identified	Quotes
Theme 2.1: Positive response	“Some read it, some don’t because of regulations. [They] found it very helpful for people in the same situation.”
	“Very well received.”
	“Positively, as I have great health care providers.”
Theme 2.2: Negative response	“Absolutely no interest.”
	“They don’t really like it. One clinic even forbid us to go there because of something I published in my blog.”
Theme 2.3: Neutral or indeterminate response	“I don’t know that he has ever visited, but I believe he thinks it is a good thing.”
	“Seemed to be a ‘passing’ interest in it.”

Table 5. Motivations for starting the blog.

Themes identified	Quotes
Theme 3.1: Reflection	“Mostly to sort out my thoughts.”
	“I thought it’d help me deal with the pain, by releasing it emotionally.”
	“The suicide of a friend.”
Theme 3.2: Communication	“I wanted a forum where I could talk honestly about the symptoms I was experiencing.”
	“Need for support and to streamline communications.”
	“I wanted to be able to share what was happening to me with friends and family but knew that I wouldn’t have the time/energy to communicate with everyone individually.”
Theme 3.3: Connection with others	
A. Loneliness, isolation	“Needing to find others who lived in chronic pain so I don’t feel so alone in my pain.”
	“I felt alone and isolated. And then when I started getting comments I realized I was helping other people cope.”
	“Share our reality with others.”
B. Education, knowledge sharing	“Goal is to contribute in whatever way I can to knowledge about chronic pain as a real disease with real physical consequences.”
	“Share knowledge on health advocacy.”
	“I blog to not only talk about things I am going thru but hopefully to help others as well.”

Table 6. Did blogging change how one deals with the challenges of chronic illness or pain?

Themes identified	Quotes
Theme 4.1: Reframing/ability to gain broader perspective on illness	“Mostly the writing helps me see possibilities open to me. I am more active as a patient in the direction of my own care.” “It has helped me to process information and be more objective, which is often helpful.” “It helps me get perspective on my own situation in a way I didn’t expect.”
Theme 4.2: Identifying patterns of illness	“It has made me more aware of issues I had not considered. By keeping a log, I can pick up patterns I may have missed.” “I find it easier to identify triggers and patterns.”
Theme 4.3: Expressive outlet	“It has become an online journal where I have an outlet to discuss what I am feeling, emotionally and physically and it has helped my friends and family be kept up to date on things.” “Starting my blog has given me an outlet to express myself, and since I have started my blog I feel better about myself.” “Much needed infusion of positive energy.”
Theme 4.4: Support and feedback	“It has helped me enormously—processing and rationalizing, sharing and having feedback and support.” “It made me feel more connected to people, especially when I met people on Twitter and Tumblr that I could talk to about my chronic illness.” “I have met others with my condition or a similar one, learned about both ideas for coping with and the potential progression of my conditions, and developed great networks to connect people with common conditions.” “It’s helped me realize that I’m not alone.”
Theme 4.5: Accountability	“I felt it made me more accountable to follow up on things with my doctors because I knew people would ask about them.” “I am a much more informed patient and I’m also much more comfortable talking to others about diabetes. I no longer feel so isolated.” “Writing a blog has made me more aware of my habits and also more accountable. Additionally, I have been able to look at the positive aspects of dealing with illness largely because of my blogging and the opportunities it has created for me.”
Theme 4.6: Ability to cope with illness	“Helped me cope with the anxiety.” “It makes a positive impact on my ability to deal with my care.” “It’s actually made me realize that my illness is impacting my life less than I had imagined.”

Table 7. Evolution of the objectives of blogging.

Themes identified	Quotes
Theme 5.1: Sense of evolving purpose	
A. Mentoring	“As my knowledge has increased I have become more of a mentor to others.” “Expanded the topics and added guest bloggers.” “The desire to make something good out of the struggles that I had faced and was facing at that time.”
B. Advocacy	“It started out with just my personal stories and then I started sharing research and attempted to be an advocate.” “My blog content has changed mostly as I have a desire to spread more awareness and create a message that others will better relate to and understand.” “I take more of a patient advocacy stance.”
Theme 5.2: Making meaning of the experience of illness/understanding illness	“As my networking has grown my knowledge has also grown but so has my understanding of my own disability.” “I’ve come to terms with being ill, so there’s less ‘oh goodness it’s weird and scary encountering (insert fact of disabled life).’ I’ve become more political. And there’s more stuff without a disability angle.” “I believe I’ve opened up and been more honest over time.”
Theme 5.3: Resolution	“I have taken my blog down now. Once I was over the worst of my medical issues I no longer felt it appropriate to continue it.” “Not posting as much as I’m dealing with everything better.”

Table 8. Sense of connection cultivated through blogging.

Themes identified	Quotes
Theme 6.1: Helping others	“First I was helped, now I am helping...a reminder that I *am* part of the world.”
	“Cancer survivors are an amazing group, so eager to help each other.”
	“Reading about other breast cancer survivors is inspiring, sad, and makes me want to work harder to end this epidemic.”
Theme 6.2: Decreased sense of isolation	“By writing openly and in detail and having interaction with others it has enhanced my sense of interaction.”
	“It has helped me learn to talk to others about my issues more openly.”
	“I’m not alone! There are others out there who are in the same situation as me!”
Theme 6.3: Shared experience	“It helps me dealing with my own situation knowing that there are people in more or less the same situation.”
	“I feel there is a strong sense of understanding between people with a common disease as well as between people with different chronic disorders.”
	“It’s been wonderful connecting with others who really get what it’s like to live with chronic pain.”
Theme 6.4: Sense of community	“Enabled me to connect with others worldwide I would never have normally met. Enabled me to become closer to friends.”
	“There is a sense of ‘we’re all in this together’ with other bloggers who are faced with similar challenges, even if we face different illnesses; we often have a lot in common. It can be a great comfort.”
	“Realize that others are going through the same things as I am.”

Table 9. Reasons of recommending blogging.

Themes identified	Quotes
Theme 7.1: Gaining understanding/perspective	“Great way to work things out. Forces you to think about what the impact of the illness is (or is not!) on your life.”
	“It takes stress away and avoids an endless story repetition to those interested.”
Theme 7.2: Emotional release	“It’s a good release of emotions and you can set the level of privacy on each post so you can choose who sees what!”
	“It helps to write down your feelings and what you are going through.”
	“I think, if nothing else, it is a good outlet and a healthy way to share your feelings.”
Theme 7.3: Decreased isolation	“Certainly to people who enjoy writing, though, and to people who are socially isolated, although never blog *just* about illness.”
	“Yes, eases sense of isolation helps in staying on top of treatment changes, enables conversations that the healthy do not want to have.”
Theme 7.4: Communication with self or others	“It can help to connect with others in a similar situation and also help you understand yourself and your illness more.”
	“Healthy communication of pain is never wrong.”
	“I think it can have an impact on helping others to ‘get it.’”
Theme 7.5: Would recommend blogging with caveat	“If they are comfortable doing it. Privacy concerns are the 800-lb gorilla in the room.”
	“Only if they want to, because it can be really demanding. But I would recommend it because it will help them to become more empowered, informed and confident.”
	“Maybe a private blog? Or a blog within an online fibromyalgia community to feel safer.”

Table 10. Perceived situations in which blogging would not be recommended.

Themes identified	Quotes
Theme 8.1: Privacy/disclosure	"I would not recommend blogging for those who do not feel comfortable sharing personal information."
	"Those with privacy concerns, such as health insurance tied to employment, need to be cautious. Those who would be harmed by public disclosure or being found by search engines."
	"Anything published on the Internet stays there forever."
Theme 8.2: Stage of disease (newly diagnosed)	"Maybe when still new to the diagnosis and searching for answers."
Theme 8.3: Mental Illness	"If you are very unwell (mentally) and might say something that could 'trigger' someone else to do something harmful toward themselves or others, you should either not write until you are a bit better or censor what you write."
	"Anyone with a mental illness who cannot handle negative comments should not blog."
Theme 8.4: Stigma of illness	"When people are uncomfortable talking about their conditions. There are also regional issues, stigmas to think about. If you don't like being found with a Google search of the condition and your name, don't blog."
Theme 8.5: Emotional vulnerability	
A. Discomfort/stress	"If it would make the person feel uncomfortable sharing medical information or if it would stress them out too much."
	"If it hurts someone else."
	"When patient is too weak or under strong emotional distress."
B. Anger	"Don't blog when angry. You'll say things you'll regret."
	"Express anger aimed at another individual."

Discussion

This project contributes to formative research on illness blogging and provides information leading to a more thorough understanding of why individuals initiate and maintain an illness blog. This convenience sample of primarily well-educated women mirrors the population who are most engaged with social media; however, this is not fully representative of the overall online population with chronic disease, based on the Pew data [18]. The Pew Internet and American Life Project has found in their surveys of Internet and social networking usage in the United States that an adult with chronic illness is less likely to have Internet access compared to an adult without chronic illness (62% versus 81%) [18]. Approximately the same proportion of patients with and without chronic conditions either consume (57%) or generate (20%) online health information.

Overall, respondents provided information that illustrates how blogging assisted them in sharing, responding, reflecting, and evolving in their understanding and experience of their own health conditions and illness challenges. This evidence demonstrates how blogging may assist some patients in articulating and sharing their illness narratives. Although health care providers only suggested blogging to 6% of respondents, the positive elements identified by bloggers suggests therapeutic potential in this social media tool. Further research is needed to develop the exploratory data reported here and to explore patient characteristics, diagnoses, or conditions that might find such a process useful. Similar to evidence that journaling provides moderation of pain and increases self-management of chronic pain, it is possible that with further research, providers may be able to better identify safe and effective conditions for recommendation of illness blogging as well as understand methods to minimize negative effects of illness blogging. Continued research may find that development of blogging as a health care tool could be a beneficial intervention for some patients with chronic pain or illness.

Respondents provided longer, more reflective responses to the open-ended survey questions than expected, identifying perceived motivators and psychosocial benefits of blogging about one's illness. Most bloggers started their blog on their own initiative, and this may suggest that this cohort already had strong illness coping strategies in place. A future research question to assess is whether introducing illness blogging by a health care provider would have a similar positive effect. Illness bloggers readily shared their blogs with friends and family, and the majority of the blogs were not password or otherwise protected from public view or search capabilities. In contrast to this open and public nature of illness blogging, a majority of respondents did not share their blogs with their health care providers. The perception that health care providers would be uninterested or lacked time, and had been hostile or were anticipated to be hostile toward the patient and/or the contents of the illness blog was expressed in a number of survey responses. Some respondents felt if their health care provider read or commented on their blog, it would stifle the free flow of emotions and ideas that blogging provides. This may be an issue to investigate further with health care providers as well as illness bloggers. The theme of accountability expressed by the bloggers was unanticipated, yet it is consistent with other responses of developing greater connections with others and deeper insight into one's self through the reflective process of blogging. The accountability expressed appeared to be a motivating factor resulting in a sense of responsibility, purpose, and attempting to validate and gain understanding of the experience of illness by the process of initiating and maintaining a blog.

The creation of online connections with others by commenting and reading others' illness blogs supports the concepts framed by the theory of Health as Expanding Consciousness, which posits that individuals strive to regain a sense of health by connecting with self and others, and finding growth, meaning, and purpose in life experiences. The process of blogging allows for the creation of real-time sharing of experience with family and a community of others in similar situations; in contrast, traditional journal writing does not permit this same process. Several survey respondents did remark on the sense of being "less alone" and less isolated by their illness when engaging in the process of illness blogging.

Limitations of the study included a convenience sample resulting in a narrow demographic cohort of self-selected, highly educated, English speakers (the survey was not translated into languages other than English), who were predominantly women. This may have been due to active participation in the survey by online communities focused on breast cancer and fibromyalgia. A sample more closely representing the population with chronic illness and pain will need to be developed to further explore blogging issues, while accounting for evidence that education and income levels of bloggers tend to be higher than the general population [26]. Although women with chronic disease are proportionally more active in generating Internet content, such as blogs, further studies may want to include bloggers with more gender-neutral conditions, such as lung or colon cancer, heart disease, and diabetes. The matrix model of our survey design may have been a specific limitation because the question regarding the impact of their illness limited respondents' ability to express if an item had both negative and positive impacts on their disease or illness which may have limited the ability of respondents to fully answer some questions. Financial aspects of illness were not addressed in this project; these might impact patients' experiences and should be included in future studies.

Although there are limitations in this formative research, we believe this study adds to the growing body of knowledge on digital communication (blogging) and how it may play a role in addressing the needs of individuals living with chronic pain or illness. The project hypothesis, that communicating the experience of illness through blogging may provide positive psychosocial benefits to some patients with chronic pain or illness, appears to be supported. Larger studies that incorporate a comparison group in a longitudinal setting need to be undertaken to generalize the benefits to a wider population.

Given the data from respondents, further research in this area is indicated in several directions. There is limited literature examining the intersection between illness narratives and blogs. Based on this data, blogging appears to decrease a sense of isolation through the establishment of online connections with others and increases a sense of purpose by assisting others facing similar situations. Although blogging consumes time and energy for the blogger, illness blogging seems to facilitate a shared reflective experience through the process of reading, writing, and commenting on illness blogs. Both Charon [4] and Pennebaker [8] describe the benefits of telling and retelling of a narrative to come to greater understanding and meaning of the event. The experience of blogging enables the blogger to narrate their own health story in real time with evolving understanding and meaning. Qualitative analysis will assist in understanding how blogging further relates to expressive writing and narrative. Health as Expanding Consciousness theory suggests that humans are seeking a sense of meaning and purpose out of disorder and hopelessness [20]. These benefits may be harder to obtain in traditional journaling or illness narratives because of the lack of real-time communication and the broad reach offered by the use of the Internet.

It is noteworthy that illness bloggers perceive a lack of interest by the health care community in their blogs. This may be an area for further research within the health care community. As we have moved forward with medical advances supporting the ability to live longer, the numbers of people living with chronic disease and pain have increased. Improving the quality of life for people with chronic disease and pain requires greater recognition of not only the physical manifestations of pain and disease, but also their psychosocial manifestations. This new paradigm of health care includes living with illness for extended lengths of time and a need for the individual and family to create new ways of understanding and processing the dynamic health-illness continuum. Charon [16] states that it is essential to find ways to bridge the divide between illness and health, to reflect and seek understanding of what has happened, and often this can be facilitated by storytelling and narratives. The Health as Expanding Consciousness theory suggests a need for finding patterns and meaning, and a connection with one's self and others are necessary tasks to complete in moving toward health. The words of a survey respondent, "First I was helped, now I am helping...a reminder that I am part of the world," describe the experience process that bloggers in this project reported. Perhaps the process of communicating the experience of chronic illness and pain through blogging may be one method to assist in moving toward a more complete, holistic model of health and healing by allowing individuals with chronic illness and pain to regain a place in the world.

Acknowledgments

We thank Gary Klatsky, PhD, as the source of inspiration behind the initial study that led to this research. Gary, who is now deceased, wrote a blog about his brain tumor and spoke to one of the authors (LG) about the reasons he started his blog.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Communicating the Experience of Chronic Pain and Illness through Blogging Survey.

[PDF File (Adobe PDF File), 193KB - jmir_v14i5e143_app1.pdf]

References

1. The American Academy of Pain Medicine. AAPM Facts and Figures on Pain, The Burden of Pain on Every Day Life URL: http://www.painmed.org/patient/facts.html#burden [WebCite Cache ID 63DcQKtMw]
2. Centers for Disease Control and Prevention. Chronic Diseases and Health Promotion URL: http://www.cdc.gov/chronicdisease/overview/index.htm [accessed 2011-11-15] [WebCite Cache ID 63DaRar6S]
3. Carr D, Loeser J, Morris D. Narrative, Pain And Suffering: Progress in Pain Research and Management. Seattle, WA: IASP Press; 2005.
4. Charon R. The patient-physician relationship. Narrative medicine: a model for empathy, reflection, profession, and trust. JAMA 2001 Oct 17;286(15):1897-1902. [Medline: 11597295]
5. Kleinman A. The Illness Narratives: Suffering, Healing, and the Human Condition. New York: Basic Books; 1988.
6. Scarry E. The Body in Pain: The Making and Unmaking of the World. New York: Oxford University Press; 1985.
7. Frank A. The Wounded Storyteller: Body, Illness, and Ethics. Chicago: University of Chicago Press; 1995.
8. Pennebaker JW. Telling stories: the health benefits of narrative. Lit Med 2000;19(1):3-18. [Medline: 10824309]
9. Baikie KA, Wilhelm K. Emotional and physical health benefits of expressive writing. Advances in Psychiatric Treatment 2005;11(5):338-346. [doi: 10.1192/apt.11.5.338]
10. Lepore S, Smyth J. The Writing Cure: How Expressive Writing Promotes Health and Emotional Well-Being. Washington, DC: American Psychological Association; 2002.
11. Pennebaker J. Writing about emotional experiences as a therapeutic process. Psychological Science 1997;8(3):162-166. [doi: 10.1111/j.1467-9280.1997.tb00403.x]
12. Pennebaker JW, Francis ME. Cognitive, emotional, and language processes in disclosure. Cognition & Emotion 1996;10(6):601-626. [doi: 10.1080/026999396380079]
13. Pennebaker JW, Kiecolt-Glaser JK, Glaser R. Disclosure of traumas and immune function: health implications for psychotherapy. J Consult Clin Psychol 1988 Apr;56(2):239-245. [Medline: 3372832]
14. Sheese BE, Brown EL, Graziano WG. Emotional expression in cyberspace: searching for moderators of the Pennebaker disclosure effect via e-mail. Health Psychol 2004 Sep;23(5):457-464. [doi: 10.1037/0278-6133.23.5.457] [Medline: 15367065]
15. Sloan DM, Marx BP. A closer examination of the structured written disclosure procedure. J Consult Clin Psychol 2004 Apr;72(2):165-175. [doi: 10.1037/0022-006X.72.2.165] [Medline: 15065952]
16. Charon R. Narrative Medicine: Honoring the Stories of Illness. Oxford: Oxford University Press; 2006.
17. Fox S. Commentary on Pew Internet: health. Washington, DC: Pew Internet & American Life Project 2012 Mar 01 webcite id: 66jCz4jad [FREE Full text]
18. Fox S, Purcell K. Chronic disease and the Internet. Washington, DC: Pew Internet & American Life Project 2010 Mar 04 [FREE Full text]
19. Heilferty CM. Toward a theory of online communication in illness: concept analysis of illness blogs. J Adv Nurs 2009 Jul;65(7):1539-1547. [doi: 10.1111/j.1365-2648.2009.04996.x] [Medline: 19457009]
20. Newman M. Health as Expanding Consciousness. URL: http://healthasexpandingconsciousness.org/home/index.php?option=com_frontpage&Itemid=1 [WebCite Cache ID 63DZBEI4V]
21. Newman M. A Developing Discipline: Selected Works of Margaret Newman. New York: National League for Nursing Press; 1995.
22. SurveyMonkey. Homepage URL: http://www.surveymonkey.com/ [accessed 2011-11-15] [WebCite Cache ID 63Dc7IVfD]
23. WEGO Health. Homepage URL: http://www.wegohealth.com/ [accessed 2011-11-15] [WebCite Cache ID 63DYDzIdn]
24. Charmaz K. Constructing Grounded Theory: A Practical Guide through Qualitative Analysis. London: SAGE; 2006.
25. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
26. Sussman M. Technorati. 2009 Oct 19. Day 1: Who Are The Bloggers? SOTB 2009 URL: http://technorati.com/social-media/article/day-1-who-are-the-bloggers1 [accessed 2012-04-06] [WebCite Cache ID 66jD2qqAw]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
IP: Internet protocol

IRB: Institutional Review Board
SSL: secure sockets layer

Edited by G Eysenbach; submitted 17.11.11; peer-reviewed by N Zhang; comments to author 11.01.12; revised version received 28.06.12; accepted 11.07.12; published 23.10.12.

Please cite as:
Ressler PK, Bradshaw YS, Gualtieri L, Chui KKH
Communicating the Experience of Chronic Pain and Illness Through Blogging
J Med Internet Res 2012;14(5):e143
URL: http://www.jmir.org/2012/5/e143/
doi:10.2196/jmir.2002
PMID:23092747

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org