Journal of Medical Internet Research

Journal Impact Factor (JIF) (2022): 7.4
Volume 16 (2014), Issue 4 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

A Smart Web Aid for Preventing Diabetes in Rural China: Preliminary Findings and Lessons (e98)
Penglai Chen, Jing Chai, Jing Cheng, Kaichun Li, Shaoyu Xie, Han Liang, Xingrong Shen, Rui Feng, Debin Wang. 3

Adolescents Just Do Not Know What They Want: A Qualitative Study to Describe Obese Adolescents' Experiences of Text Messaging to Support Behavior Change Maintenance Post Intervention (e103)
Kyla Smith, Deborah Kerr, Ashley Fenner, Leon Straker. 16

Factors Related to Sustained Use of a Free Mobile App for Dietary Self-Monitoring With Photography and Peer Feedback: Retrospective Cohort Study (e109)
Elina Helander, Kirsikka Kaipainen, Ilkka Korhonen, Brian Wansink. 48

Can an Internet-Based Health Risk Assessment Highlight Problems of Heart Disease Risk Factor Awareness? A Cross-Sectional Analysis (e106)
Justin Dickerson, Catherine McNeal, Ginger Tsai, Cathleen Rivera, Matthew Smith, Robert Ohsfeldt, Marcia Ory. 61

Usage and Effectiveness of a Fully Automated, Open-Access, Spanish Web-Based Smoking Cessation Program: Randomized Controlled Trial (e111)
Guillermo Mañanes, Miguel Vallejo. 83

Adherence to Self-Monitoring via Interactive Voice Response Technology in an eHealth Intervention Targeting Weight Gain Prevention Among Black Women: Randomized Controlled Trial (e114)
Dori Steinberg, Erica Levine, Ilana Lane, Sandy Askew, Perry Foley, Elaine Puleo, Gary Bennett. 101

A Spanish Pillbox App for Elderly Patients Taking Multiple Medications: Randomized Controlled Trial (e99)
José Mira, Isabel Navarro, Federico Botella, Fernando Borrás, Roberto Nuño-Solinís, Domingo Orozco, Fuencisla Iglesias-Alonso, Pastora Pérez-Pérez, Susana Lorenzo, Nuria Toro. 113

Would You Tell Everyone This? Facebook Conversations as Health Promotion Interventions (e108)
Jonathan Syred, Carla Naidoo, Sarah Woodhall, Paula Baraitser. 132

Mapping Physician Twitter Networks: Describing How They Work as a First Step in Understanding Connectivity, Information Flow, and Message Diffusion (e107)
Ranit Mishori, Lisa Singh, Brendan Levy, Calvin Newport. 141

Still a Hard-to-Reach Population? Using Social Media to Recruit Latino Gay Couples for an HIV Intervention Adaptation Study (e113)
Omar Martinez, Elwin Wu, Andrew Shultz, Jonathan Capote, Javier López Rios, Theo Sandfort, Justin Manusov, Hugo Ovejero, Alex Carballo-Dieguez, Silvia Chavez Baray, Eva Moya, Jonathan López Matos, Juan DelaCruz, Robert Remien, Scott Rhodes. 152

Real-Time Sharing and Expression of Migraine Headache Suffering on Twitter: A Cross-Sectional Infodemiology Study (e96)
Thiago Nascimento, Marcos DosSantos, Theodora Danciu, Misty DeBoer, Hendrik van Holsbeeck, Sarah Lucas, Christine Aiello, Leen Khatib, MaryCatherine Bender, UMSoD (Under)Graduate Class Of 2014, Jon-Kar Zubieta, Alexandre DaSilva. 166

Performance of eHealth Data Sources in Local Influenza Surveillance: A 5-Year Open Cohort Study (e116)
Toomas Timpka, Armin Spreco, Örjan Dahlström, Olle Eriksson, Elin Gursky, Joakim Ekberg, Eva Blomqvist, Magnus Strömgren, David Karlsson, Henrik Eriksson, James Nyce, Jorma Hinkula, Einar Holm. 177

Electronic Versus Paper-Based Assessment of Health-Related Quality of Life Specific to HIV Disease: Reliability Study of the PROQOL-HIV Questionnaire (e115)
Martin Duracinsky, Christophe Lalanne, Cécile Goujard, Susan Herrmann, Christian Cheung-Lung, Jean-Paul Brosseau, Yannick Schwartz, Olivier Chassany. 210

User-Generated Online Health Content: A Survey of Internet Users in the United Kingdom (e118)
Braden O'Neill, Sue Ziebland, Jose Valderas, Francisco Lupiáñez-Villanueva. 222

The Impact of Search Engine Selection and Sorting Criteria on Vaccination Beliefs and Attitudes: Two Experiments Manipulating Google Output (e100)
Ahmed Allam, Peter Schulz, Kent Nakamoto. 233

Executive Functioning in Alcoholics Following an mHealth Cognitive Stimulation Program: Randomized Controlled Trial (e102)
Pedro Gamito, Jorge Oliveira, Paulo Lopes, Rodrigo Brito, Diogo Morais, Diana Silva, Ana Silva, Sara Rebelo, Marta Bastos, Alberto Deus. . .
2 5 9

Web-Based Virtual Patients in Nursing Education: Development and Validation of Theory-Anchored Design and Activity Models (e105)
Carina Georg, Nabil Zary. 272

Reviews

Mobile Applications for Diabetics: A Systematic Review and Expert-Based Usability Evaluation Considering the Special Requirements of Diabetes Patients Age 50 Years or Older (e104)
Madlen Arnhold, Mandy Quade, Wilhelm Kirch. 28

Effectiveness and Cost-Effectiveness of eHealth Interventions in Somatic Diseases: A Systematic Review of Systematic Reviews and Meta-Analyses (e110)
Niels Elbert, Harmieke van Os-Medendorp, Wilco van Renselaar, Anne Ekeland, Leona Hakkaart-van Roijen, Hein Raat, Tamar Nijsten, Suzanne Pasmans. 187

Viewpoints

Collaborative Biomedicine in the Age of Big Data: The Case of Cancer (e101)
Abdul Shaikh, Atul Butte, Sheri Schully, William Dalton, Muin Khoury, Bradford Hesse. 127

Is Biblioleaks Inevitable? (e112)
Adam Dunn, Enrico Coiera, Kenneth Mandl. 253

Original Paper

A Smart Web Aid for Preventing Diabetes in Rural China: Preliminary Findings and Lessons

Penglai Chen[1*], MD; Jing Chai[1*], MD; Jing Cheng[1], MD; Kaichun Li[2], MD; Shaoyu Xie[2], MD; Han Liang[1], MD; Xingrong Shen[1], MD; Rui Feng[1], MD; Debin Wang[1], MD, PhD

[1]School of Health Services Management, Anhui Medical University, Hefei, China
[2]Lu’an Center for Diseases Prevention and Control, Lu'an, China
[*]these authors contributed equally

Corresponding Author:
Debin Wang, MD, PhD
School of Health Services Management
Anhui Medical University
81 Meishan Road
Hefei, 230032
China
Phone: 86 55165116395
Fax: 86 55165116395
Email: dbwang@vip.sina.com

Abstract

Background: Increasing cases of diabetes, a general lack of routinely operational prevention, and a long history of separating disease prevention and treatment call for immediate engagement of frontier clinicians. This applies especially to village doctors who work in rural China where the majority of the nation’s vast population lives.

Objective: This study aims to develop and test an online Smart Web Aid for Preventing Type 2 Diabetes (SWAP-DM2) capable of addressing major barriers to applying proven interventions and integrating diabetes prevention into routine medical care.

Methods: Development of SWAP-DM2 used evolutionary prototyping. The design of the initial system was followed by refinement cycles featuring dynamic interaction between development of practical and effective standardized operation procedures (SOPs) for diabetes prevention and Web-based assistance for implementing the SOPs. The resulting SOPs incorporated proven diabetes prevention practices in a synergetic way. SWAP-DM2 provided support to village doctors ranging from simple educational webpages and record maintenance to relatively sophisticated risk scoring and personalized counseling. Evaluation of SWAP-DM2 used data collected at baseline and 6-month follow-up assessment: (1) audio recordings of service encounters; (2) structured exit surveys of patients’ knowledge, self-efficacy, and satisfaction; (3) measurement of fasting glucose, body mass index, and blood pressure; and (4) qualitative interviews with doctors and patients. Data analysis included (1) descriptive statistics of patients who received SWAP-DM2–assisted prevention and those newly diagnosed with prediabetes and diabetes; (2) comparison of the variables assessed between baseline and follow-up assessment; and (3) narratives of qualitative data.

Results: The 17 participating village doctors identified 2219 patients with elevated diabetes risk. Of these, 84.85% (1885/2219) consented to a fasting glucose test with 1022 new prediabetes and 113 new diabetes diagnoses made within 6 months. The prediabetic patients showed substantial improvement from baseline to 6-month follow-up in vegetable intake (17.0%, 43/253 vs 88.7%, 205/231), calorie intake (1.6%, 4/253 vs 71.4%, 165/231), leisure-time exercises (6.3%, 16/253 vs 21.2%, 49/231), body weight (mean 62.12 kg, SD 9.85 vs mean 58.33 kg, SD 9.18), and body mass index (mean 24.80 kg/m^2, SD 3.21 vs mean 23.36 kg/m^2, SD 2.95). The prediabetic patients showed improvement in self-efficacy for modifying diet (mean 5.31, SD 2.81 vs mean 8.53, SD 2.25), increasing physical activities (mean 4.52, SD 3.35 vs mean 8.06, SD 2.38), engaging relatives (mean 3.93, SD 3.54 vs mean 6.93, SD 2.67), and knowledge about diabetes and risks of an imbalanced diet and inadequate physical activity. Most participating doctors and patients viewed SWAP-DM2 as useful and effective.

Conclusions: SWAP-DM2 is helpful to village doctors, acceptable to patients, and effective in modifying immediate determinants of diabetes at least in the short term, and may provide a useful solution to the general lack of participation in diabetes prevention by frontier clinicians in rural China.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN): 66772711; http://www.controlled-trials.com/ISRCTN66772711 (Archived by WebCite at http://www.webcitation.org/6OMkAqyEy).

(J Med Internet Res 2014;16(4):e98) doi:10.2196/jmir.3228

KEYWORDS

diabetes mellitus; prediabetic state; Internet; prevention; evaluation; eHealth

Introduction

The rates of type 2 diabetes mellitus (DM2) are increasing and it is expanding rapidly from primarily affecting people in developed nations to also afflicting people in the developing world [1-3]. It is predicted that diabetes will claim up to 366 million people worldwide by 2030 [4]. This epidemic is also growing in China [5-7]. A recent nationwide investigation revealed that age-standardized prevalence of total diabetes in China is as high as 11.4% in urban areas and 8.2% in rural areas [8-10]. Diabetes interacts with other major risk factors (eg, hypertension, dyslipidemia) and increases the risk of a variety of morbidities (eg, chronic kidney disease, end-stage renal disease, atherosclerosis, coronary heart disease, and cerebral ischemia) leading to tremendous physical, psychological, and socioeconomic suffering and burden [11-13]. Diabetes develops from prediabetes, a lesser degree of hyperglycemia [14,15], and it is well established that the risk of this progression can be modified substantially (by 20% to 60%) regardless of nation and population group [16-20]. Encouraged by high efficacies, research efforts have been invested in this regard and a whole range of information, education, and communication are available, including guidelines, protocols, tool kits, and best practices [21,22].

Unfortunately, proven interventions against diabetes are not being practiced routinely in China. This is especially true in the resource-poor rural areas where more than 75% of the vast population lives. Previously published studies and our own preliminary investigations all suggest that rural village doctors seldom participate in identifying and preventing diabetes [23,24]. As a result, 42% to 82% of rural villagers newly screened with diabetes had never been diagnosed with the disease before and knowledge about prediabetes or diabetes among rural villagers is extremely low. The primary reason for this gap between proven technologies and application may relate to a general lack of necessary knowledge and skills among village doctors in rural China. Less than 12% of village doctors have received formal training on diabetes, and only 43% knew basic knowledge about diabetes prevention [25,26]. Another barrier preventing routinely implementing proven interventions may be lack of incentives [27,28]. Persistent promotion of lifestyle modification, the key to diabetes prevention, requires sustained momentum. This calls for continuous monitoring, supervising, and rewarding. However, all these are generally missing with the current health care systems in China, especially those in resource-poor areas.

In response to these challenges and others, we started a quasi-cluster randomized controlled trial (ISRCTN66772711) in Lu’an, Anhui, China. It aimed at devising, implementing, and evaluating an intervention package to reduce progression from prediabetes to diabetes among male and female farmers aged 40 and older and establishing a sustainable mechanism for integrating diabetes prevention within routine medical service. The trial utilized a batched implementation strategy in which villages were recruited in blocks to practice planned intervention with former blocks informing later ones. For each block, measurement occured at baseline and every 12 months (for plasma glucose) or monthly (for knowledge, behavior, body weight, and blood pressure) after baseline. The intervention package is known as educating doctors and electronic supports, counseling diabetes prevention, recipe for lifestyle modification, operational toolkit, performance-based reimbursement for doctors, and screening service (eCROPS). The overall trial protocol has been published elsewhere [29]. This paper documents the “e” component of the intervention package. The Smart Web Aid for Preventing Type 2 Diabetes (SWAP-DM2) was designed to tackle existing barriers to applying proven interventions and integrating diabetes prevention with routine medical care by using evidence-based measures in an innovative and synergetic way.

Methods

Development of Initial SWAP-DM2

The SWAP-DM2 inherited the strategies and theoretical frameworks we used successfully in producing a practical computerized expert system for routine human immunodeficiency virus (HIV) voluntary counseling and testing that simplified the counseling process and leveraged essential procedures and best practices [30]. It utilized Microsoft Visual Studio 2008 as the platform, SQL Server 2008 as the data management tool, C# as the programming language, and evolutionary prototyping as the overall development approach in which design of the initial type system was followed by continuous refinement cycles. The whole process featured dynamic interaction between development of practical and effective standardized operation procedures (SOPs) and Web-based assistance for implementing these SOPs.

The initial SOPs behind SWAP-DM2 were derived through evidence- and theory-based consensus. First, our research team on diabetes prevention conducted a systematic literature review of related guidelines, protocols, theories, and research articles and then worked out a long list of proven intervention approaches and models. Second, an expert panel consisting of experienced counselors, psychologists, clinicians, and nurses who work with patients with diabetes, as well as experts on nutrition and physical activities and epidemiologists conveyed for a consensus meeting and produced a short list of proven interventions from the long list through clarification, brainstorming, and rounds of voting. Third, our technical groups on diabetes prevention and Web program development worked together and translated the short list of intervention approaches and models into primitive SOPs. Fourth, the expert panel gathered again and revised the primitive SOPs into initial practical ones. Motivational interviewing played a guiding role in this process because the core part of the SOP development concerned counseling lifestyle management and the theory has been applied successfully for promoting behavior changes in various population groups (Multimedia Appendix 1) [31].

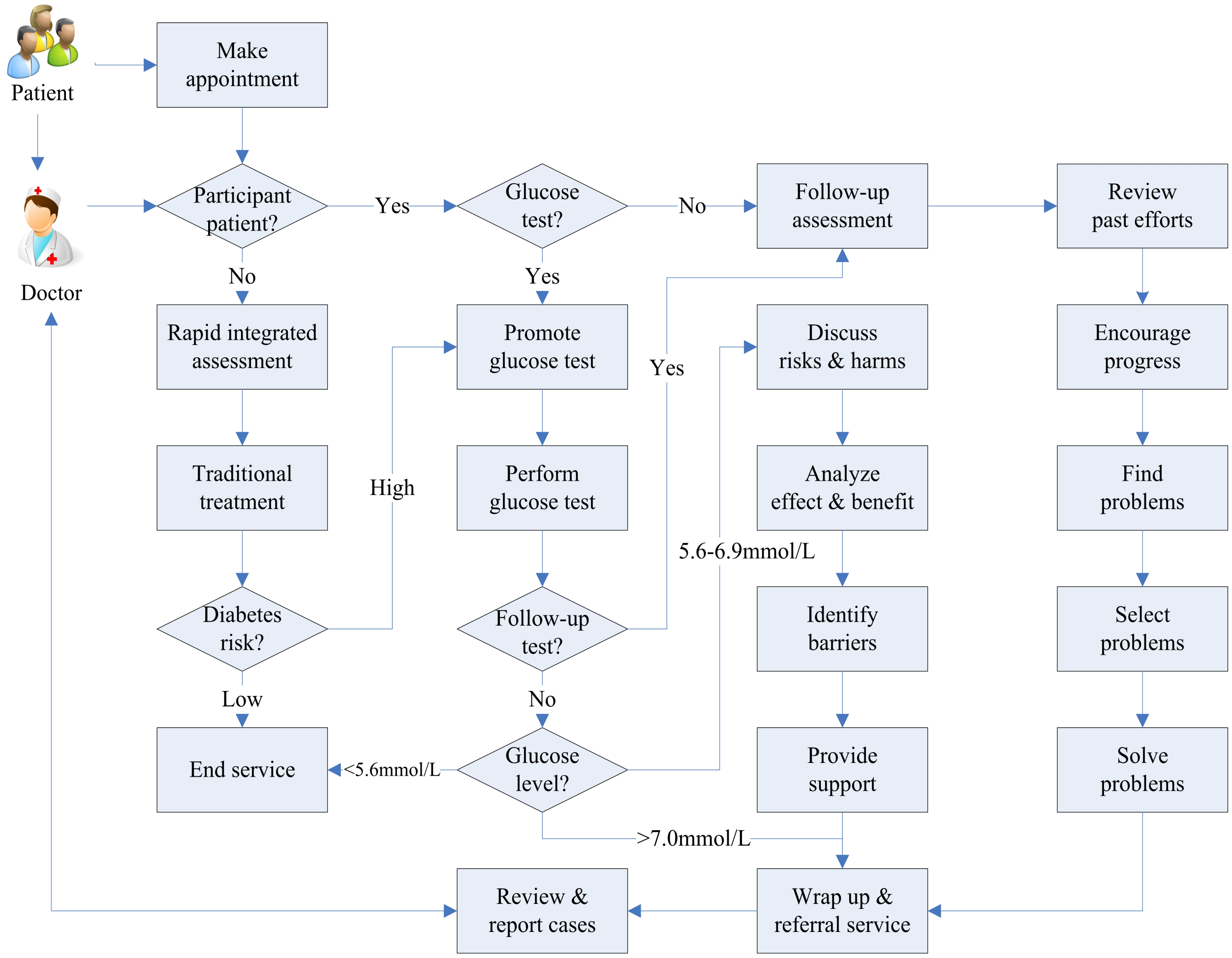
The prototype SWAP-DM2 built upon the previously mentioned SOPs through close collaboration between software programmers and diabetes prevention researchers within our team. First, our diabetes prevention researchers clarified on an item-by-item basis the SOPs to our Web programmers and addressed all enquiries raised by the latter. Following this, the same 2 groups continued working together and performed an individual needs analysis in which each item of the SOPs underwent a semistructured process for identifying potential Web-based electronic supports for implementing the SOP by using nominal group techniques [32,33]. Then, the programmers grouped and combined all the needs identified into pragmatic applications and program codes.

SWAP-DM2: Mediated Diabetes Prevention

As summarized in Multimedia Appendix 2, the SWAP-DM2 provides a package of support for village doctors in delivering diabetes prevention ranging from simple educational webpages and record maintenance to relatively sophisticated risk scoring and personalized counseling. The essence of the Web is that it leverages the use of SOPs and best diabetes prevention in a user-friendly way. The users (village doctors in this case) need to mainly follow the steps proposed and make minimum responses by either clicking the mouse or pressing an access key depending on the users' preferences. Figure 1 depicts the general process of SWAP-DM2–mediated prevention and Multimedia Appendix 3 provides sample application webpages.

For a given patient presenting to a clinic, SWAP-DM2 automatically classifies (after inputting a unique identification number) the patient as participant or nonparticipant of the diabetes prevention project and then proposes standard service for each kind of patient accordingly. If the patient was a nonparticipant, the system provided SOPs for performing an integrated rapid assessment (ie, assessing the patient's risk for developing diabetes by using a very short instrument while the patient received traditional medical service), which in turn enabled the system to automatically assign the patient to either a high- or low-risk nonparticipant patient group. For a high-risk nonparticipant patient, SWAP-DM2 lead to standard procedures for promoting and performing a glucose test, which then allowed for further classification of the patient as suspected diabetes (fasting glucose ≥7.0 mmol/L), prediabetes (fasting glucose 5.6-6.9 mmol/L), or normal (fasting glucose <5.6 mmol/L). For a normal patient, SWAP-DM2 tells the doctor to end the service for the patient. For suspected diabetes, the system suggests an SOP referring the patient to higher-level diabetes diagnosis and treatment services. As for prediabetes, SWAP-DM2 proposes a 4-step SOP for promoting the patient to participate in consecutive cycles of lifestyle management. These 4 steps are (1) discussing risks to the patient and the harms diabetes causes, (2) analyzing effectiveness and benefits of lifestyle management, (3) identifying barriers to lifestyle management, and (4) providing supports for overcoming the barriers. Each time a participant patient presented, SWAP-DM2 mediated a tailored round of lifestyle management. Each round of lifestyle management comprised 6 core steps (ie, perform follow-up assessment, review previous efforts, encourage progress, identify problems, select problems, and solve problems) for all participants and 2 additional steps (ie, promote and perform glucose test) for selective patients. *Tailored* here means that detailed content of each round of lifestyle management differed from others and depended on the patient's previous involvement in and performance on lifestyle management.

Figure 1. Flowchart of the Smart Web Aid for preventing type 2 diabetes (SWAP-DM2).



SWAP-DM2 Assessment Process and Measures

The initial SWAP-DM2 described previously has been refined through repeated cycles of testing and assessment. The pilot work started in early June 2013 and the assessment occurred at baseline (in the week before training of participating village doctors) and every month after SWAP-DM2 application (in the last week of the month). Village doctors from randomly selected clinics in rural Lu'an (one of the largest prefectures in Anhui province) were invited to participate in testing the system.

Assessment is an integral part of the refinement of SWAP-DM2. This study did not collect any additional information; rather, it utilized data generated during the baseline and the 6-month follow-up assessment of SWAP-DM2 application scheduled during the last week of the month. These assessments consisted of (1) prevention case recording, in which all the patient-doctor encounter episodes happened at the participating clinics during the selected week were, after informed consent, audio recorded; (2) exit survey using a structured questionnaire (ie, after completion of each service during the assessment week), the eligible patient (aged 40 to 70 years and tested as prediabetics) was led to a private room and interviewed by a trained interviewer using a structured questionnaire to solicit information about his/her knowledge about diabetes and self-efficacy (self-perceived ability) to practice indicative lifestyle modification against diabetes and satisfaction toward the service just received (Multimedia Appendix 4); (3) measurements including body height and weight, fasting glucose (baseline and annual follow-up evaluations), and blood pressure; and (4) qualitative interview (follow-up evaluation only) in which all participating village doctors and 1-2 patients from each doctor's clinic were interviewed by trained researchers using a checklist of open-ended questions about SWAP-DM2 (eg, its effects on diabetes intervention, areas to improve, barriers to utilization). The exit survey questionnaire covered all key aspects, according to the motivational interviewing theory mentioned previously, of diabetes-related lifestyle management (see Multimedia Appendix 3). It tested moderately reliable with a standardized Cronbach alpha of .74, .85, .70, .77, and .98 for knowledge about diabetes and prediabetes, harms of inadequate physical activity and imbalanced diet, self-efficacy, diabetes-related behaviors, and satisfaction toward Web aid-supported service, respectively.

Data was collected using both quantitative and qualitative methods. The audio-recorded cases of service delivery were used to generate quantitative service quality data through (1) translating the audio recordings into textual transcripts; (2) mixing the baseline case transcripts with the follow-up ones with all identification information being removed except for a unique number to ensure blind assessment; (3) sending the mixed transcripts to 2 researchers and asking them check each of the cases independently against a list of essential prevention procedures and judge whether these essential prevention procedures had been delivered; and (4) having the 2 researchers sit together and reach consensus over their disagreements.

Data Analysis

Data analysis used SPSS 16.0 (SPSS Inc, Chicago, IL, USA) and included (1) descriptive numbers and percentages of patients who received SWAP-DM2–assisted prevention and newly diagnosed prediabetes and diabetes; and (2) comparison, using chi-square tests for percentages or *t* tests for quantitative ratings, between the baseline and 6-month follow-up assessments in terms of improvements in delivery of essential prevention procedures, knowledge about diabetes, lifestyle management practices, body mass index (BMI), and blood pressure, etc; and (3) narratives of qualitative data.

Human Subject Protection

This study involves recruitment, intervention, and assessment of villagers and village doctors. So it adheres to rigorous human subject protection principles and procedures. The study protocol had been reviewed and approved by the Biomedical Ethics Committee of Anhui Medical University. Participation of villagers and village doctors was 100% voluntary. Written informed consent was sought from all participants.

Results

Patient and Doctor Participants

A total of 17 village doctors participated in pilot-testing SWAP-DM2 and they identified 2219 patients at risk for being diagnosed with prediabetes or diabetes using the rapid assessment function of the system during this preliminary evaluation period. Of these 2219 patients, 1885 (85.94%) received a free fasting glucose test; 1022 (54.22%) and 103 (5.46%), were diagnosed with prediabetes and diabetes, respectively. Most patients diagnosed with prediabetes (95.3%, 974/1022) did not know their glucose status before (Table 1).

The total number of patients encountered and identified as having prediabetes during baseline and 6-month follow-up assessment were 253 and 231, respectively. The intervention duration of the follow-up assessment for participants ranged from 1 month to 6 months. No statistically significant differences in gender, age, and education level were found between these 2 groups. Approximately two-thirds of these participating patients were females, 171 (67.59%) and 160 (69.26%) for the baseline and follow-up groups, respectively.

Table 1. Description of patients assessed by the Smart Web Aid for Preventing Type 2 Diabetes (SWAP-DM2).

Service output	Total	Gender		Age (years)[a]		
		Male	Female	≥50	51-60	61-70
Patients assessed with elevated prediabetes risk,[b] n	2219	807	1412	825	590	804
Patients took fasting glucose test, n (%)	1885 (84.94)	675 (83.64)	1210 (85.69)	679 (82.30)	515 (87.28)	691 (85.95)
Patients tested with prediabetes, n (%)	1022 (54.22)	348 (51.56)	674 (55.70)	339 (49.93)	261 (50.68)	422 (61.07)
Number of newly diagnosed prediabetes, n (%)	974 (95.30)	333 (95.69)	641 (95.10)	327 (96.46)	249 (95.40)	398 (94.31)
Number of newly diagnosed diabetes	103 (5.46)	39 (5.78)	64 (5.29)	39 (5.74)	26 (5.05)	38 (5.50)

[a] Age ranges are approximate.

[b] Prediabetes denotes fasting glucose ≥5.6 mmol/L and <7.0 mmol/L.

Delivery of Essential Prevention Procedures

Before the SWAP-DM2 application, the proportion of patients who received the essential prevention procedures assessed was extremely low. The most commonly delivered essential prevention procedure was measuring blood pressure, accounting for 43.9% (111/253) of the patients, followed by assessing diet behavior (17.0%, 43/253), and physical activity (13.0%, 33/253). Compared with baseline, SWAP-DM2 utilization significantly increased delivery of almost all essential prevention procedures listed in Table 2 except for glucose measurement, which decreased from 9.1% (23/231) to 1.3% (3/231). Assessing diet behavior (17.0%, 43/253 to 100.0%, 231/231), physical activity (13.0%, 33/253 to 95.7%, 221/231), and BMI (10.7%, 27/253 to 100%, 231/231) witnessed the greatest improvement, followed by counseling barriers to (0.8%, 2/253 vs 81.4%, 188/231) and skills for (3.2%, 8/253 vs 90.9%, 210/231) modifying diet. No statistically significant difference was observed in delivery of the essential prevention procedures at baseline or follow-up between male and female patients.

Table 2. Essential prevention procedures delivered in traditional and SWAP-DM2–mediated service.

Essential prevention procedures	Male, n (%)			Female, n (%)			Total, n (%)		
	Baseline (n=82)	Follow-up (n=71)	*P*	Baseline (n=171)	Follow-up (n=160)	*P*	Baseline (n=253)	Follow-up (n=231)	*P*
Assessing diet behavior	16 (19.5)	71 (100)	<.001	27 (15.8)	160 (100)	<.001	43 (17.0)	231 (100)	<.001
Assessing physical activity	17 (20.7)	70 (98.6)	<.001	16 (9.4)	151 (94.4)	<.001	33 (13.0)	221 (95.7)	<.001
Measuring body mass index	9 (11)	71 (100)	<.001	18 (10.5)	160 (100)	<.001	27 (10.7)	231 (100)	<.001
Measuring blood pressure	33 (40.2)	64 (90.1)	<.001	78 (45.6)	150 (93.8)	<.001	111 (43.9)	214 (92.6)	<.001
Measuring blood glucose	7 (8.5)	0 (0)	.89	16 (9.4)	3 (1.9)	.54	23 (9.1)	3 (1.3)	.54
Counseling diabetes susceptibility	0 (0)	48 (67.6)	<.001	3 (1.8)	114 (71.3)	<.001	3 (1.2)	162 (70.1)	<.001
Counseling risks of unhealthy diet	5 (6.1)	43 (60.6)	<.001	10 (5.8)	96 (60.0)	<.001	15 (5.9)	139 (60.2)	<.001
Counseling risks of inadequate activity	4 (4.9)	42 (59.2)	<.001	8 (4.7)	94 (58.8)	<.001	12 (4.7)	136 (58.9)	<.001
Counseling barriers to modifying diet	1 (1.2)	58 (81.7)	<.001	1 (0.6)	130 (81.3)	<.001	2 (0.8)	188 (81.4)	<.001
Counseling barriers to increasing activity	0 (0)	49 (69.0)	<.001	1 (0.6)	112 (70.0)	<.001	1 (0.4)	161 (69.7)	<.001
Counseling skills to modify diet	0 (0)	63 (88.7)	<.001	8 (4.7)	147 (91.9)	<.001	8 (3.2)	210 (90.9)	<.001
Counseling skills to increase activity	0 (0)	59 (83.1)	<.001	3 (1.8)	137 (85.6)	<.001	3 (1.2)	196 (84.9)	<.001

Changes in Patients’ Immediate Indicators

As shown in Table 3, SWAP-DM2 helped to raise the prediabetic patients’ knowledge about the risks of diabetes, imbalanced diet, and inadequate physical activity. It also worked well in improving the patients’ self-efficacy in eating a healthier diet and increasing physical activities. Most importantly, SWAP-DM2–mediated prevention substantially improved key diabetes-related behaviors among the prediabetic patients, with the greatest improvement observed in increasing vegetable intake and a relatively smaller change for increasing exercise. Body weight and BMI of the prediabetic patients also showed statistically significant improvement. The patients’ satisfaction toward the service as a whole and their doctor’s responsiveness remained unchanged, whereas their satisfaction toward the service technology increased slightly. Most of these changes showed no statistically significant differences between gender groups. Self-efficacy for increasing physical activity and engaging relatives, however, had greater improvement in males than females, mean 8.64 (SD 2.31) vs mean 7.80 (SD 2.41, *P*=.01) and mean 7.80 (SD 2.41) vs mean 6.54 (2.70, *P*<.001), respectively.

Patients’ and Doctors’ Views Toward SWAP-DM2

All village doctors who tested the SWAP-DM2 and a convenience sample of 20 prediabetic patients who experienced the Web-facilitated diabetes prevention service participated in our qualitative interviews. The doctors viewed SWAP-DM2 as helpful and easy:

> *It’s an innovative and real expert system.*
>
> *It standardizes diabetes prevention and proposes pertinent procedures and information.*
>
> *With the Web, identifying high-risk patients and counseling lifestyle modification becomes easy.*
>
> *By following its steps, you won’t miss important things.*
>
> *It helps communicate balanced information comprising not only benefits but also “dis-benefits,” not only barriers but also measures to overcome barriers; and this is especially helpful in motivating sustainable lifestyle changes.*
>
> *Perhaps the biggest advantage is its ability in tailoring prevention to individual needs and bridging current and past services into continuous prevention.*

The doctors also raised several concerns about SWAP-DM2:

> *Although desktop computers and Internet are available at most village clinics, electricity supply is not totally secured.*
>
> *The Web adopts a step-by-step approach in counseling lifestyle changes; patients’ responses may go “awry” or may not quite fit the proposed procedures occasionally and this means doctors*

> *should have enough experience in leading the communication.*
>
> *The Web system does not work so well during outreach, although it is accessible via mobile phones; mobile communication service charges are not refundable by government or health insurance agencies for most village doctors at present.*

The patients' views toward SWAP-DM2–assisted prevention were also mostly positive:

> *A Web aid plus a doctor must be better than a doctor without any help; more importantly, it is a smart Web designed by experienced experts with a credited medical college.*
>
> *I had never thought of taking glucose test before. Yet I felt most necessary to do so when my doctor told me, after having asked a few questions according to the screen and entered my responses that I was at increased risk for diagnosing glucose impairment since...*
>
> *I felt pressing when my doctor pointed to the screen and said, "Your body weight keeps going up and let's decide on something about it."*
>
> *At each follow-up visit, my doctor always refers to the Web and says, "Let's see what we had decided to do last time," this reminds me repeatedly that the Web has a good memory and I should keep my words.*

With regard to concerns that computer operation may depersonalize service provision, most patients disagreed:

> *I am used to computer-aided services; it happens almost everywhere like supermarkets, banks and so on.*
>
> *I did not feel any uneasiness.*

However, some patients mentioned "my doctor referred to a computer from time to time, it was more or less different from a free discussion" or "being an old and illiterate farmer, I do not know how the Web had helped me."

Table 3. Patients’ knowledge, self-efficacy, satisfaction, and behavioral outcome measures.

Variable assessed	Male			Female			Total		
	Baseline (n=82)	Follow-up (n=71)	*P*	Baseline (n=171)	Follow-up (n=160)	*P*	Baseline (N=253)	Follow-up (N=231)	*P*
Knowledge about risks of diabetes/prediabetes, n (%)									
Leads to eye, kidney, heart lesions	2 (2.4)	55 (77.5)	<.001	5 (2.9)	129 (80.6)	<.001	7 (2.8)	184 (79.7)	<.001
Affects long-term objectives and development	5 (6.1)	39 (54.9)	<.001	7 (4.1)	83 (51.9)	<.001	12 (4.7)	122 (52.8)	<.001
Affects family and social relationships	3 (3.7)	31 (43.7)	<.001	3 (1.8)	60 (37.5)	<.001	6 (2.4)	91 (39.4)	<.001
Induces psychological and economic burdens	3 (3.7)	42 (59.2)	<.001	2 (1.2)	77 (48.1)	<.001	5 (2.0)	119 (51.5)	<.001
Knowledge about risks of imbalanced diet, n (%)									
Leads to overweight or obesity	2 (2.4)	43 (60.6)	<.001	8 (4.7)	108 (67.5)	<.001	10 (4.0)	151 (65.4)	<.001
Causes hypertension	10 (12.2)	60 (84.5)	<.001	6 (3.5)	118 (73.8)	<.001	16 (6.3)	178 (77.1)	<.001
Leads to cerebral and cardiovascular diseases	4 (4.9)	33 (46.5)	<.001	3 (1.8)	62 (38.8)	<.001	7 (2.8)	95 (41.1)	<.001
Induces diabetes	5 (6.1)	49 (69)	<.001	7 (4.1)	102 (63.8)	<.001	12 (4.7)	151 (65.4)	<.001
Leads to cancer	0 (0.0)	19 (26.8)	<.001	3 (1.8)	56 (35.0)	<.001	3 (1.2)	75 (32.5)	<.001
Knowledge about risks of inadequate physical activity, n (%)									
Leads to overweight or obesity	5 (6.1)	63 (88.7)	<.001	15 (8.8)	151 (94.4)	<.001	20 (7.9)	214 (92.6)	<.001
Causes hypertension	4 (4.9)	53 (74.7)	<.001	8 (4.7)	120 (75.0)	<.001	12 (4.7)	173 (74.9)	<.001
Leads to cerebral and cardiovascular diseases	3 (3.7)	31 (43.7)	<.001	5 (2.9)	66 (41.3)	<.001	8 (3.2)	97 (42.0)	<.001
Induces diabetes	5 (6.1)	34 (47.9)	<.001	4 (2.3)	65 (40.6)	<.001	9 (3.6)	99 (42.9)	<.001
Leads to cancer	1 (1.2)	28 (39.4)	<.001	0 (0.0)	45 (28.1)	<.001	1 (0.4)	73 (31.6)	<.001
Reduces body immunity	2 (2.4)	36 (50.7)	<.001	5 (2.9)	84 (52.5)	<.001	7 (2.8)	120 (51.9)	<.001
Self-efficacy ratings, [a] mean (SD)									
Modifying diet	5.38 (2.79)	8.59 (2.23)	<.001	5.28 (2.85)	8.50 (2.28)	<.001	5.31 (2.81)	8.53 (2.25)	<.001
Increasing physical activities	4.85 (3.26)	8.64 (2.31)	<.001	4.36 (3.39)	7.80 (2.41)	<.001	4.52 (3.35)	8.06 (2.38)	<.001
Refusing snacks	8.97 (2.00)	9.88 (1.88)	<.001	8.64 (2.29)	9.43 (2.15)	<.001	8.75 (2.20)	9.57 (2.08)	<.001
Engaging relatives in diabetes prevention	4.41 (3.21)	7.80 (2.41)	<.001	3.70 (3.58)	6.54 (2.70)	<.001	3.93 (3.54)	6.93 (2.67)	<.001
Behaviors practiced, n (%)									
Reduced calorie intake	2 (2.4)	55 (77.5)	<.001	2 (1.2)	110 (68.8)	<.001	4 (1.6)	165 (71.4)	<.001
Increased vegetable intake	15 (18.3)	64 (90.1)	<.001	28 (16.4)	141 (88.1)	<.001	43 (17.0)	205 (88.7)	<.001
Increased leisure-time exercises	5 (6.1)	15 (21.1)	<.001	11 (6.4)	34 (21.3)	<.001	16 (6.3)	49 (21.2)	<.001
Immediate outcome measures, mean (SD)									
Body weight (kg)	67.31 (10.03)	63.68 (9.35	.023	59.63 (9.00)	55.95 (9.11)	<.001	62.12 (9.85)	58.33 (9.18)	<.001
Body mass index (kg/m^2)	24.32 (3.15)	23.01 (2.86)	.008	25.03 (3.32)	23.52 (2.97)	<.001	24.80 (3.21)	23.36 (2.95)	<.001
Systolic blood pressure (mm Hg)	141.21 (21.07)	137.23 (19.56)	.23	135.57 (21.80)	131.54 (20.07)	.08	137.40 (21.50)	133.29 (19.80)	.03
Diastolic blood pressure (mm Hg)	87.89 (13.63)	87.01 (12.89)	.68	84.4 (11.98)	82.92 (11.65)	.26	85.90 (12.67)	84.18 (12.02)	.13

Variable assessed	Male			Female			Total		
	Baseline (n=82)	Follow-up (n=71)	*P*	Baseline (n=171)	Follow-up (n=160)	*P*	Baseline (N=253)	Follow-up (N=231)	*P*
Satisfaction ratings, [a] **mean (SD)**									
Toward service provided	9.28 (1.53)	9.11 (1.48)	.49	8.82 (1.86)	8.73 (1.82)	.66	8.97 (1.77)	8.85 (1.73)	.45
Toward service techniques used	8.86 (1.89)	9.20 (1.76)	.25	8.71 (1.94)	9.10 (1.78)	.06	8.76 (1.92)	9.13 (1.77)	.03
Toward doctor's responsiveness	9.49 (1.07)	9.40 (1.05)	.60	9.05 (1.54)	9.04 (1.52)	.95	9.20 (1.42)	9.15 (1.41)	.70

[a] Maximum=10.

Discussion

Effectiveness and Acceptability of SWAP-DM2

By using SWAP-DM2, 17 village doctors identified 2219 patients with elevated prediabetes and diabetes risk, encouraged 85.0% of them perform a fasting glucose test, and diagnosed 1022 prediabetes and 103 new diabetes cases within only 6 months. More importantly, these prediabetic patients showed substantial improvement from baseline to 6 months after intervention in terms of vegetable intake, calorie intake, leisure-time activities, and even body weight and BMI. The prediabetic patients also witnessed great improvements in self-efficacy in modifying diet, increasing physical activities, etc. These findings are consistent and point to an encouraging implication that SWAP-DM2 is effective, at least in the short term. The primary reason underlying this effectiveness may be that SWAP-DM2 incorporates, in a synergetic way, a variety of proven strategies and techniques. For example, it ensures delivery of essential prevention procedures (Table 3) through required SOPs, leverages behavior changes via convincing evidences, such as changes and trends in the patient's own BMI, fasting glucose, etc, and facilitates counseling lifestyle modification by using motivational interviewing. In contrast to most other essential prevention procedures, glucose testing decreased from baseline to 6-month follow-up. This was because the SOPs required that the test occur only annually for patients with elevated risk. The nonsignificant changes in the patients' satisfaction variables may be due primarily to the fact that the same satisfaction ratings at baseline were already quite high and thus room for improvement was limited.

This preliminary evaluation also revealed indications that SWAP-DM2 was acceptable to both villagers and village doctors. These included the positive comments solicited during the qualitative interviews and the frequent use of the Web as indicated by the large volume of SWAP-DM2–mediated risk assessments, glucose tests, and counseling sessions delivered during the short trial period, and the more than 88.5% satisfaction of the villagers toward the Web-assisted services. A number of factors may have contributed to this acceptance: (1) SWAP-DM2 transforms complex diabetes education or counseling into step-by-step procedures and makes it easy to learn and practice; (2) SWAP-DM2 enriches service content and procedures with little added workload by real-time data recording and retrieving, and automatic behavior risk identification, classification, and lifestyle modification planning; and (3) SWAP-DM2 gains credibility (for doctors) and confidence (in patients) by providing highly professional and tailored suggestions and by demonstrating high-tech (ie, Internet and expert system) use. There existed a sharp discrepancy between the number of female patients identified as being prediabetic and female participants in both the baseline and follow-up assessments than males. This may not necessarily indicate that the SWAP-DM2–aided prevention was more acceptable among females that males. Rather, it may be explained by the fact that most male villagers were pursuing temporary jobs in cities during the trial period.

Limitations

First, this study documented only a segment of an ongoing project [29]. Although the overall design of this project belongs to a quasi-randomized controlled trial, findings presented in this paper allowed for only baseline vs follow-up rather than intervention vs control comparisons. Therefore, readers are cautioned about a variety of potential confounding factors in interpreting the findings (eg, national or regional diabetes education programs during the same period of this study). Second, the evaluation of SWAP-DM2 covered only 6 months of intervention and used mostly short-term variables. Whether the Web-mediated prevention continues to work smoothly in the long run when doctors and patients have lost curiosity about Web-based systems and whether the immediate indicators (eg, knowledge, self-efficacy, healthy lifestyle practices, BMI) ultimately lead to desired outcomes (eg, reduced diabetes progression) needs further investigation. Some previous studies showed that body weight, fasting glucose, etc, decreased from baseline to their lowest level at approximately 6 months after intervention and then began to rise again slowly [34]. Third, there may be some seasonal effects. The villagers may have different amounts of farm work to do in different seasons. Their food availability and eating and cooking habits may also change across seasons. This may result in quite different findings over the 6 months. Fourth, the SWAP-DM2-mediated prevention service (including glucose tests, blood and body weight measurements, etc) was provided free of charge to the villagers and the participating doctors were reimbursed, through a temporary project, for their time and efforts invested at a minimum rate (approximately US $5 per prevention case per year). This may raise sustainability concerns, although it would not be very difficult to get permanent government support if this prevention proves to be cost-effective. In addition, readers are reminded of the measurement limitations because most of the patient outcomes were derived from self-repot data that were prone to various biases (eg, social desirability and recency biases).

Lessons From SWAP-DM2 Development and Application

Given the number of elements involved and the complex relationships between them, it is almost impossible to distinguish effects of any specific elements except the overall SWAP-DM2–supported intervention as a whole. Therefore, the key to development of similar systems and transferring lessons from this study to future interventions is leveraging smart Web applications in selecting and combining proven elements into an integrated package. Another strategy getting around the complexity just mentioned toward effectiveness is rapid prototyping followed by continuous refinement. This includes not only the application and assessment cycles as described in the methodology section, but also the inbuilt self-learning mechanisms with the Web-based system as shown in Multimedia Appendix 2 (eg, learned tailoring of risk assessment and behavior counseling). Projects of this kind depend heavily on effectively bridging 2 heterogeneous aspects of expertise (ie, diabetes prevention and information technology). Although repeated interaction between these 2 specialty groups helped in our case, the leading role of a diabetes prevention expert who has adequate knowledge of Web-based applications of modern information technologies seemed to be even more important. Finally, necessary structural changes should accompany application of soft-systems such as this. As introduced in the umbrella protocol of this study, a set of performance-based incentives were implemented alongside SWAP-DM2 application. These may have played some role in the short-term—yet encouraging—findings of this study. In addition, approaches for promoting application of the system also merits careful consideration. For instance, the village doctors were often misled by the seemingly complex workflows and functions as described in Figure 1 and Multimedia Appendix 2. They should only be used to provide readers with an overview of what SWAP-DM2 can do in overall rather than what it does in a single encounter. A typical SWAP-DM2–assisted intervention session involves only part of the steps and takes 10 to 40 minutes depending on the patient's needs. Although the system "behind" the interface is quite sophisticated, the Web pages were designed to be as smooth and simple as possible. Therefore, stepwise learning by doing may be an effective way to gain and train village doctors.

Conclusions

Our preliminary findings suggest that SWAP-DM2 is helpful to village doctors, acceptable to patients, and effective in modifying immediate determinants of diabetes at least in the short term. Rapidly growing diabetes cases contrasted by a general lack of routinely operational prevention against the epidemic calls for immediate engagement of frontier clinicians, especially village doctors who work in rural China where the majority of the nation's vast population lives [35-37]. Given the long history of separating disease prevention and treatment and the extremely limited resources in rural China, this depends heavily on effectively tackling a series of barriers [38,39]. SWAP-DM2 may provide a useful solution in reaching this end.

Acknowledgments

Development of the primitive project protocol was supported by the Natural Science Foundation of China (grant number: 81172201). Refinement and Implementation of the study is lead and supported by Lu'an CDC.

Authors' Contributions

Jing Chai and Penglai Chen contributed equally in facilitating intervention protocol, SOPs, and Web development, and in drafting this manuscript. Kaichun Li and Shaoyu Xie led field data collection. Han Liang and Xingrong Shen programmed the website. Jing Cheng supervised field data collection. Rui Feng analyzed the data and revised the manuscript. Debin Wang provided expertise for the overall design of the study and finalized the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Guiding Models and principles for developing standardized operation procedures (SOPs) of counseling diabetes prevention.

[PDF File (Adobe PDF File), 40KB - jmir_v16i4e98_app1.pdf]

Multimedia Appendix 2

Main functions of Smart Web Aid for Preventing Diabetes (SWAP-DM2).

[PDF File (Adobe PDF File), 48KB - jmir_v16i4e98_app2.pdf]

Multimedia Appendix 3

Sample application webpages of Smart Web Aid for Preventing Type 2 Diabetes (SWAP-DM2).

[PDF File (Adobe PDF File), 153KB - jmir_v16i4e98_app3.pdf]

Multimedia Appendix 4

Exit survey for Smart Web Aid for Preventing Type 2 Diabetes (SWAP-DM2) assessment.

[PDF File (Adobe PDF File), 30KB - jmir_v16i4e98_app4.pdf]

References

1. Aponte J. Prevalence of normoglycemic, prediabetic and diabetic A1c levels. World J Diabetes 2013 Dec 15;4(6):349-357 [FREE Full text] [doi: 10.4239/wjd.v4.i6.349] [Medline: 24379926]
2. Wang Q, Ma A, Han X, Zhao S, Cai J, Ma Y, et al. Prevalence of type 2 diabetes among newly detected pulmonary tuberculosis patients in China: a community based cohort study. PLoS One 2013;8(12):e82660 [FREE Full text] [doi: 10.1371/journal.pone.0082660] [Medline: 24367535]
3. Abu-Rmeileh NM, Husseini A, Capewell S, O'Flaherty M, MEDCHAMPS project. Preventing type 2 diabetes among Palestinians: comparing five future policy scenarios. BMJ Open 2013;3(12):e003558 [FREE Full text] [doi: 10.1136/bmjopen-2013-003558] [Medline: 24362011]
4. Rathmann W, Giani G. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care 2004 Oct;27(10):2568-9; author reply 2569. [Medline: 15451946]
5. Chen AH, Tseng CH. The role of triglyceride in cardiovascular disease in asian patients with type 2 diabetes--a systematic review. Rev Diabet Stud 2013;10(2-3):101-109. [doi: 10.1900/RDS.2013.10.101] [Medline: 24380086]
6. Liu J, Zhao D, Liu J, Qi Y, Sun J, Wang W. Prevalence of diabetes mellitus in outpatients with essential hypertension in China: a cross-sectional study. BMJ Open 2013;3(11):e003798 [FREE Full text] [doi: 10.1136/bmjopen-2013-003798] [Medline: 24259390]
7. Yang SH, Dou KF, Song WJ. Prevalence of diabetes among men and women in China. N Engl J Med 2010 Jun 24;362(25):2425-6; author reply 2426. [Medline: 20578276]
8. Li MZ, Su L, Liang BY, Tan JJ, Chen Q, Long JX, et al. Trends in prevalence, awareness, treatment, and control of diabetes mellitus in mainland china from 1979 to 2012. Int J Endocrinol 2013;2013:753150 [FREE Full text] [doi: 10.1155/2013/753150] [Medline: 24288530]
9. Wang KW, Shu ZK, Cai L, Wu JQ, Wei W. Assessment of the magnitude of contextual and individual demographic effects on diabetes mellitus and glucose intolerance in rural Southwest China: a multilevel analysis. PLoS One 2013;8(7):e68553 [FREE Full text] [doi: 10.1371/journal.pone.0068553] [Medline: 23874667]
10. Zhao XL, He XY, Zhang HS, Lu B, Li JH, Li YM, et al. A better instrument for screening diabetes in rural areas of China: an equation developed from multivariate logistic regression or a simplified scoring form. Biomed Environ Sci 2013 Jun;26(6):496-499 [FREE Full text] [doi: 10.3967/0895-3988.2013.06.011] [Medline: 23816584]
11. Chen SC, Tseng CH. Dyslipidemia, kidney disease, and cardiovascular disease in diabetic patients. Rev Diabet Stud 2013;10(2-3):88-100. [doi: 10.1900/RDS.2013.10.88] [Medline: 24380085]
12. Reinehr T. Type 2 diabetes mellitus in children and adolescents. World J Diabetes 2013 Dec 15;4(6):270-281 [FREE Full text] [doi: 10.4239/wjd.v4.i6.270] [Medline: 24379917]
13. Mezuk B, Chen Y, Yu C, Guo Y, Bian Z, Collins R, et al. Depression, anxiety, and prevalent diabetes in the Chinese population: findings from the China Kadoorie Biobank of 0.5million people. J Psychosom Res 2013 Dec;75(6):511-517. [doi: 10.1016/j.jpsychores.2013.09.008] [Medline: 24290039]
14. Perreault L, Færch K. Approaching Pre-diabetes. J Diabetes Complications 2014;28(2):226-233. [doi: 10.1016/j.jdiacomp.2013.10.008] [Medline: 24342268]
15. Shen L, Zhang YG, Liu M, Qiang DC, Sun XL, Liu L, et al. Increased arterial stiffness in subjects with pre-diabetes among middle aged population in Beijing, China. Biomed Environ Sci 2013 Sep;26(9):717-725 [FREE Full text] [doi: 10.3967/0895-3988.2013.09.002] [Medline: 24099605]
16. Kyrios M, Moore SM, Hackworth N, Buzwell SA, Crafti N, Critchley C, et al. The influence of depression and anxiety on outcomes after an intervention for prediabetes. Med J Aust 2009 Apr 6;190(7 Suppl):S81-S85. [Medline: 19351299]
17. Talbot P, Dunbar MJ. Nova Scotia Prediabetes Project: upstream screening and community intervention for prediabetes and undiagnosed type 2 diabetes. Chronic Dis Inj Can 2011 Dec;32(1):2-11 [FREE Full text] [Medline: 22153171]
18. Kuo YL, Wu SC, Hayter M, Hsu WL, Chang M, Huang SF, et al. Exercise engagement in people with prediabetes - a qualitative study. J Clin Nurs 2013 Dec 12. [doi: 10.1111/jocn.12424] [Medline: 24330396]
19. Anderson ZL, Scopelliti EM, Trompeter JM, Havrda DE. Management of Prediabetes: A Comparison of the Treatment Approaches Utilized by a Family Practice Clinic and an Internal Medicine/Endocrinology Practice. J Pharm Pract 2013 Dec 10. [doi: 10.1177/0897190013514089] [Medline: 24326412]
20. Rowan CP, Riddell MC, Jamnik VK. The Prediabetes Detection and Physical Activity Intervention Delivery (PRE-PAID) Program. Can J Diabetes 2013 Dec;37(6):415-419. [doi: 10.1016/j.jcjd.2013.09.003] [Medline: 24321723]
21. Maindal HT, Bonde A, Aagaard-Hansen J. Action research led to a feasible lifestyle intervention in general practice for people with prediabetes. Prim Care Diabetes 2014 Apr;8(1):23-29. [doi: 10.1016/j.pcd.2013.11.007] [Medline: 24361372]

22. Lu YH, Lu JM, Wang SY, Li CL, Zheng RP, Tian H, et al. Outcome of intensive integrated intervention in participants with impaired glucose regulation in China. Adv Ther 2011 Jun;28(6):511-519. [doi: 10.1007/s12325-011-0022-4] [Medline: 21533568]
23. Lin CW, Abdul SS, Clinciu DL, Scholl J, Jin X, Lu H, et al. Empowering village doctors and enhancing rural healthcare using cloud computing in a rural area of mainland China. Comput Methods Programs Biomed 2014 Feb;113(2):585-592. [doi: 10.1016/j.cmpb.2013.10.005] [Medline: 24296075]
24. Yan X, Liu T, Gruber L, He M, Congdon N. Attitudes of physicians, patients, and village health workers toward glaucoma and diabetic retinopathy in rural China: a focus group study. Arch Ophthalmol 2012 Jun;130(6):761-770. [doi: 10.1001/archophthalmol.2012.145] [Medline: 22801838]
25. Hu J, Gruber KJ, Liu H, Zhao H, Garcia AA. Diabetes knowledge among older adults with diabetes in Beijing, China. J Clin Nurs 2013 Jan;22(1-2):51-60. [doi: 10.1111/j.1365-2702.2012.04273.x] [Medline: 23216552]
26. Wang H, Zhang X, Zhang J, He Q, Hu R, Wang L, et al. Factors associated with prevalence, awareness, treatment and control of hypertension among adults in Southern China: a community-based, cross-sectional survey. PLoS One 2013;8(5):e62469 [FREE Full text] [doi: 10.1371/journal.pone.0062469] [Medline: 23671599]
27. Shen H, Edwards H, Courtney M, McDowell J, Wei J. Barriers and facilitators to diabetes self-management: perspectives of older community dwellers and health professionals in China. Int J Nurs Pract 2013 Dec;19(6):627-635. [doi: 10.1111/ijn.12114] [Medline: 24330214]
28. Tulloch H, Sweet SN, Fortier M, Capstick G, Kenny GP, Sigal RJ. Exercise facilitators and barriers from adoption to maintenance in the diabetes aerobic and resistance exercise trial. Can J Diabetes 2013 Dec;37(6):367-374. [doi: 10.1016/j.jcjd.2013.09.002] [Medline: 24321716]
29. Feng R, Li K, Cheng J, Xie S, Chai J, Wei P, et al. Toward integrated and sustainable prevention against diabetes in rural China: study rationale and protocol of eCROPS. BMC Endocr Disord 2013;13(1):28 [FREE Full text] [doi: 10.1186/1472-6823-13-28] [Medline: 23919331]
30. Chai J, Wang D, Zhou M, Xu W, Liang G, Shen YF. Developing and piloting an expert system for better routine voluntary HIV counseling and testing in China: preliminary results and lessons. AIDS Care 2012;24(4):424-433. [doi: 10.1080/09540121.2011.613907] [Medline: 22149044]
31. Lange N, Tigges BB. Influence positive change with motivational interviewing. Nurse Pract 2005 Mar;30(3):44-53. [Medline: 15753814]
32. Miller D, Shewchuk R, Elliot TR, Richards S. Nominal group technique: a process for identifying diabetes self-care issues among patients and caregivers. Diabetes Educ 2000 Apr;26(2):305-10, 312, 314. [Medline: 10865596]
33. Tani S, Marukami T, Matsuda A, Shindo A, Takemoto K, Inada H. Development of a health management support system for patients with diabetes mellitus at home. J Med Syst 2010 Jun;34(3):223-228. [Medline: 20503606]
34. Diabetes Prevention Program Research Group, Knowler WC, Fowler SE, Hamman RF, Christophi CA, Hoffman HJ, et al. 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. Lancet 2009 Nov 14;374(9702):1677-1686 [FREE Full text] [doi: 10.1016/S0140-6736(09)61457-4] [Medline: 19878986]
35. Shi L, Shu XO, Li H, Cai H, Liu Q, Zheng W, et al. Physical activity, smoking, and alcohol consumption in association with incidence of type 2 diabetes among middle-aged and elderly Chinese men. PLoS One 2013;8(11):e77919 [FREE Full text] [doi: 10.1371/journal.pone.0077919] [Medline: 24223743]
36. Liang H, Chen X, Chen Q, Wang Y, Wu X, Li Y, et al. The metabolic syndrome among postmenopausal women in rural Canton: prevalence, associated factors, and the optimal obesity and atherogenic indices. PLoS One 2013;8(9):e74121 [FREE Full text] [doi: 10.1371/journal.pone.0074121] [Medline: 24040183]
37. Perria C, IMPLEMEG Study Group. Strategies for the introduction and implementation of a guideline for the treatment of type 2 diabetics by general practitioners (GPs) of the Lazio region of Italy (IMPLEMEG study): protocol for a cluster randomised controlled trial [ISRCTN80116232]. BMC Health Serv Res 2004 Jun 14;4(1):13 [FREE Full text] [doi: 10.1186/1472-6963-4-13] [Medline: 15196307]
38. Appiah B, Hong Y, Ory MG, Helduser JW, Begaye D, Bolin JN, et al. Challenges and opportunities for implementing diabetes self-management guidelines. J Am Board Fam Med 2013 Feb;26(1):90-92 [FREE Full text] [doi: 10.3122/jabfm.2013.01.120177] [Medline: 23288286]
39. Tian M, Chen Y, Zhao R, Chen L, Chen X, Feng D, et al. Chronic disease knowledge and its determinants among chronically ill adults in rural areas of Shanxi Province in China: a cross-sectional study. BMC Public Health 2011;11:948 [FREE Full text] [doi: 10.1186/1471-2458-11-948] [Medline: 22192681]

Abbreviations

BMI: body mass index
DM2: type 2 diabetes mellitus
eCROPS: educating doctors and electronic supports, counseling diabetes prevention, recipe for lifestyle management, operational toolkit, performance-based reimbursement for doctors, and screening service
SOP: standardized operation procedure

SWAP-DM2: Smart Web Aid for Preventing Type 2 Diabetes

Edited by G Eysenbach; submitted 05.01.14; peer-reviewed by L Ruggiero, L Ning; comments to author 31.01.14; revised version received 05.02.14; accepted 13.03.14; published 01.04.14.

Please cite as:
Chen P, Chai J, Cheng J, Li K, Xie S, Liang H, Shen X, Feng R, Wang D
A Smart Web Aid for Preventing Diabetes in Rural China: Preliminary Findings and Lessons
J Med Internet Res 2014;16(4):e98
URL: http://www.jmir.org/2014/4/e98/
doi:10.2196/jmir.3228
PMID:24691410

Original Paper

Adolescents Just Do Not Know What They Want: A Qualitative Study to Describe Obese Adolescents' Experiences of Text Messaging to Support Behavior Change Maintenance Post Intervention

Kyla L Smith[1], BSc, PostGradDipDiet; Deborah A Kerr[2], BAppSc, GradDipDiet, MSc, PhD; Ashley A Fenner[3], BSc, BA; Leon M Straker[1], BAppSc, MSc, PhD

[1]School of Physiotherapy and Exercise Science, Faculty of Health Sciences, Curtin University, Perth, Australia

[2]School of Public Health, Faculty of Health Sciences, Curtin University, Perth, Australia

[3]Health Psychology and Behavioural Medicine Research Group, School of Psychology and Speech Pathology, Faculty of Health Sciences, Curtin University, Perth, Australia

Corresponding Author:
Kyla L Smith, BSc, PostGradDipDiet
School of Physiotherapy and Exercise Science
Faculty of Health Sciences
Curtin University
Building 408
GPO Box U1987
Perth, 6845
Australia
Phone: 61 8 9266 3694
Fax: 61 8 9266 3699
Email: kyla.smith@curtin.edu.au

Abstract

Background: Adolescents are considered a hard to reach group and novel approaches are needed to encourage good health. Text messaging interventions have been reported as acceptable to adolescents but there is little evidence regarding the use of text messages with overweight and obese adolescents to support engagement or behavior change after the conclusion of a healthy lifestyle program.

Objective: The intent of this study was to explore the opinions of overweight adolescents and their parents regarding the use of text messages as a support during the maintenance period following an intervention.

Methods: This paper reports on the findings from focus groups conducted with adolescents (n=12) and parents (n=13) who had completed an eight-week intensive intervention known as Curtin University's Activity, Food and Attitudes Program (CAFAP). Focus groups were conducted three months post intensive intervention. Participants were asked about their experiences of the prior three-month maintenance phase during which adolescents had received tri-weekly text messages based on the self-determination theory and goal-setting theory. Participants were asked about the style and content of text messages used as well as how they used the text messages. Data were analyzed using content and thematic analyses.

Results: Two clear themes emerged from the focus groups relating to (1) what adolescents liked or thought they wanted in a text message to support behavior change, and (2) how they experienced or responded to text messages. Within the "like/want" theme, there were five sub-themes relating to the overall tone of the text, frequency, timing, reference to long-term goals, and inclusion of practical tips. Within the "response to text" theme, there were four sub-themes describing a lack of motivation, barriers to change, feelings of shame, and perceived unfavorable comparison with other adolescents. What adolescents said they wanted in text messages often conflicted with their actual experiences. Parent reports provided a useful secondary view of adolescent experience.

Conclusions: The conflicting views described in this study suggest that overweight and obese adolescents may not know or have the ability to articulate how they would best be supported with text messages during a healthy lifestyle maintenance phase. Further, supporting both engagement and behavior change simultaneously with text messaging may not be possible. Intervention texts should be personalized as much as possible and minimize feelings of guilt and shame in overweight and obese adolescents. Future research with text messaging for overweight and obese adolescents should incorporate clear intervention aims and evaluation methods specifically related to adolescent engagement or behavior change.

Trial Registration: Australian New Zealand Clinical Trials Registry: ACTRN12611001187932; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?ACTRN=12611001187932 (Archived by WebCite at http://www.webcitation.org/6LGSbk8d9).

(J Med Internet Res 2014;16(4):e103) doi:10.2196/jmir.3113

KEYWORDS

telemedicine; text messaging; adolescent; obesity

Introduction

The provision of effective health-related interventions for adolescents is a difficult task, particularly when traditional methods of communication are unlikely to be engaging [1]. Current mobile telephone usage trends show adolescents are increasingly using text messaging as a preferred method of communication [2]. Text messaging, also known as short message service (SMS), may therefore provide an acceptable means of delivering health messages in adolescent populations [3-6].

In adolescent weight management programs, text messaging shows promise as a feasible and acceptable method of communication [4,7]. Evidence from healthy weight adolescent groups support this concept of “acceptability” [8], yet it remains unknown whether text messaging is an effective means for fostering engagement or supporting behavior change.

Given the significant dropout rates reported in pediatric weight management programs [9] and lack of evidence for long-term weight maintenance in adolescence [10], there is much to be gained from developing an effective method for keeping participants engaged with a program. Text messaging may be a useful way of maintaining this engagement. De Niet et al [6] found that children and adolescents randomized to a text message treatment group were 3.5 times less likely than the control group to drop out of the maintenance period of a healthy lifestyle intervention but also showed that text message engagement declined over the 9-month period. Older pre-adolescent children (12 year olds sent 0.5 texts per week) were less likely to send responses compared with younger children (7 year olds sent 0.8 texts per week) [6]. This decline in adherence to monitoring over time has been documented previously [11], but adherence remains higher than other traditional methods of self-monitoring. Kornman et al [7] described only a modest level of adolescent engagement with their text messaging and email adjunct to the Loozit program, with adolescents responding to 22.0% of text messages over a 10-month maintenance period. The authors found that text messages asking for a response had more replies and text messages had a more immediate response than the emails. Whether text messaging can be used to maintain engagement of overweight adolescents in a healthy lifestyle program remains unclear.

Studies evaluating the effectiveness of text messages offer inconsistent support for the use of text messaging to support behavior change and/or weight maintenance [12-16]. Adolescents in the Loozit program rated the text message support during the 10-month maintenance period as “somewhat helpful”; however, the results suggested that the use of text messages developed for this intervention did not have a significant effect on primary outcomes at 12 or 24 months post program [7,15,16]. Other short-term studies assessing the effectiveness of text messaging interventions with adolescents suggest success only in those who had not previously engaged in the targeted behaviors before the intervention commenced [12] or who received additional intensive support (ie, intensive insulin therapy) in addition to text messages [13]. There remains a gap in the evidence regarding the effectiveness of text messaging interventions to support behavior change in overweight and obese adolescents.

There is also limited evidence regarding the best way to construct and send text messages for use with adolescents [17]. Factors to consider include the message tone, content and length, as well as the timing and frequency of text message delivery. Several studies have reported the wording and timing of delivered messages [3,6,7,12], but only two studies have considered the format of text messages to be sent regularly to support behavior change in adolescents [8] or obese adolescents [17]. Rigorous pilot testing of text messages has provided descriptions of what adolescents think they want [8,17], but whether adolescents actually do know and can articulate what they want in text messages and whether this is desirable for either engagement or intervention efficacy is unknown. There are currently no studies that have completed a detailed post-intervention evaluation of overweight and obese adolescent experiences of text messages, including perceptions of content, style, or usefulness. A lack of synchrony between what adolescents think they want and may actually like or find useful is suggested by evidence that overweight teenagers have indicated a strong preference for directive text messages, which seems to contradict their desire for autonomy [17].

Parents have not traditionally been involved in the development or refinement of text messaging interventions for adolescents. However, parents may play an important role in supporting adolescent engagement in research [1], providing healthful environments [18], modelling healthy behaviors [19], and supporting adolescent behavior change [20]. The potential for parents to offer valuable insight into adolescent experiences of text messaging interventions warrants exploration.

Current recommendations for research into text messaging interventions advise researchers to design their study with a strong theoretical foundation [21,22] and report on process evaluation after the intervention, to inform others about the best way to structure text messages, and effectively utilize this form of communication with adolescents [7,17,21,23]. Therefore, the aim of this study was to explore opinions of overweight and obese adolescents and their parents, who have participated in a multi-disciplinary healthy lifestyle program, regarding the use of text messages as a support during the maintenance period.

Methods

Study Design

This study relates to focus groups held during the maintenance phase of a multi-disciplinary healthy lifestyle program for overweight and obese adolescents aged 12-16 years and their parents. A wait-list controlled trial of Curtin University's Activity, Food and Attitudes Program (CAFAP) has been conducted and is described elsewhere [24] along with its theoretical underpinnings [20]. Briefly, participants enrolled in CAFAP and were placed on a three-month waiting list to allow for a control period. Following this waiting list period, adolescents and parents completed the intensive phase of the program, involving twice-weekly group sessions for eight weeks. Sessions covered healthy eating, increasing physical activity, reducing sedentary behavior, and setting goals for healthy behavior change. Parents were also trained in behaviors around need satisfaction and goal setting to support their adolescent's lifestyle changes [20]. The intensive phase was followed by a tapered maintenance phase over 12 months. The first three months of the maintenance phase were considered to be "high intensity" where adolescents received three text messages per week and one phone coaching session every two weeks. The next three-month period included a tapering of contact to "medium intensity", where adolescents received a weekly text message and a monthly phone coaching session. The "low intensity" phase of the maintenance period occurred between six and 12 months post intensive program and included monthly text messages and quarterly phone coaching sessions. This study was approved by the Curtin University Human Research Ethics Committee (HR105/2011).

This qualitative study was performed three months after the eight-week intervention period of CAFAP, which coincided with the conclusion of the high intensity maintenance period. The intent of the study was to explore adolescents' perceptions of text messages received, in combination with the observations and experiences of the parents, during the previous three months. Four focus groups were conducted with overweight and obese adolescents between September 2012 and April 2013 to cover the three waves of participants who completed the program at different times. Four focus groups were also completed by the parents of these adolescents. Focus groups were small (3-4) to encourage rich discussion of the text messages.

Recruitment

Three months after completion of the intensive face-to-face CAFAP sessions, adolescents from each wave were invited by flyer, email, and text message to participate in a 60-minute focus group. Adolescents were offered a $30AUD gift voucher for participating in the focus group. Of the 35 adolescents invited, 16 agreed to participate and 12 attended the focus groups. Parents were recruited once their adolescent had agreed to participate in the focus group. All adolescents had a parent participate in the focus groups at the same time as they completed their own focus group.

Text Message Development and Programming

CAFAP used an automated text messaging system to send predetermined but semi-tailored text messages to adolescents. During the first three months of the 12-month maintenance period, adolescents were sent a text message at 6pm on two weekdays and noon on one weekend day. The timing and frequency of text messages in this study was chosen as it reflected a midway point between previous adolescent trials [7,13] and incorporated evidence from associated formative research [25]. Adolescents chose the days that would best suit them at the conclusion of the intensive intervention phase. Two versions of the message plan were developed using the same text messages but in a different order to ensure that adolescents who were close friends would receive different messages. Texts were limited to 320 characters minus auto text ("Hi [first name], [text message], from the CAFAP Team"). Texts were sent to the adolescent's personal phone in most instances, except where the adolescent did not have a personal phone or did not use it consistently. Some families requested the text also be sent to a parent's phone to keep the parent involved. If text messages were not able to be delivered, research staff would follow up to manually send the text or contact the participant by phone if the text remained undelivered.

The contact during the maintenance period continued with the same theoretical base and key messages as during the intensive face-to-face contact period. Text contact was thus based on self-determination theory [26] and goal-setting theory [27] and focused on eating more fruit and vegetables, eating less junk food, being less sedentary, and being more physically active [20,24]. Message development also incorporated recommendations from current literature, by including a positive tone [6,17,23] and friendly but professional language [17]. Messages were constructed so as to be perceived as deriving from the research team [7,17] and were semi-tailored [28] by using first names and references to CAFAP or past text message contact [4-7,13,17,21]. The word "you" was frequently used to make facts relatable to adolescents [8] and a maximum of two reflective questions were included in each message [17]. Messages that provided options for behaviors were always framed to give adolescents a choice (eg, "You might like to…") and recipes or testimonials from other teenagers (eg, "Other teens have found it helpful to…") were used to enhance self-efficacy [17]. Suggestions for activities were based on affective beliefs to emphasize the enjoyment and social nature of participation [12] and reaffirmed the benefits of healthy eating and physical activity. Negative wording or the mention of triggers for unhealthy behaviors (eg, consumption of junk food) was avoided [17]. Adolescents were able to reply to text messages and receive a response, but were not expressly asked to reply. This approach was chosen to align with the theoretical underpinnings of text message development relating to fostering a sense of autonomy [20,26] and formative work suggesting adolescents may be reluctant to spend their own money on responses [25]. Text messages were designed to prompt behaviors and offer adolescents options for healthy behaviors, but they did not have to act on the texts if that was not their choice. The number of text messages sent to adolescents was measured on the SMS database. Five overall categories of texts were developed (general, goal setting, healthy eating, physical activity, and sedentary behavior) with messages tailored for weekday evenings and weekends. See Table 1 for examples of CAFAP text messages.

Table 1. Text message examples used to represent the content of text messages used in the intervention and typical adolescent responses from focus groups.

Category	Text message	Key strategies	Examples of adolescent responses[a]
Healthy eating	After dinner can be a good time to eat some fruit. If you had less than 2 bits of fruit today, you might like some tinned apricots & yoghurt for dessert?	Message including a ‘helpful tip’	*I liked it except I didn’t have any of that stuff.* [Adolescent 10 (A10)] *I’d prefer if you told me to have strawberries.* [A12] *It’s ok but I don’t really eat after dinner.* [A4]
Sedentary behavior	A CAFAP teen has found using an egg-timer is a good way of knowing it’s time for an active break after playing on the computer for 30 mins. You could choose to try it too.	Testimonial message	*Most probably wouldn’t influence my behavior because, I don’t know, I get addicted to my computer.* [A3]
Physical activity	What are the reasons you want to be more active? You might like to think about these when the going gets tough!	Intrinsic motivation, reflecting goal-setting process used during the program	*It makes you actually think about it. You have it there but you don’t actually think about it.* [A7]
Physical activity	How many steps have you done today? How about trying to do a few extra thousand steps this afternoon?	Reflective questioning	*There’s nothing wrong with it, but like saying a few extra thousand like is too much.* [A5] *It would make you think about doing more steps.* [A1]
Goal setting	Have a look at your goals for physical activity, sedentary behavior and healthy eating. Plan something fun to do this weekend to help meet your goals.	Bigger picture, reinforcing key CAFAP areas and goal setting.	*I have homework on the weekend, like every day… If I’m not doing homework I’m sleeping. Like pretty much.* [A11]
Healthy eating	How many bits of junk food have u had today? You could challenge yourself to see if you can go the rest of the day without any junk food. Some cut up fruit might be a sweet treat instead!	Reflective questioning and ‘helpful tip’	*I felt really guilty because I think I’d had ummm I’d had something, maybe a donut for breakfast.* [A2]

[a] Adolescents were assigned a code to distinguish between responders. The code ranges from A1-A12.

Focus Group Protocol

Overview

Facilitators had completed formal training with a qualitative research expert covering focus group conduct prior to involvement in these focus groups and had prior experience in conducting focus groups with similar participants.

Adolescent Focus Groups

Adolescent focus groups were 60 minutes long and conducted by one author (KS). Each focus group was audiotaped with participant consent. In the adolescent groups, actual text messages were used to guide the discussion at the start of the focus group based on the work of Woolford et al [17] and previous difficulties noted by the research team when adolescents were asked to answer questions that required them to reflect on their experiences. Six messages were chosen from the database of texts sent during the high intensity maintenance phase to reflect the different content and key strategies used in the main text categories as outlined in Table 1.

The aim of using the six text messages was to explore the adolescents’ response to the message content. For each of the six messages, adolescents were asked:

- Did you like/dislike this message?
- What would you first think when you got this message?
- Is it a realistic message for you? (Is it *right* for you?)
- What’s good about it?
- What’s not so good about it?
- Did it make you think about changing your behavior (diet/activity/habit)?
- How could we make it better?

For the second part of the focus group, adolescents were asked about another six text messages (see Table 2) to understand their responses to the specific strategies used in the messages. The final part of the focus group included questions about general responses, timing, and how the text messages were used by adolescents.

Table 2. Text message examples used to represent the style of text messages sent in the high intensity maintenance period.

Question to adolescents	Example text message	Examples of adolescent responses[a]
What did you think of messages that asked you a question?	What was your sedentary behavior goal for today? Did you achieve your goal?	*For Monday to Fridays, it's a bit hard because we have loads of double periods…and we're sitting down for long periods of time so we've kinda destroyed our sedentary goal.* [Adolescent 12 (A12)]
What did you think of messages that reminded you about why being healthy was important?	Do you remember that the benefits of being more active include having a healthier heart and body chemistry, feeling less tired, sleeping better, being happier, and thinking better?	*Yeah. It helps you because it asks you a question but it instantly answers it for you.* [A4] *It's good to remember. Feeling less tired would be good.* [A2]
What did you think of the 'big picture' messages?	Remember the key messages of CAFAP are: eat more fruit and veg, eat less junk food, be less inactive, and be more active.	*It actually like reminds you that you're like on Earth and you're sitting there playing video games…And I've gotta like do stuff and be proactive like normal people.* [A9]
What did you think of the messages that gave you healthy tips?	Think about how many veggies u had today. If u had less than your goal, try to add in 1 more piece tomorrow. How about some veggie sticks with one of the yummy dips u made at CAFAP?	*It was helpful. I often followed the tips.* [A8] *I found it good because I wouldn't think of something like that.* [A6]
What did you think of the messages that included ideas from other teenagers?	Some teenagers have told us that cut up fruit salad for recess helps them reach their healthy eating goals. What about taking some tomorrow?	*These sound like healthy teenagers and sometimes I get a bit annoyed because I'm jealous. They have a better state of mind than me.* [A4]
What do you think about messages that are about what *not* to do?	Don't eat junk food today or don't spend the afternoon lying on the couch.[b]	*If you tell someone not to do something, they're just going to do it, everyone knows that.* [A5]

[a] Adolescents were assigned a code to distinguish between responders. The code ranges from A1-A12.

[b] This text message was created to be tested in the focus groups based on current literature [17] but was not actually used in the intervention.

Parent Focus Groups

Parent groups were also 60 minutes long, conducted by one of two authors (LS or AF) and audiotaped with participant consent. This secondary view from parents is potentially helpful to attempt to validate the information provided by adolescents, particularly from interviews or focus groups where there is potential for participants to express a more socially acceptable view and obscure their true opinions. The group was asked three main questions regarding the experiences of their family in the previous three months, their opinions on the usefulness of the text message support, and any suggestions to better support adolescents after completion of the intensive eight-week program.

Data Analysis

Following focus group discussions, responses to questions were transcribed verbatim by KS, with confidentiality ensured by coding of transcripts with subject identifier codes. The data was sorted and coded and assigned to categories based on similar phrases and topics. Coding was completed separately by two authors (KS and DK), with peer review checks by LS and AF to ensure the overall credibility of findings and interpretations [29]. Initial thematic analysis was based on the structure of the research questions [30] to identify theoretical constructs that described the experience of adolescents and parents. Themes were allowed to emerge using a mostly inductive approach [30]. Differences in interpretation were resolved by consensus. Underlying similarities and differences were evaluated and used to form the fundamental impression of the focus group discussions [31]. The data were triangulated with adolescent and parent interpretations compared [31] to give greater context to the data. By comparing and contrasting results from two groups with different viewpoints, we attempted to overcome the intrinsic biases associated with single group observations and explain the situation more fully [32]. Categories were amalgamated using Microsoft Excel and the major themes detailed using description and quotes from participants to support these findings [29].

Results

Demographics

Twelve adolescents with a mean BMI (body mass index) *z* score of 2.05 (SD 0.35) participated in the focus groups. The mean age of the adolescent participants was 14.3 (SD 1.5) years, with females overrepresented (92%, 11/12) when compared to all CAFAP participants who participated in the text message intervention (77% female, 33/43). A total of 13 parents, including 12 mothers and one father, participated in the parent focus groups. The majority of participants were white Australians from middle-low socioeconomic areas. Details regarding household characteristics were not further explored. The focus group participants included adolescents who had varying levels of success in adopting healthy behaviors at the three-month assessment and were likely a good representation of the overall group.

Text Message Descriptive Statistics

In this intervention, 37% (16/43) of adolescents did not have access to their own phone regularly and had the text messages sent to their parent's phone. These adolescents made up 33%

(4/12) of the focus group participants. Of the parents, 60% (26/43) of parents in the intervention received a copy of the text messages, which included some parents of adolescents who had their own phone. A total of 2240 text messages were sent to adolescents (not including additional messages sent to parents only), eliciting 152 replies at a response rate of 6.79%.

Participant Opinions

Overview

All attendees were generous with their feedback and engaged meaningfully in the discussions. Two distinct themes emerged from the adolescent and parent discussions relating to (1) what adolescents liked or thought they wanted in a text message to support behavior change, and (2) how they experienced or responded to text messages. Within the "like/want" theme, there were five sub-themes relating to the overall tone of the text, frequency, timing, reference to long-term goals, and inclusion of practical tips. Within the "response to text" theme, there were four sub-themes describing a lack of motivation, barriers to change, feelings of shame, and perceived unfavorable comparison with other adolescents. Themes are described and supported with quotes from adolescents (A) and parents (P) below.

More Casual and Personalized Text Messages Are Preferred

Adolescents were unanimous in their reaction to the tone of the texts, highlighting a need to make the messages less formal and more relatable to them. Parents expressed a similar view, indicating that the tone of the messages needed to be more personal.

> *More smiley faces. Smiley faces are good.* [A1]
>
> *They're really proper. You need to abbreviate and stuff. Make it seem more human.* [A3]
>
> *They sound a bit rehearsed at times. A bit impersonal. Sometimes more casual…instead of Hi XX, it could be like Hey or Hey XX. More chatty. Because the ones that were being sent sound like for an adult. They're too formal.* [A5]
>
> *Yes—make 'em more 21st century.* [A4]
>
> *Make them custom for each person, because each person does a different amount of each thing…So for the stuff that we already do, the text messages should be like different for everyone.* [A9]
>
> *The SMS were too impersonal. It was like it was coming from a machine.* [P7]

Tri-Weekly Text Messages Were Too Frequent

Adolescents and parents all agreed that the messages were sent too frequently and as a result the reaction to the texts became increasingly negative over time.

> *If it had been say one a whole week, say on a Wednesday because that's the middle of the week, to see how you're going but to help as well, that'd be good. But three continuous in a week, that's just like Shut Up!* [A5]
>
> *It's just another form of nagging to them.* [P7]
>
> *They kept on coming. Like why can't you let us do what we're doing and then we tell you eventually when we come back.* [A4]
>
> *I know the messages got to a stage where, it's almost like she would duck.* [P10]
>
> *With the messages, she was fine with them to begin with, but after that she only wanted to know about a few.* [P6]
>
> *Towards the end I would see CAFAP and put it to one side.* [A2]

Lack of Consistent Response on the Best Time to Receive Messages

Adolescent views varied regarding the best time to receive text messages, although most thought the timing needed to be specific for each individual.

> *7 o'clock in the morning would be better for me because it's just before I go to school. It's when you check your phone.* [A8]
>
> *No, no, no, no, no. If you text messaged me before school I would absolutely call you and go off at you because you ain't texting me before school. I am a bad morning person and getting a text message from CAFAP would just blow it. It's a teenager thing.* [A5]
>
> *Wednesdays are a good day.* [A1]
>
> *Fridays, so you get them before the weekend.* [A8]
>
> *Weekends are more helpful.* [A2]
>
> *Have more texts in the holidays.* [A6]
>
> *Don't text me during school holidays because I will not get them. I go into hibernation.* [A12]

Adolescents Wanted To Be Reminded To Think About Their Reasons for Wanting To Change

The messages that included triggers for adolescents to think about their long-term goals, reasons for wanting to make changes, and their decision to participate in CAFAP were generally appreciated**.** The supportive nature of these texts seemed to be better received than text messages related to behavior change.

> *I mean cuz, it's like where do you want to be from here, now? What are you going to do to get to that spot? And it just kind of motivates you to say "Oh, that's my dream and I'm getting there."* [A10]
>
> *Maybe like "Why did you choose to do this program? Remember those and keep going." I like that.* [A11]
>
> *It reminds you that you've gotta do that because you've done the program but now you've got to do it yourself.* [A12]

Practical Tips Were Valuable Inclusions in Text Messages

A strong theme that emerged was the preference of adolescents to receive practical and relevant examples of behavior change. This approach of providing positive ideas for adolescents to choose to engage in seemed to be better received than the reflective questioning style used in the intervention or negative framed messages (what not to do) that were tested in the focus groups but not used in the intervention.

> *Having a healthy tip that you can actually do is good.* [A1]
>
> *The being happier and thinking better…sometimes I need an example. Like what am I going to do?* [A2]

Text Messages Were Not Effective Motivators for Change for Adolescents

Adolescents reported that although the text messages often acted as a reminder or an awareness raiser, they were not able to motivate behavior change. Adolescents were quick to come up with reasons for why the text message wasn't applicable to them or why they wouldn't be able to use the healthy tips. Adolescents appeared not to be interested in thinking too deeply about the text messages or having to adapt the content to suit their lives. Parents highlighted that adolescent lack of motivation was a common barrier and thought the text messages were not successful in encouraging behavior change.

> *That's a good message because it tells you what you should do and should not do, but you're still going to be inactive and you're still going to eat junk food.* [A4]
>
> *Not really good. What's wrong with it is (1) I don't have an egg timer, (2) I don't really plan on buying one, and (3) I don't think that anyone would stay on the computer for [just] 30 minutes, I think they'd want to be on there for longer.* [A5]
>
> *She does talk about things like going to Zumba classes with a friend but it's all talk, nothing's been done.* [P3]
>
> *Don't think it made any difference. But then it's hard to tell what's going on in there sometimes.* [P12]
>
> *She's made all of these wonderful decisions; she's at that point where she's gotta keep motivated to stay there. And that's where she's having a little bit of trouble.* [P9]

Lack of Time and Tiredness Were Barriers for Adolescents to Participate in Healthy Behaviors

Adolescents repeatedly emphasized their lack of free time to plan for or perform healthy behaviors. They identified many cases where they disregarded text messages because they didn't feel there was enough time to implement any of the strategies offered. Adolescents also reported being too tired to participate in healthy behaviors and this was supported by parent opinions.

> *I don't have time; I don't even eat breakfast in the morning. We just don't have that much time.* [A10]
>
> *Homework. I didn't have time to think about it [the text messages] at all. I've got like five projects at the moment.* [A9]
>
> *I answer it [the text] in my head but the thing is I'm already tired. So by the time you're on your weekend, you just crash.* [A2]
>
> *But the exercise is not there. I think part of it is their school is really full on. They're too tired.* [P11]

Adolescents Stated Some Texts Induced Feelings of Shame

Adolescents reported a sense of guilt or shame associated with messages that reminded them about healthy behaviors that they were not implementing. Parents also reiterated this sense of shame emerging for their adolescent, stemming from a number of texts.

> *Thanks for making me feel like crap, cuz I'm not the healthiest and I'm not the fittest and I feel really tired. And maybe I'm not as happy as I could be. [That] would just make me feel bad.* [A3]

(in response to the text: Do you remember that the benefits of being more active include having a healthier heart and body chemistry, feeling less tired, sleeping better , being happier, and thinking better?)

> *That just makes me feel really depressed if I've eaten a lot of junk that day.* [A12]
>
> *It's good to have healthy reminder but I notice that my daughter does respond with a level of guilt or shame. It's a reminder in a way that she's not doing, she could be doing more.* [P1]

Hearing About What Other Teenagers Are Doing Was Not Motivating For Adolescents

Similarly, the majority of teens described their dislike for messages that included ideas or experiences from other teens, reiterating this sense of shame that others were improving their health while they were not.

> *I think "Oh, other teens are better than me? Like OK maybe I should catch up a bit because next time I see them they'll all be skinny and I'll be like…still…still here."* [A2]
>
> *I think what my daughter rolled her eyes at the most was "the other teens" and she felt like she's being compared again…she would just sort of turn off.* [P4]

A minority reported that they enjoyed hearing about how other teens were experiencing similar challenges; however, this was disparate to the perceptions of their parents' reports, as well as the facilitators' knowledge of the adolescents' progress.

Discussion

Principal Findings

Overview

The present study contributes to the evidence base around experiences of overweight and obese adolescents and their parents in response to text messages designed to support behavior change in the maintenance period of a larger intervention. Adolescents described a sense of shame in response to the text message intervention, which was also observed by their parents, presenting a new issue for health researchers using text messaging with overweight and obese adolescents. In this study, overweight and obese adolescents' stated preferences for text messages differed from how they actually responded to such text messages. This suggests that pre-intervention testing of messages may not adequately simulate what happens in “real life”. There may be differences between messages that adolescents say they like and those that are actually helpful in supporting behavior change. The consensus from adolescents was that the text messages were occasionally useful but sent too frequently and did not substantially help them to change their behaviors. Parents agreed that the text messages were too frequent and possibly too impersonal. They acknowledged there were some positive responses to the messages but the majority of the adolescents did not like to be reminded about healthy behaviors. These findings emphasize the uniqueness of overweight and obese adolescents and suggest a need to strengthen future intervention aims and evaluation methods.

Shame

Overweight and obese adolescents experience a sense of shame regarding their health and/or body [33,34]. The results of this study show that regular health-related text messages have the potential to heighten this sense of shame, which is a new issue for consideration by researchers using text messaging with overweight and obese adolescents. Based on the theoretical underpinnings of CAFAP [20], text messages were specifically worded to promote autonomy and support adolescent choices, yet many adolescents perceived the text messages as reminders of what they should be doing but weren’t doing. This response from overweight and obese adolescents differs to that described in similar studies, albeit shorter term, in adolescents of differing weight statuses [8] or diabetic adolescents [13] who reported to enjoy receiving health-related text messages. Similarly, suggestions about what other teens had found useful were not well-received as they had been in the pilot phase of a previous study [17], rather, these adolescents reported they disliked being compared to others. This response is potentially a significant barrier for overweight and obese adolescents to maintain healthy behaviors, with feelings of shame being related to poor mental health [34] and in turn to poorer self-efficacy and reduced ability to engage in lifestyle changes [35]. These feelings or circumstances may be unique to overweight and obese adolescents and suggest that results from trials involving non-obese adolescents may not be easily generalized to this group.

Want Versus Response

The results from this study suggest a notable difference between what overweight and obese adolescents say they prefer in a text message as opposed to what is actually useful to them in the maintenance phase following intervention. Adolescents reported that texts relating to their long-term goals were useful to motivate them, yet they had difficulty with maintaining behavior change and parents reported significant issues with adolescent motivation. Adolescents were able to point out flaws in the text message content and suggest improvements; however, many of these included strategies previously used or previously criticized by the adolescents themselves. Similarly, adolescents expressed a desire for practical tips to use in their daily lives reflecting current evidence [8,17], yet were quick to highlight reasons that would prevent them from regularly using the tips. They identified a number of barriers including lack of time or lack of resources (eg, particular foods not available), but weren’t able to suggest many ideas to overcome these. Interpretation of these results suggest that adolescents are able to identify idealistic concepts for text message style and content that may make theoretical sense, but have difficulty envisioning their actual response to such messages. This has potential implications for researchers doing pre-intervention testing with overweight and obese adolescents because the views they express may indicate that the intervention will be useful but might not be a true reflection of their experience in a text message intervention. Currently, the majority of the evidence around text message development for overweight and obese adolescents is based on the assumption that adolescents know what will be helpful to them in a text message and can articulate that; however, these results suggest that this may not be the case. Qualitative post-intervention research may provide the most contextual data regarding overweight and obese adolescents’ reaction to text messages, but even within these results there are inconsistencies between what adolescents say they want and what they actually respond well to.

Purpose of Text

Text messages are often used with adolescents to (1) encourage behavior change or maintenance, and (2) to foster engagement between adolescents and interventions; however, this research suggests that it may be difficult to concurrently achieve both of these aims in a weight-related intervention using text messages. The overweight and obese adolescents in this study did not like receiving text messages reminding them to maintain healthy behaviors, and thus such messages have the potential to damage the therapeutic relationship established during a face-to-face intervention and hence ongoing engagement.

In this study, overweight and obese adolescents’ responses to the text messages became increasingly negative over time. This suggests that their motivation may decline over time leading to failure to maintain behavior changes. Many parents in this study reiterated that adolescents didn’t like to be prompted about healthy things they could be doing, although they felt that it was sometimes useful to encourage behavior change. Despite adolescents reporting their dislike of the text messages, the healthy behavior focus may have had a positive effect on adolescents’ health. The effectiveness of using texts as healthy behavior prompts is not clear, with some evidence suggesting that this may be effective in promoting behavior change in the long-term [36,37], while other evidence suggests that eliciting a negative or shameful response to a message results in a reduction in self-efficacy and lack of behavior change [35]. This delicate balance between prompting behavior change and avoiding a shameful response may account for some of the differences reported in the literature relating to effectiveness of text messaging interventions in overweight and obese adolescents [4,17].

Conversely, messages not related to weight or healthy behaviors may be more effective at keeping adolescents engaged with an intervention. In this study, adolescents expressed a desire for more casual and positive text messages, which may have been more suited to preserving links with the intervention rather than eliciting behavior change. If the intervention aim was to preserve links with adolescents, then text messages would need to be carefully constructed to foster this sense of engagement but could be completely unrelated to the behavioral aims of the intervention. Message development based on self-determination theory would therefore aim to promote a sense of connection for the adolescent [26]. To our knowledge, the use of text messaging interventions to foster engagement has not been tested independent of behavior change and should be explored in future research.

The current findings suggest that future studies would benefit from generating clear aims for text messaging interventions specifically relating to either behavior change or engagement, and designing the texts with these aims in mind. Similarly, evaluation methods need to be strengthened and appropriate measures used to assess whether the text messages actually achieve their intended aims.

Timing

Adolescents expressed a number of differing opinions regarding the timing of the text messages, suggesting that it may be most acceptable if timing is personalized to each participant's preferences. In the current intervention, messages were sent three times per week. Adolescents and parents consistently reported that the frequency of the text messages was excessive and they found that the messages became boring over time. This is in contrast to other studies with even greater frequencies of contact [4,5,8,12,13] where adolescents were reported to be accepting of daily texts, although 20% of participants in one study indicated that they got bored of receiving the same or similar text messages [13]. Interestingly, only one of these studies [4] was directed at overweight and obese adolescents, suggesting that this group may have unique perceptions regarding the frequency of supportive text messages and this may be related to how well they are achieving their lifestyle goals. Ngyuen et al [16] sent a text message once per month on average to overweight and obese adolescents but suggested that this dose was possibly too low to have an impact. The desire for less frequent text messages in this study may be related to the sense of guilt overweight and obese adolescents describe when they are reminded about changes they aren't making, although it remains unclear as to whether this can prompt behavior change. These results highlight how messages can possibly invoke an undesirable response in overweight and obese adolescents. Text messages become personal because they are sent to the participant's phone and differ from other health messages on media such as television where the individual may be able to more easily ignore a message as they are perceived as not relevant to them. When an individual receives and reads the text message, they will make a decision as to how relevant or useful the information is for them. From the responses obtained, it appears that, in overweight and obese adolescents, health messages may invoke a response that reduces their motivation to change rather than stimulating them to continue to take action to increase their physical activity or improve their diet. Health researchers may therefore need to modify the logistics of text message delivery to best suit the aims of their intervention, be that engagement or behavior change.

Style

Focus group participants were unanimous in their call for more casually worded text messages. Adolescents have previously expressed a preference for health care providers to communicate via text using a natural tone while avoiding the use of text message "slang" as frequently used in adolescent-to-adolescent contact [17]. Despite attempts to use less formal language than in the previous study, our research suggests that the tone we used was perceived as too professional by adolescents. The addition of symbols and emoticons (eg, exclamation marks and smiley faces) to text messages may help to convey a less formal tone. Although most text messaging studies have used tailored messages [4-7,13,17,21,28], our research suggests that overweight and obese adolescents want them to be further customized to their individual goals and experiences. Adolescents also reported that negatively worded texts (eg, "don't eat junk food") were likely to increase the likelihood of actually performing those behaviors, supporting recent findings by Woolford et al [17]. These authors [17] suggested that overweight and obese adolescents often want to avoid the psychological work associated with reflection and instead prefer tips and testimonials from other adolescents, yet our results suggest that overweight and obese adolescents in actuality do not find these helpful to assist with behavior change.

Parents

No other studies to our knowledge have included parent perceptions of their overweight and obese adolescent's responses to receiving supportive text messages. Including parent views helps to understand the differences in opinions that adolescents might express to parents when at home, as opposed to researchers. This triangulation of the data provides a richer sense of what the "real" truth may be. The views expressed by parents in this study were generally aligned with the opinions described by adolescents and were also useful in identifying inconsistencies described by individual adolescents whose views consistently differed from the overall group.

Limitations

As with all qualitative studies, there are potential limitations associated with participants reporting a view that may be perceived as socially acceptable, rather than their true opinion. It is possible that overweight and obese adolescents in particular may be prone to report in this way. The focus group facilitators in this study were known to the participants, which may have influenced participants to offer a more positive view on the use of text messages; however, the sometimes negative opinions expressed by parents and adolescents suggest this was not the case. The small size of the focus groups may have either encouraged or impeded discussion. It is also not possible to determine directly how the messages affected their behavior and if they took action. For example, some irritation with the messages may still have served as a reminder to increase their physical activity or improve their diet. The strengths of this study include an appropriate qualitative design to address the research question, a range of informants with directly relevant opinions for the research question, the inclusion of parents and adolescents, and provision of novel, in-depth information about the challenging topic of providing support to overweight and obese adolescents following an intensive intervention period.

Implications for Research and Practice

This study presents several implications to consider in future research. First, overweight and obese adolescents may not know or have the ability to articulate how they would best be supported during the maintenance phase. Second, researchers should generate clear aims when planning a text message intervention, and develop appropriate methods of measuring success relative to the specific aims (for example engagement or behavior change). Third, timing and content of messages may need to be more individualized and may require ongoing input from adolescents to understand their preferences for receiving text messages. Fourth, reports of parent experiences may offer a useful secondary view of adolescent responses to text message interventions. Last, overweight and obese adolescents might be more susceptible to feelings of guilt or shame than non-obese adolescents when reminded about maintaining behavior change.

Conclusions

Findings from this study suggest that text messages may have a useful role in interventions for overweight and obese adolescents but that adolescent opinions pre-intervention may not be very predictive of their actual experience and that it may be difficult to achieve both enhanced engagement and behavior change simultaneously. The results emphasize the importance of message tone, content, and timing as these factors appear to impact on how the messages will be received by overweight and obese adolescents. The potential to initiate feelings of shame in adolescent recipients of text messaging was identified as an important issue. The study also supported the use of process evaluation to help inform further interventions, and highlights the value of parent and adolescent reports. Future research should be clear on the engagement or efficacy aims for an intervention, develop messages to accommodate adolescent preferences in conjunction with an appropriate theory, and employ appropriate evaluation measures of engagement or effectiveness.

Acknowledgments

This trial was funded by a Healthway Health Promotion Research Project Grant #19938. Professor Leon Straker was supported by a National Health and Medical Research Council senior research fellowship. No funding or other input to the study was received from any pharmaceutical company. The authors would like to thank the adolescents and parents who shared their valuable insights with the team.

Authors' Contributions

KS designed and completed the adolescent focus groups, transcribed the data, analyzed the results, and drafted the manuscript. LS conceived of the study, participated in its design and coordination, including facilitating parent focus groups, and helped to review the manuscript. DK contributed to the development of the focus groups, identification of categories, and helped to review the manuscript. AF contributed to the development of the focus groups, facilitated the parent focus groups, and helped to review the manuscript.

Conflicts of Interest

None declared.

References

1. Steinbeck K, Baur L, Cowell C, Pietrobelli A. Clinical research in adolescents: challenges and opportunities using obesity as a model. Int J Obes (Lond) 2009 Jan;33(1):2-7. [doi: 10.1038/ijo.2008.263] [Medline: 19065147]
2. Lenhart A. Pew Internet & American Life Project. Washington DC; 2012. Teens, smartphones and texting URL: http://www.pewinternet.org/2012/03/19/teens-smartphones-texting/ [accessed 2013-11-13] [WebCite Cache ID 6OTroiRNX]
3. Franklin V, Waller A, Pagliari C, Greene S. "Sweet Talk": text messaging support for intensive insulin therapy for young people with diabetes. Diabetes Technol Ther 2003;5(6):991-996. [doi: 10.1089/152091503322641042] [Medline: 14709202]
4. Woolford SJ, Clark SJ. Tailored mobile phone text messages as an adjunct to obesity treatment for adolescents. Journal of Adolescent Health 2009 Feb;44(2):S1. [doi: 10.1016/j.jadohealth.2008.10.012]
5. Shapiro JR, Bauer S, Hamer RM, Kordy H, Ward D, Bulik CM. Use of text messaging for monitoring sugar-sweetened beverages, physical activity, and screen time in children: a pilot study. J Nutr Educ Behav 2008 Dec;40(6):385-391 [FREE Full text] [doi: 10.1016/j.jneb.2007.09.014] [Medline: 18984496]
6. de Niet J, Timman R, Bauer S, van den Akker E, de Klerk C, Kordy H, et al. Short message service reduces dropout in childhood obesity treatment: a randomized controlled trial. Health Psychol 2012 Nov;31(6):797-805. [doi: 10.1037/a0027498] [Medline: 22468714]
7. Kornman KP, Shrewsbury VA, Chou AC, Nguyen B, Lee A, O'Connor J, et al. Electronic therapeutic contact for adolescent weight management: the Loozit study. Telemed J E Health 2010 Aug;16(6):678-685. [doi: 10.1089/tmj.2009.0180] [Medline: 20575613]

8. Hingle M, Nichter M, Medeiros M, Grace S. Texting for health: the use of participatory methods to develop healthy lifestyle messages for teens. J Nutr Educ Behav 2013;45(1):12-19. [doi: 10.1016/j.jneb.2012.05.001] [Medline: 23103255]
9. Skelton JA, Beech BM. Attrition in paediatric weight management: a review of the literature and new directions. Obes Rev 2011 May;12(5):e273-e281 [FREE Full text] [doi: 10.1111/j.1467-789X.2010.00803.x] [Medline: 20880126]
10. Butryn ML, Wadden TA, Rukstalis MR, Bishop-Gilyard C, Xanthopoulos MS, Louden D, et al. Maintenance of weight loss in adolescents: current status and future directions. J Obes 2010;2010:789280 [FREE Full text] [doi: 10.1155/2010/789280] [Medline: 21274275]
11. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013;15(4):e32 [FREE Full text] [doi: 10.2196/jmir.2283] [Medline: 23587561]
12. Sirriyeh R, Lawton R, Ward J. Physical activity and adolescents: an exploratory randomized controlled trial investigating the influence of affective and instrumental text messages. Br J Health Psychol 2010 Nov;15(Pt 4):825-840. [doi: 10.1348/135910710X486889] [Medline: 20156396]
13. 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]
14. Shrewsbury V, Chou A, Steinbeck K, Nguyen B, Baur L, Lee A, et al. Adolescent engagement in additional therapeutic contact for overweight management via short message service and electronic mail: the Loozit study. Journal of Adolescent Health 2010 Feb;46(2):S15-S16. [doi: 10.1016/j.jadohealth.2009.11.036]
15. Nguyen B, Shrewsbury VA, O'Connor J, Steinbeck KS, Lee A, Hill AJ, et al. Twelve-month outcomes of the Loozit randomized controlled trial: a community-based healthy lifestyle program for overweight and obese adolescents. Arch Pediatr Adolesc Med 2012 Feb;166(2):170-177. [doi: 10.1001/archpediatrics.2011.841] [Medline: 22312175]
16. Nguyen B, Shrewsbury VA, O'Connor J, Steinbeck KS, Hill AJ, Shah S, et al. Two-year outcomes of an adjunctive telephone coaching and electronic contact intervention for adolescent weight-loss maintenance: the Loozit randomized controlled trial. Int J Obes (Lond) 2013 Mar;37(3):468-472. [doi: 10.1038/ijo.2012.74] [Medline: 22584456]
17. Woolford SJ, Barr KL, Derry HA, Jepson CM, Clark SJ, Strecher VJ, et al. OMG do not say LOL: obese adolescents' perspectives on the content of text messages to enhance weight loss efforts. Obesity (Silver Spring) 2011 Dec;19(12):2382-2387 [FREE Full text] [doi: 10.1038/oby.2011.266] [Medline: 21869762]
18. Pearson N, Biddle SJ, Gorely T. Family correlates of fruit and vegetable consumption in children and adolescents: a systematic review. Public Health Nutr 2009 Feb;12(2):267-283. [doi: 10.1017/S1368980008002589] [Medline: 18559129]
19. Hanson NI, Neumark-Sztainer D, Eisenberg ME, Story M, Wall M. Associations between parental report of the home food environment and adolescent intakes of fruits, vegetables and dairy foods. Public Health Nutr 2005 Feb;8(1):77-85. [Medline: 15705248]
20. Fenner AA, Straker LM, Davis MC, Hagger MS. Theoretical underpinnings of a need-supportive intervention to address sustained healthy lifestyle changes in overweight and obese adolescents. Psychology of Sport and Exercise 2013 Nov;14(6):819-829. [doi: 10.1016/j.psychsport.2013.06.005]
21. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]
22. Militello LK, Kelly SA, Melnyk BM. Systematic review of text-messaging interventions to promote healthy behaviors in pediatric and adolescent populations: implications for clinical practice and research. Worldviews Evid Based Nurs 2012 Apr;9(2):66-77. [doi: 10.1111/j.1741-6787.2011.00239.x] [Medline: 22268959]
23. de Niet J, Timman R, Bauer S, van den Akker E, Buijks H, de Klerk C, et al. The effect of a short message service maintenance treatment on body mass index and psychological well-being in overweight and obese children: a randomized controlled trial. Pediatr Obes 2012 Jun;7(3):205-219. [doi: 10.1111/j.2047-6310.2012.00048.x] [Medline: 22492669]
24. Straker LM, Smith KL, Fenner AA, Kerr DA, McManus A, Davis MC, et al. Rationale, design and methods for a staggered-entry, waitlist controlled clinical trial of the impact of a community-based, family-centred, multidisciplinary program focussed on activity, food and attitude habits (Curtin University's Activity, Food and Attitudes Program--CAFAP) among overweight adolescents. BMC Public Health 2012;12:471 [FREE Full text] [doi: 10.1186/1471-2458-12-471] [Medline: 22721261]
25. Smith KL, Straker LM, McManus A, Fenner AA. Barriers and enablers for participation in healthy lifestyle programs by adolescents who are overweight: a qualitative study of the opinions of adolescents, their parents and community stakeholders. BMC Pediatr 2014;14:53 [FREE Full text] [doi: 10.1186/1471-2431-14-53] [Medline: 24552207]
26. Deci EL, Ryan RM. The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. Psychological Inquiry 2000 Oct;11(4):227-268. [doi: 10.1207/S15327965pli1104_01]
27. Locke EA, Latham GP. A theory of goal setting and task performance. Englewood Cliffs, New Jersey: Prentice Hall; 1990.
28. Mauriello LM, Ciavatta MM, Paiva AL, Sherman KJ, Castle PH, Johnson JL, et al. Results of a multi-media multiple behavior obesity prevention program for adolescents. Prev Med 2010 Dec;51(6):451-456 [FREE Full text] [doi: 10.1016/j.ypmed.2010.08.004] [Medline: 20800079]
29. Gliner JA. Reviewing qualitative research - proposed criteria for fairness and rigor. Occup Ther J Res 1994;14(2):78-90.

30. Broom A. Using qualitative interviews in CAM research: a guide to study design, data collection and data analysis. Complement Ther Med 2005 Mar;13(1):65-73. [doi: 10.1016/j.ctim.2005.01.001] [Medline: 15907681]
31. Polgar S, Thomas SA. Introduction to research in the health sciences. 3rd ed. Melbourne: Churchill Livingstone; 1995.
32. Cohen L, Manion L, Morrison K. Research methods in education. 6th edition. London: Routledge; 2007.
33. Morinder G, Biguet G, Mattsson E, Marcus C, Larsson UE. Adolescents' perceptions of obesity treatment--an interview study. Disabil Rehabil 2011;33(12):999-1009. [doi: 10.3109/09638288.2010.520800] [Medline: 20874445]
34. Sjöberg RL, Nilsson KW, Leppert J. Obesity, shame, and depression in school-aged children: a population-based study. Pediatrics 2005 Sep;116(3):e389-e392 [FREE Full text] [doi: 10.1542/peds.2005-0170] [Medline: 16140683]
35. Melnyk BM, Small L, Morrison-Beedy D, Strasser A, Spath L, Kreipe R, et al. Mental health correlates of healthy lifestyle attitudes, beliefs, choices, and behaviors in overweight adolescents. J Pediatr Health Care 2006 Dec;20(6):401-406. [doi: 10.1016/j.pedhc.2006.03.004] [Medline: 17071371]
36. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]
37. Fjeldsoe B, Neuhaus M, Winkler E, Eakin E. Systematic review of maintenance of behavior change following physical activity and dietary interventions. Health Psychol 2011 Jan;30(1):99-109. [doi: 10.1037/a0021974] [Medline: 21299298]

Abbreviations

AUD: Australian dollars
BMI: body mass index
CAFAP: Curtin University’s Activity, Food and Attitudes Program
SMS: short message service

Edited by G Eysenbach; submitted 21.11.13; peer-reviewed by L Militello, D Schoffman; comments to author 11.02.14; revised version received 19.02.14; accepted 13.03.14; published 08.04.14.

Please cite as:
Smith KL, Kerr DA, Fenner AA, Straker LM
Adolescents Just Do Not Know What They Want: A Qualitative Study to Describe Obese Adolescents’ Experiences of Text Messaging to Support Behavior Change Maintenance Post Intervention
J Med Internet Res 2014;16(4):e103
URL: http://www.jmir.org/2014/4/e103/
doi:10.2196/jmir.3113
PMID:24713407

Review

Mobile Applications for Diabetics: A Systematic Review and Expert-Based Usability Evaluation Considering the Special Requirements of Diabetes Patients Age 50 Years or Older

Madlen Arnhold[1], Dipl-Soz; Mandy Quade[1], MSc; Wilhelm Kirch[1], DMD, MD

Research Association Public Health Saxony and Saxony-Anhalt, Medizinische Fakultät Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

Corresponding Author:
Madlen Arnhold, Dipl-Soz
Research Association Public Health Saxony and Saxony-Anhalt
Medizinische Fakultät Carl Gustav Carus
Technische Universität Dresden
Fiedlerstraße 33
Dresden, 01307
Germany
Phone: 49 351 458 13457
Fax: 49 351 458 5338
Email: Madlen.Arnhold@tu-dresden.de

Abstract

Background: A multitude of mhealth (mobile health) apps have been developed in recent years to support effective self-management of patients with diabetes mellitus type 1 or 2.

Objective: We carried out a systematic review of all currently available diabetes apps for the operating systems iOS and Android. We considered the number of newly released diabetes apps, range of functions, target user groups, languages, acquisition costs, user ratings, available interfaces, and the connection between acquisition costs and user ratings. Additionally, we examined whether the available applications serve the special needs of diabetes patients aged 50 or older by performing an expert-based usability evaluation.

Methods: We identified relevant keywords, comparative categories, and their specifications. Subsequently, we performed the app review based on the information given in the Google Play Store, the Apple App Store, and the apps themselves. In addition, we carried out an expert-based usability evaluation based on a representative 10% sample of diabetes apps.

Results: In total, we analyzed 656 apps finding that 355 (54.1%) offered just one function and 348 (53.0%) provided a documentation function. The dominating app language was English (85.4%, 560/656), patients represented the main user group (96.0%, 630/656), and the analysis of the costs revealed a trend toward free apps (53.7%, 352/656). The median price of paid apps was €1.90. The average user rating was 3.6 stars (maximum 5). Our analyses indicated no clear differences in the user rating between free and paid apps. Only 30 (4.6%) of the 656 available diabetes apps offered an interface to a measurement device. We evaluated 66 apps within the usability evaluation. On average, apps were rated best regarding the criterion “comprehensibility” (4.0 out of 5.0), while showing a lack of “fault tolerance” (2.8 out of 5.0). Of the 66 apps, 48 (72.7%) offered the ability to read the screen content aloud. The number of functions was significantly negative correlated with usability. The presence of documentation and analysis functions reduced the usability score significantly by 0.36 and 0.21 points.

Conclusions: A vast number of diabetes apps already exist, but the majority offer similar functionalities and combine only one to two functions in one app. Patients and physicians alike should be involved in the app development process to a greater extent. We expect that the data transmission of health parameters to physicians will gain more importance in future applications. The usability of diabetes apps for patients aged 50 or older was moderate to good. But this result applied mainly to apps offering a small range of functions. Multifunctional apps performed considerably worse in terms of usability. Moreover, the presence of a documentation or analysis function resulted in significantly lower usability scores. The operability of accessibility features for diabetes apps was quite limited, except for the feature “screen reader”.

(J Med Internet Res 2014;16(4):e104) doi:10.2196/jmir.2968

KEYWORDS

mobile applications; apps; mobile health; mHealth; diabetes mellitus; market analysis; systematic review; elderly; usability test; expert review

Introduction

Compared to early mobile phones, today's smartphones and tablet PCs offer a considerably wider range of functionalities. Mobile applications (apps) are increasingly used in managing various tasks in daily life. Currently, more than 900,000 apps are available in the Apple App Store (operating system: iOS, developer: Apple) and more than 700,000 apps in the Google Play Store (operating system: Android, developer: Google) [1]. The number of health-related apps increased to 31,000 in 2013 [2].

Within the health care sector, apps are supporting the management of illnesses, thereby promoting health awareness and well-being [3-5]. Specifically, a multitude of apps have been developed to assist patients in the management of diabetes mellitus type 1 or 2 [3,6]. For example, a topic-specific search in the Google Play Store resulted in more than 1000 hits. The ability, however, to sort the results according to individual needs is lacking. One reason for the large number of diabetes apps is the high and steadily increasing diabetes prevalence, especially among people older than 50 years [7-9]. In 2012, 371 million people between the ages of 20 and 79 suffered from diabetes worldwide and this number is estimated to increase to 552 million people by 2030 [10]. The high self-therapy potential certainly has a major influence on the high number of currently available apps.

We carried out a systematic review of all currently available diabetes apps for the operating systems iOS and Android, between February 2013 and April 2013. Our review aimed to provide an overview of the number of newly released apps, range of functions, target user groups, languages, acquisition costs, popularity/user ratings, the ability to connect to measurement devices, and the connection between acquisition costs and user ratings.

Diabetes prevalence increases with age. Thus, the elderly are a large target group that could benefit from diabetes apps. However, several studies have shown a lack of acceptance and a subpar use of innovative mobile technologies among this age group [11-15]. As one possible reason, Holzinger et al [11] and Mallenius et al [13] have pointed out the insufficient consideration of usability requirements of the elderly. Their experiences in handling mobile devices and apps are frequently limited. Inhibition thresholds and entry barriers are therefore particularly pronounced among this age group. In addition, cognitive and physical skills are declining with age [11] and result in needs that are considerably different from those of young users. Hence, this age group would benefit from apps that consider their specific usability requirements.

In order to better assess and quantify usability for the elderly, we carried out an expert-based usability evaluation based on a representative 10% sample of diabetes apps available as of April 2013. Therewith, we examined to what extent existing diabetes applications serve the usability requirements of diabetes patients aged 50 or older.

Until now, just a few reviews of diabetes apps had been conducted [3,5,6,16]. They differ from the review presented here in several ways: they considered a broader range of health care applications, they were restricted to one operating system, they reviewed solely the offered functionalities, or they were done more than one year ago. There is an absence of usability evaluation for diabetes apps [16,17]. Especially, formative usability evaluations of health apps are rare. To our knowledge, just one article has been published that links a diabetes app review with a formative usability evaluation [16], but their evaluation is limited to Android apps and gives no special consideration to the requirements of elderly diabetes patients as we do.

Methods

Systematic Review

Search and Screening Strategy

Our review focused on the leading operating systems for mobile devices, iOS and Android. The analysis was carried out using the Apple App Store for iOS apps and the Google Play Store for Android apps. We focused exclusively on diabetes apps available in English and German.

As a first step, we identified keywords to ensure that every relevant diabetes app was detected. Therefore, we chose the following German and English keywords, directly related to diabetes mellitus: Diabetes, Blood Sugar/Blutzucker, Glucose/Glukose. Every hit was reviewed in terms of its relevance and explicit link to diabetes mellitus. This pre-selection was necessary due to the growing number of misleading descriptions (spam techniques) for apps, caused partly by non-existent or low admission requirements for novel apps. In the Google Play Store, no admission requirements currently exist for newly developed apps, whereas iOS apps are first internally reviewed by an app review board. All apps with an explicit link to diabetes mellitus were included in the analysis. The basis for the systematic and comparative market analysis was defined by categories and respective subcategories/specifications outlined in Table 1.

We considered all the available information given by both the stores and the apps and collected the information for all categories and subcategories/specifications. In some cases, the structure of the app stores and the provided information differed strongly from one another, so we applied different approaches for the analysis of iOS and Android apps.

Table 1. Categories and respective subcategories/specifications extracted from diabetes apps.

Category	Subcategory/specifications
General information	
	App name
	App language
	Date of release/date of latest update (the acquisition of the release date was only possible for iOS apps; for Android apps, only the date of the latest update could be recorded)
	Availability of a desktop application
Operating system	
	App exclusively for the iOS operating system
	App exclusively for the Android operating system
	App for both operating systems available
Developer information	
	Name of the developer
Acquisition costs	
	Freeware
	Exact price
	Availability as “lite” version (paid apps sometimes offer free or cheaper lite versions with limited functionality)
Popularity/user ratings	
	Number of downloads/installations
	User rating
	Number of user ratings
Range of functions (multiple selection possible)	
	Documentation function
	Information function
	Data forwarding/communication function
	Analysis function
	Recipe suggestions
	Reminder function/timer
	Advisory function/therapy support
Target user groups	
	Patients
	Physicians/qualified health personnel
	Both user groups
Interfaces	
	Availability of an interface/connectivity to an external sensor(s)/device

Search and Screening Strategy for iOS Apps

The analysis of iOS apps was conducted using the information available in the Apple App Store. In contrast to the Google Play Store, the Apple App Store offers several options for filtering the search results by choosing thematic subcategories. The results can additionally be sorted by relevance, popularity, user rating, and date of release. During the survey period, a sorting function was only available for the iPad, so the whole iOS app survey was performed via the iPad.

For the analysis, we chose the subcategories “Health and Fitness” and “Medicine”. Subsequently, the displayed apps were sorted by their date of release. The date of release served as an objective characteristic, which was necessary for a reliable and reproducible acquisition of all diabetes apps. The number of hits given by the Apple App Store corresponded exactly to the number of relevant apps.

We checked every app hit with regard to its availability for iPad and iPhone. Additionally, we verified whether the app was offered exclusively for the operating system iOS or also for Android. The market analysis of diabetes apps for iOS resulted in 390 hits.

Search and Screening Strategy for Android Apps

By using the information available in the Google Play Store, the analysis of Android apps was conducted. To date, this app store offers no option to filter the search results for apps according to individual needs. Furthermore, the given “numbers of hits” is not only the number of apps but also the number of detected search terms in the app title and the app description. Thus, the search term “diabetes” led to more than 1000 hits in the Google Play Store. Keeping the limitations in mind, the number of available apps was a considerable overestimation.

In order to ensure a representative analysis despite missing selection criteria, we defined one day (03/06/2013) to record all found apps with title and developer. This definition will enable future app review processes. Additionally, every app was crosschecked for availability of an iOS version. Altogether, we found 380 diabetes apps available for the operating system Android.

Expert-Based Usability Evaluation

To examine the usability of currently available diabetes applications for the elderly, we performed an expert-based usability evaluation. With this method, usability experts put themselves in the role of potential or current users to examine products in terms of usability. We performed a summative evaluation as we exclusively included apps whose development was already finished [18].

Due to the high number of apps available for review, the usability evaluation was based on a representative 10% sample of existing diabetes apps as of April 2013. The sample was chosen on a random basis. The evaluation was performed by three independent experts, as suggested by Nielsen [19] and Barnum [20]. They were chosen due to their comprehensive experience in handling and testing mobile devices and applications with regard to usability for the elderly and operability of accessibility features. In addition, they had specific expertise in the field of diabetes and diabetes management. They were already involved in the accompanying systematic review and a survey of diabetes patients aged 50 or older and physicians investigating the acceptance factors of diabetes applications. The authors of this article were not involved in the usability evaluation for reasons of independence.

The basis for the usability evaluation was defined by a specially created set of usability criteria considering interaction processes, interface design, and comprehensibility of content (Table 2). Therefore, we reviewed usability guidelines (ISO, DIN) with explicit regard to the requirements of the elderly concerning mobile applications. Additionally, we considered usability requirements that have been proven as relevant in previous studies with this age group (Table 2). This guideline-based approach of usability testing is in accordance with the proposals of Nielsen [19], and Sarodnick and Brau [18]. We divided the selected criteria into main and subcriteria, added a clear description of their specific characteristics and defined respective assessment criteria. The experts rated each subcriterion and the expression of its characteristics by means of a 5-point Likert scale to grade the evaluation or by means of a dichotomous scale [16,19-21]. The main criteria were not evaluated themselves; their scores were calculated from the mean of the respective subcriteria.

To lower barriers for persons with reduced or limited cognitive and physical skills, iOS and Android offer different accessibility features. We tested the operability of three features for each tested app in a separate test run. We have chosen features that are relevant to the elderly and were offered by both operating systems:

- Screen reader—Voice over (iOS)/Talk back (Android): dichotomous scale
- “Larger Type” as an additional measurement for “possibility to flexibly adapt the size of operating elements and displayed images”: dichotomous scale
- “Invert colors” as an additional measurement for “sufficient color contrast”: 5-point Likert scale

According to the methodical approach of Barnum, the evaluators run through typical scenarios of use to conduct their evaluation [20]. They were asked to take the perspective of a diabetes patient aged 50 or older. Each expert tested the main functionalities of the app, listed in the app description (eg, record of blood glucose data and/or medication, plotting graphs, search for information on diabetes mellitus, etc). All Android apps were tested on a Samsung Galaxy Note 10.1. All iOS apps were tested on an iPad 4thgeneration. If apps were offered for multiple platforms by the manufacturer, they were tested on an iPad 4thgeneration.

The chosen method offers a high level of validity and comparability due to its guideline-based approach and closed response categories [22]. Additionally, a user-based test would not have been able to represent the pronounced heterogeneity among the age group 50 or older regarding health status, skills and preferences, experience in technology use, sociodemographics, etc, which is much more pronounced than for younger age groups [13,15,23-25]. At the same time, it would have been rather difficult to find test persons whose characteristics corresponded exactly to the distribution within the basic population.

Table 2. Evaluated usability and assessment criteria for diabetes apps for the elderly.

Main criterion/subcriteria	Description of characteristics	Assessment criteria
Comprehensibility		
Use of understandable semantics		
	Avoidance of foreign language and technical terms Use of generally intelligible symbols and terms If necessary, provision of additional explanations [14,26,27]	5-point Likert scale (1=does not apply at all; 5=does fully apply)
Simple comprehensibility and interpretability of displayed images and depictions		
	Self-explanatory images and depictions, understandable without further support and explanations [12]	5-point Likert scale (1=does not apply at all; 5= does fully apply)
Simple, self-explanatory menu structures		
	Easily understandable and internally consistent menu structures	5-point Likert scale (1=does not apply at all; 5=does fully apply)
	Avoidance of strong hierarchical menu structures and too many functionalities [11,15,28]	
Presentation (Image and Text)		
Sufficient color contrast		
	Clear, distinguishable colors for images and depictions or choice of color-neutral depictions Avoidance of too glaring colors [12,26]	5-point Likert scale (1=does not apply at all; 5=does fully apply)
Large size of operating elements		
	Sufficient size of screen as well as input and output fields [13,27,28]	5-point Likert scale (1=does not apply at all; 5=does fully apply)
Ability to adapt the size of operating elements and displayed images		
	Ability to adapt size of operating elements and displayed images according to individual needs, capabilities, and preferences [14,26]	Dichotomous scale (applicable, not applicable)
Usability		
Instant and easily understandable feedback		
	Instant response to entered data, including easily understandable error messages in case of erroneous data input [15]	5-point Likert scale (1=does not apply at all; 5=does fully apply)
Intuitive usability		
	Ability to use the application without prior knowledge Ease of learning Fast achievement of a first feeling of success [15,29]	5-point Likert scale (1=does not apply at all; 5=does fully apply)
Simple recognition of click-sensitive areas		
	Simple distinction between click-sensitive and non-click-sensitive areas, also without prior knowledge of the features of the touchscreen technology [12]	5-point Likert scale (1=does not apply at all; 5=does fully apply)
General characteristics		
High fault tolerance/efficient fault management		
	Reducing probability of erroneous data input by limiting choice to meaningful values Efficient proofreading mode and/or helpful user feedback, for example, in case of erroneous data input [27,30]	5-point Likert scale (1=does not apply at all; 5=does fully apply)

Main criterion/subcriteria	Description of characteristics	Assessment criteria
Password-protected services		
	Avoidance of registration at online platforms (but partly contrary to data protection regulations) [13]	Dichotomous scale (applicable, not applicable)

Results

Systematic Review

Search and Screening

In total, we examined 656 apps during the review process. As a result, we created three data sets (Multimedia Appendix 1), which separated the currently available diabetes apps into apps available exclusively for the operating system iOS (276 apps), apps available exclusively for the operating system Android (266 apps), and apps available for both operating systems (114 apps).

Annual Development of App Releases

The first diabetes app for iOS (according to Apple App Store as of April 2013) was developed and released on July 17, 2008 (name: Glucose-Charter, developer: e-agent). The first Android diabetes app (according to Google Play Store as of April 2013) followed on November 8, 2009 (name: Body Sugar, developer: Adibu). The number of diabetes apps released annually increased during the last five years, from 6 in 2008 to 267 in 2012. In the first four months of 2013, 149 new diabetes apps were released. The number of apps for Android more than doubled each year (Figure 1); however, this was not by publication date (unavailable in Google Play Store) but rather the date of the last update. More than half of the iOS diabetes apps (50.7%, 140/276) were specially designed for use on the iPhone. Only 87/276 (31.5%) were designed for both iPhone and iPad. Due to a lack of information in the Google Play Store, this subdivision into smartphone and tablet PC apps could not be made for Android apps.

Figure 1. Annual release figures for diabetes apps.

Operating Language

The majority (85.4%, 560/656) of the examined apps were in English, especially the apps running exclusively on an Android operating system, (90.2%, 240/266). Apps with German as operating language were of relatively low number (14.6%, 96/656) (Table 3). Some apps offered the opportunity to choose between several languages after download.

Table 3. Language of available diabetes apps as of April 2013.

		Operating system			
Category	Subcategory	iOS (n=276)	Android (n=266)	iOS and Android (n=114)	Total (n=656)
Language, n (%)					
	English	229 (83.0)	240 (90.2)	91 (79.8)	560 (85.4)
	German	47 (17.0)	26 (9.8)	23 (20.2)	96 (14.6)

Acquisition Costs

The acquisition costs and the ratio of free to paid apps differed strongly between the two operating systems (Table 4). While most of the iOS apps required payment (62.3%, 172/276), the vast majority of Android and Android/iOS apps were free (63.5%, 169/266 and 69.3%, 79/114). Nonetheless, some of the free apps and the inexpensive apps worked with specially designed test strips or were able to be linked to measurement devices. In these cases, the apps could not have been used without compatible devices.

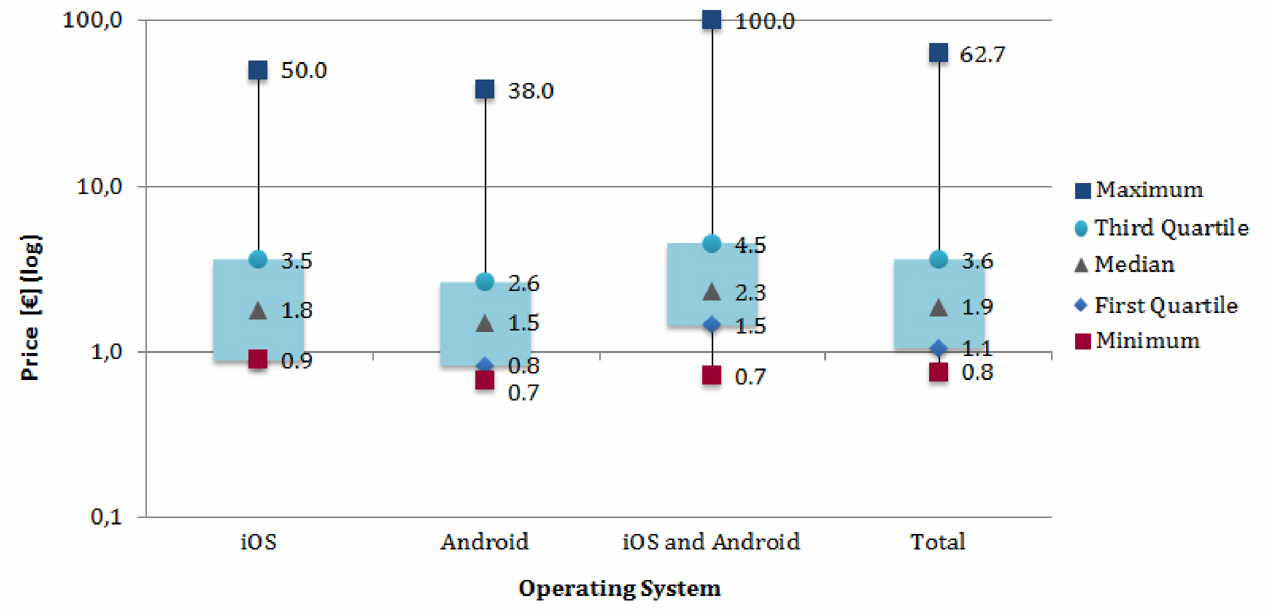
The analysis of app price distribution revealed that a greater number of free apps were available across all apps (53.7%, 352/656). This appeared to be driven by Android apps where 63.5% (169/266) were free compared with 36.5% (97/266) paid. The reverse trend was observed for iOS where only 37.7% (104/276) were free compared with 62.3% (172/276) paid (Table 4).

The price of paid apps differed strongly between the operating systems (Figure 2). The vast majority (69.7%, 212/304), were in the price range of €0.01 to €3.00. The median price varied between €1.50 and €2.30, depending on the operating system. The apps designed for both operating systems tended to be the apps with the highest price level (Figure 2). The analysis also showed that some costly apps offer free or cheaper “lite” versions with limited functionalities (5.3%, 35/656).

Table 4. Price distribution of apps and annual proportions of free apps since 2008.

Category	Subcategory	Operating system			
		iOS (n=276)	Android (n=266)	iOS and Android (n=114)	Total (n=656)
Price distribution of diabetes apps and “lite” versions, n (%)					
	Free	104 (37.7)	169 (63.5)	79 (69.3)	352 (53.7)
	Paid	172 (62.3)	97 (36.5)	35 (30.7)	304 (46.3)
	Paid/Lite version available	18 (6.5)	11 (4.1)	6 (5.3)	35 (5.3)
Development share of free diabetes apps since 2008, n (%)					
	2013 (by April)	6/20 (30.0)	60/87 (69.0)	33/42 (78.6)	99/149 (66.4)
	2012	58/104 (55.8)	79/108 (73.7)	40/55 (72.7)	177/267 (66.3)
	2011	23/71 (32.4)	27/58 (46.6)	6/16 (37.5)	56/145 (38.6)
	2010	13/52 (25.0)	3/12 (25.0)	0/1 (0.0)	16/65 (24.6)
	2009	3/23 (13.0)	0/1 (100.0)	0/0 (0.0)	3/24 (12.5)
	2008	1/6 (16.7)	0/0 (0.0)	0/0 (0.0)	1/6 (16.7)

Figure 2. Price distribution of paid diabetes apps available as of April 2013.



Range of Functions/Functionality

Examining the range of functions of diabetes apps demonstrated that most were limited to one function (54.1%, 355/656). Only 185/656 (28.2%) combined two functions, and three or more functions were offered by 116/656 (17.7%) of the apps available as of April 2013 (Table 5). Apps developed exclusively for iOS tended to offer a wider range of functions compared to Android apps.

A total of 348/656 (53.0%) apps and thus the majority of diabetes apps available as of April 2013 offered a documentation function (Figure 3). By using this feature, the measured blood glucose values can be recorded and either summarized in a table or plotted as a graph. Hence, the app enables the user to monitor the disease progression.

The recording of the blood glucose values mainly occurred via manual data input. Only a small number of apps offered the option to transfer the data wirelessly and automatically from the measuring device via Bluetooth to the mobile device.

The documentation function may be linked with an analysis function, which opens up the possibility to analyze the recorded data and to graphically display the results (Multimedia Appendix 2); 117/656 (17.8%) of the diabetes apps offered this service (Figure 3). The documentation function includes the recording and monitoring of individual eating habits (eg, the bread unit intake). Some apps, additionally, log the frequency of the user's physical activity or the individual medical therapy (type and frequency). The documentation function was frequently linked with a reminder function, which reminds the user of its periodic, pre-defined medication (11.4%, 75/656) (Figure 3). According to the holistic setting approach, some of the available apps already offered the opportunity to track the course of disease for affected family members.

In total, 226 (34.5%) of the examined diabetes apps offered an information function, including the ability to inform about the illness, its diagnosis, the course of the disease, various treatment options, medication, and secondary diseases (Figure 3). Sometimes those apps provided information on the nutrient content of diverse foods and beverages and calorie consumption during various sporting activities (Multimedia Appendix 3).

A data forwarding/communication function was offered by 204/656 (31.1%) apps. With this function, the user has the opportunity to send the recorded data via email to the attending physician, family members, and/or friends (Multimedia Appendix 4). The reports can be sent frequently or on demand. Some of the apps were connected to special diabetes forums, where the users can upload their individual blood glucose values and discuss them with other diabetes patients (name: Diabesties, developer: Ayogo Health).

Surprisingly, only 58/656 (8.8%) of the diabetes apps provided an advisory function or any other kind of therapeutic support (Figure 3). Only a limited number of apps used the recorded data to create individualized advice to optimize the patients measuring, medication, eating habits, or activity behavior. One reason may be a previously required certification as a medical product for that kind of support.

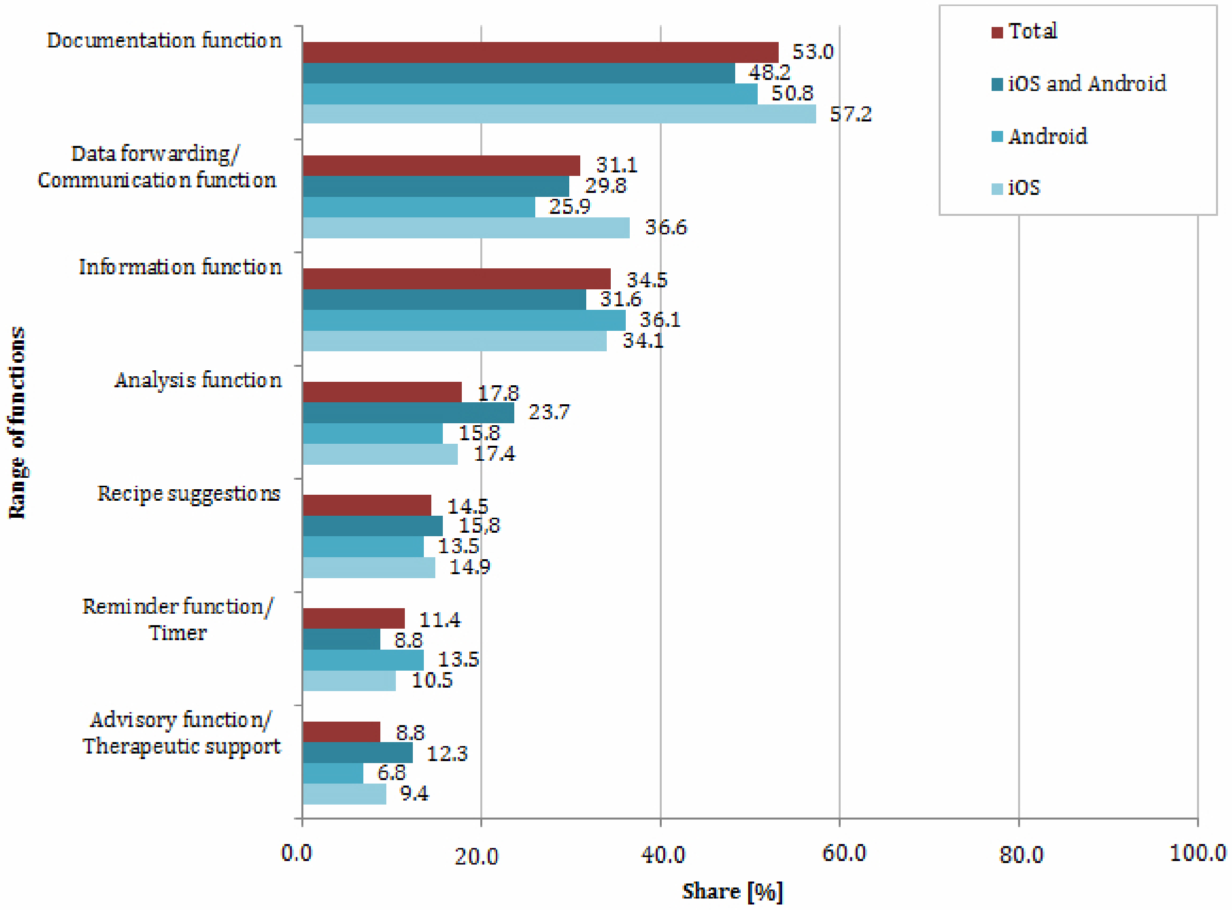
Besides the previously described functions, 95/656 (14.5%) of the apps included suggestions for recipes suitable for the needs of diabetics (Figure 3).

As an example, Multimedia Appendix 5 shows screenshots of a highly reviewed app linking a documentation, analysis, communication, and information function (name: IBG Star Diabetes Manager mg/dl, developer: Sanofi Diabetes).

Table 5. Number of functions, target user groups, and popularity/user ratings of diabetes apps available as of April 2013.

Category	Subcategory	Operating system			
		iOS (n=276)	Android (n=266)	iOS and Android (n=114)	Total (n=656)
Number of functions per diabetes app, n (%)					
	1 function	134 (48.6)	156 (58.6)	65 (57.0)	355 (54.1)
	2 functions	87 (31.5)	71 (26.7)	27 (23.7)	185 (28.2)
	3 functions	36 (13.0)	25 (9.4)	13 (11.4)	74 (11.3)
	4 functions	15 (5.4)	11 (4.1)	9 (7.9)	35 (5.3)
	> 4 functions	4 (1.4)	3 (1.1)	0 (0.0)	7 (1.1)
Target user groups, n (%)					
	Patients	263 (95.3)	260 (97.7)	107 (93.9)	630 (96.0)
	Physicians/qualified health personnel	19 (6.9)	17 (6.4)	14 (12.3)	50 (7.6)
	Patients and physicians/qualified health personnel	6 (2.2)	11 (4.1)	7 (6.1)	24 (3.7)
Popularity/user ratings					
	Share of apps with rating, n (%)	31 (11.2)	189 (71.0)	75 (65.8)	295 (45.0)
	Median number of ratings	9.0	6.0	6.0	7.0
	Median number of stars (max 5)	3.5	4.0	4.0	3.8

Figure 3. Range of functions of diabetes apps available as of April 2013.



Target User Groups

The vast majority (96.0%, 630/656) of the examined apps were designed specifically for patients, 24/656 (3.7%) apps addressed both patients and physicians/qualified health personnel, and only 50/656 (7.6%) were specifically designed for the target group physicians/qualified health personnel (Table 5). One reason might be the high potential for individual treatment and management of diabetes mellitus by the patients themselves. Particularly, patients suffering from type 2 diabetes have the opportunity to impact the course of the disease in a positive manner by a change in their lifestyles [31-33]. In this regard, diabetes mellitus differs from other chronic diseases such as cancer or dementia.

Popularity and User Rating

User ratings are a highly valuable and realistic evaluation of the additional benefits of apps. It is thus remarkable that just 31/656 (11.2%) of the apps designed exclusively for iOS were rated by users. In comparison, 189/266 (71.0%) of the Android apps and 75/114 (65.8%) of the apps running on both operating systems were rated (Table 5). One reason might be the rating procedure of the Apple App Store, which is more complicated than the procedure of the Google Play Store and requires several steps to rate an app. The median of the amount of provided ratings varied between six (iOS & iOS/Android apps) and nine ratings (Android apps). With a maximum of five stars for an app evaluation, the median rating varied from 3.5 (iOS apps) to 4 (Android & iOS/Android apps) stars (Table 5). Thus, 50% of the diabetes apps earned ratings of more than 3.5 to 4 stars, corresponding to a moderate to good rating.

Not only was the lower number of rated iOS apps conspicuous, the median of awarded stars was also lower than for Android apps (Table 5). That was surprising due to the higher access restrictions (peer review-based admission procedure for new apps) by the Apple App Store compared to non-existent restrictions by the Google Play Store.

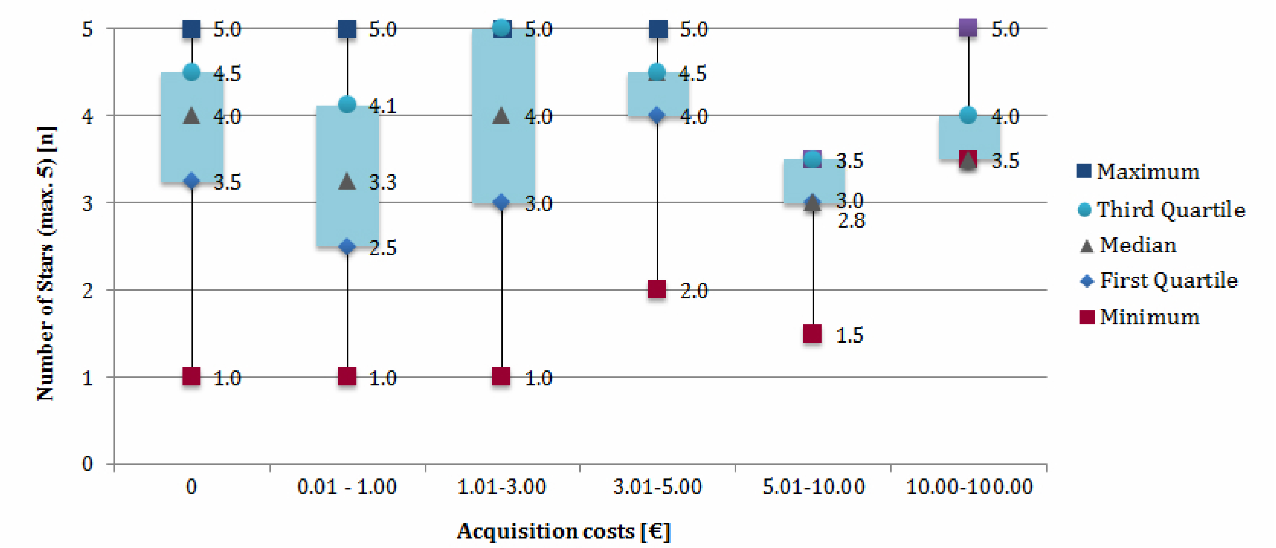
Except for the ratings, the Google Play Store gave information about the number of downloads (ie, the number of installations) as another indicator of the app popularity. This information was not given by the Apple App Store. Hence, it was not possible to compare this indicator between both operating systems. But it has been shown that the number of downloads tended to correlate with the number of ratings and awarded stars.

Connection Between Acquisition Costs and User Ratings

During the analysis, the question arose of whether there is a connection between the price of an app and the level of user ratings. The results indicated that there existed a positive correlation between the acquisition costs and the number of given stars, for the price range of €0.01 to €5.00 (Figure 4). If the price exceeded €5.00, the correlation tended to inverse and the apps received worse evaluations. However, compared to free apps, no clear differences in the number of given stars could be found.

In general, free apps were rated more frequently than paid apps. With a share of 56.5% (204/361), they received the highest number of given ratings compared to just 27.5% (28/102; price range: €0.01-€1.00) up to 41.7% (5/10; price range: €10.00-€100.00) of the paid apps. However, it has to be considered that the number of free apps was considerably higher than the number of price-intensive apps.

Figure 4. Distribution of user rating differentiated by acquisition costs as of April 2013.



Availability of Interfaces to External Sensors/Devices

Contrary to our initial expectations, only a limited number of diabetes apps possessed an interface to an external sensor or a measuring device (eg, for the measurement of blood glucose). Predominantly, apps developed for both operating systems were able to connect with an external sensor/device (7.9%, 9/114). Rarely, iOS apps (2.5%, 7/276) offered this feature compared to Android apps (5.3%, 14/266).

The majority of apps that were able to connect to an external measuring device transmitted the data via a Bluetooth interface. This interface enabled a wireless data transfer to the mobile device or to a PC. Some of the measuring devices already offered an automated transmission of the measured values in real time. There were two options for data synchronization: (1) wireless transfer of measured values to a mobile device and synchronization with the Internet, mostly to an online patient diary (registration required), and (2) wireless transfer of measured values to a PC, transfer of data to an online platform (registration required), and synchronization with a mobile device in the second step (eg, System Health Vault via Microsoft).

Expert-Based Usability Evaluation

In total, we evaluated 66 out of 656 diabetes apps within the usability evaluation (Multimedia Appendix 6): 29 apps available exclusively for the operating system iOS, 28 apps available exclusively for the operating system Android, and 9 apps available for both operating systems.

For all main and subcriteria, we averaged the evaluations of all three experts. The values of the main criteria represent the mean of the corresponding subcriteria (Table 6). The total usability score was calculated from all categories, which were determined by means of a 5-point Likert scale.

Analyzing the results, the majority of evaluations were in the range of 3.0 to 4.0, which corresponded to a moderate to good rating of the apps included in the 10% sample. All tested apps received the best rating for the subcriteria “use of understandable semantics” and “simple comprehensibility and interpretability of displayed images and depictions” with a total average value of 4.1 (Table 6). Independent of the operating system, all apps received the worst rating for the subcriterion “fault tolerance” (2.8) followed by “simple recognizability of click-sensitive areas” (3.0). It has to be mentioned that it was only possible to evaluate fault tolerance for apps offering data input (36.4%, 24/66) (Table 6). The values determined for all other subcriteria varied between 3.1 and 4.0. Comparing the total usability score between the different operating systems, we found no clear differences with values varying between 3.3 and 3.4 for both iOS/Android apps. The worst-rated app with a usability score of 2.5 was HealthFile LIVE! (developer: WakefieldSoft LLC) for Android (Multimedia Appendix 6). The best-rated app was Diabetic Recipes Volume I (developer: ECI) for Android with a usability score of 4.1. Password-protected services were offered by an average of 18.2% of the apps.

In the second run, we evaluated the three chosen accessibility features. The results show that their operability was rather limited. The highest values were observable for the screen reader features Voice over (iOS) and Talkback (Android); 25 (86.2%) of the 29 iOS apps offered the ability to read the screen content aloud compared to 19 (67.9%) of the 28 Android apps, and just 4 (44.4%) of the 9 apps designed for both operating systems. The feature “invert colors” showed no considerable improvement of color contrast compared to the results of our evaluation without testing this feature. The results for testing the feature “large type” differed widely. While none of the iOS apps offered this feature, 11/27 (40.7%) Android apps offered contents in large font (Table 6).

While conducting our systematic review, we hypothesized that usability decreases with an increasing number of functions. Hence, we additionally investigated the relationship between the main usability criteria and the number of functions by conducting several correlation analyses. The results shown in Table 7 indicate statistically significant negative relationships between the number of functions and all usability criteria, except “fault tolerance”. Thus, the number of functions and all usability criteria were significantly negative correlated with coefficients varying between −.29 and −.25 implicating that diabetes apps offering a wider range of functions performed worse in terms of usability.

Furthermore, we analyzed the relationship between the usability score and specific functions based on the differences in functionality we found in our systematic review. Therefore, we conducted multiple linear regression analysis to control for potential confounding effects of other functions offered by the same app (Table 8). It showed significant results only for two types of functions. The presence of a documentation function reduced the usability score on average by 0.36 points while the usability score of apps offering an analysis function was on average reduced by 0.21 points. According to R^2, 25% of the variance of the usability score was explained by the model. All correlation and regression analyses were conducted with the statistical software Stata 11.1.

Table 6. Usability scores from expert-based usability evaluation by operating system, shown as mean values.

Main criterion	Subcriteria	Operating system			
		iOS (n=29)	Android (n=28)	iOS and Android (n=9)	Total (n=66)
		mean (SD)			
Comprehensibility		4.1 (0.53)	4.0 (0.43)	3.7 (0.35)	4.0 (0.48)
	Use of understandable semantics	4.3 (0.58)	4.0 (0.45)	3.8 (0.45)	4.1 (0.54)
	Simple comprehensibility and interpretability of displayed images and depictions	4.2 (0.54)	4.1 (0.53)	4.0 (0.37)	4.1 (0.51)
	Simple, self-explanatory menu structures	3.7 (0.82)	3.9 (0.84)	3.3 (0.66)	3.7 (0.82)
Presentation (Image and Text)		3.4 (0.36)	3.6 (0.38)	3.2 (0.36)	3.5 (0.40)
	Sufficient color contrast	3.5 (0.52)	3.8 (0.47)	3.1 (0.89)	3.6 (0.60)
	Sufficient color contrast with accessibility feature “invert colors”	3.2 (0.65)	3.9 (0.55)	3.4 (0.56)	3.5 (0.68)
	Big size of operating elements	3.4 (0.69)	3.2 (0.57)	3.1 (0.18)	3.3 (0.59)
	Ability to adapt the size of operating elements and displayed images[a], n (%)	8 (27.6%)	4 (14.3%)	2 (22.2%)	14 (21.2%)
	Ability to adapt the size of operating elements and displayed images with accessibility feature “large type”[a], n (%)	0 (0.0%)	11 (40.7%)[b]	3 (37.5%)[b]	14 (21.2%)
Usability		3.4 (0.43)	3.2 (0.44)	3.2 (0.38)	3.3 (0.43)
	Instant and easily understandable feedback	3.3 (0.66)	3.3 (0.53)	3.5 (0.47)	3.3 (0.58)
	Intuitive usability	3.6 (0.68)	3.5 (0.72)	3.3 (0.56)	3.5 (0.68)
	Simple recognition of click-sensitive areas	3.1 (0.65)	2.8 (0.45)	2.9 (0.48)	3.0 (0.55)
	Accessibility Features: Voice over (iOS), Talkback (Android)[a], n (%)	25 (86.2%)	19 (67.9%)	4 (44.4%)	48 (72.7%)
General characteristics		2.5 (0.95)	2.8 (0.87)	3.5 (0.43)	2.8 (0.89)
	Fault tolerance/Efficient fault management	2.5 (0.95)	2.8 (0.87)	3.5 (0.43)	2.8 (0.89)
	Password-protected services[a], n (%)	5 (17.2%)	4 (14.3%)	3 (33.3%)	12 (18.2%)
Number of functions per app		1.6 (0.82)	1.7 (0.85)	1.6 (1.13)	1.7 (0.89)
Total Usability Score		3.3 (0.40)	3.3 (0.38)	3.4 (0.48)	3.3 (0.39)

[a] The values of this subcriterion show means of frequencies.

[b] One observation was missing for this subcriterion and the corresponding operating system. Accordingly n is reduced by 1.

Table 7. Spearman's rank correlation coefficients comparing number of functions with main usability criteria scores.

Number of functions	Main usability criteria scores			
	Comprehensibility	Presentation	Usability	Fault tolerance
1	−.29* (P=.02)	−.25* (P=.046)	−.25* (P=.04)	.46** (P<.001)

*5% significance level

**1% significance level

Table 8. Multiple regression analysis: relationship between usability score and functions.[a]

Variable	Coefficient (b)	95% CI	t	P
Information function	−.11	−0.29 to 0.07	−1.23	.22
Recipe suggestions	.06	−0.15 to 0.27	0.58	.56
Documentation function	−.36	−0.57 to −0.15	−3.43	.001[b]
Analysis function	−.21	−0.39 to −0.02	−2.23	.03[c]
Reminder function/timer	−.04	−0.42 to 0.33	−0.23	.82
Advisory function/therapeutic support	−.12	−0.38 to 0.14	−0.90	.37
Data forwarding/communication function	.04	−0.20 to 0.27	0.31	.76
Intercept	3.72	3.53 to 3.91	38.97	<.001[b]
	n=66	$F_{7,58}$=3.46	R^2=.25	

[a] Ordinary Least Squares regression with robust standard errors

[b] 1% significance level

[c] 5% significance level

Discussion

Systematic Review

The systematic review showed that a large number of diabetes apps are available. Providers may be entering the market as a result of the rising number of patients suffering from diabetes. For users, especially patients, it becomes increasingly difficult to find an app in this plethora of options that is suitable for one's own needs. This problem is caused by a lack of effective search criteria and filter functions in the app stores. More frequently, apps are chosen that appear first in the search results for diabetes apps. The sorting criteria in the app stores are not apparent. New apps from relatively unknown developers could have difficulties being listed among the first results.

At the same time, many apps offered similar functionalities, mostly a documentation function, which is consistent with earlier findings of Martínez-Pérez et al [3], Eng et al [5], Chomutare et al [6], and Demidowich et al [16]. Differences were found mostly in the design and the menu structure. Additionally, the majority of diabetes apps offered only one or two functions. An application that simultaneously informs and contributes to successful treatment by combining documentation, reminder, and advisory functions was not available as of April 2013. Such a multifunctional app would have a clear additional benefit, especially for newly diagnosed and elderly diabetes patients. At the same time, simple, understandable design, content, and menu navigation are needed. But several apps showed a lack of suitability and usability for its main target group diabetics, which is in accordance with the findings of Demidowich et al [16]. Some were apparently developed without intensive cooperation or prior (usability) tests with patients or professional health care personnel. The obligation for certification as a medical product does currently not exist, even though some diabetes apps are already certified, especially those linked to an external measurement device, eg, iBGStar Diabetes Manager (iOS), Bodytel Mobile (iOS and Android), or Diabetes Companion by mySugr (iOS and Android). Peer review processes of health apps are already offered by several platforms such as iMedical Apps [34], JMIR mHealth [35], or HealthOn [36]. This structure offers substantial and valuable support for users and their decision-making processes, but also in terms of quality assurance and improvement. In contrast, the Google Play Store, as one of the leading app stores, does not currently apply a peer review-based admission procedure for new apps. This lack of certification results in a lack of "[…] demonstrated safety and effectiveness, especially where information and trends are not just presented to patients, but used to make treatment recommendations" [37].

As an example, one app mainly offered labels like "after breakfast" or "after lunch" for the documentation of measured blood glucose values, which implies postprandial states. But, for most diabetics, the blood glucose values *before* eating are decisive to adjust the amount of insulin. Another app offered the feature to plot a graph labeled "HbA1c (glycated hemoglobin) according to day-time". This neglects that the HbA1c is a value for long-term blood glucose. As a further example, some apps provided no option to modify, once entered, values at a later point in time. This misconception prevents a subsequent data amendment for example in the run-up of a doctors’ visit. At the same time, there is no opportunity to correct wrong values, for instance, as a result of erroneous data input.

Taking a look into the future, we expect that the data forwarding function, especially to the attending physician, will gain significantly more importance. A regular transmission of data to their physician linked with frequent feedback can be a valuable therapy support, particularly for people in rural regions that are or will become affected by a shortage of doctors [38,39]. Nevertheless, there still exist open questions concerning data security, network coverage, interoperability, documentation requirements, and coverage in health care plans, etc [38,40,41].

Additionally, the automated transmission of measured values in real time from the measuring device to the mobile device will probably spread and is an important driver for the perceived ease of use as El-Gayar et al point out [42]. At the moment, the input of the measured blood glucose values occurs manually in most of the cases, as earlier findings of Eng et al [5] and Chomutare et al [6] confirm. Some manufacturers already offer blood glucose meters that allow real-time data transmission of measured blood glucose data via Bluetooth to a mobile device (eg, GlucoTel [43], iBGStar [44]) (Figure 5). This function simplifies the process of documenting for the patient and, at the same time, it increases the reliability of the entered data and subsequent analysis.

Notwithstanding the functions offered by diabetes apps, their effects on patients’ self-management and, accordingly, on important indicators, as for example the HbA1c value, have to be evaluated. A comprehensive, representative, and long-term study investigating these health effects is lacking so far. But different studies focusing on the outcomes of mobile phone interventions, such as SMS, point out a slightly positive influence as shown in the reviews of Holtz et al [45] and Free et al [46].

Figure 5. Glucose meters with automated transmission of blood glucose values to mobile devices.

Source: [43, 44]

Expert-Based Usability Evaluation

As a supplement to our systematic review, we conducted an expert-based usability evaluation to examine the usability of currently available diabetes apps for patients aged 50 or older. Therefore, we focused on the age group with the highest diabetes prevalence. The results show moderate to good evaluations (range 3.0-4.0) for all reviewed usability criteria, which is in accordance with the results of Demidovich et al [16]. An exception was “fault tolerance” (Table 6). The main criteria, “comprehensibility”, rated best with a score of 4.0. In particular, the elderly benefit from easy, understandable semantics and easy, comprehensible, and interpretable images and depictions, due to their frequently limited experience in handling mobile devices and apps. Hence, it can lower inhibition thresholds, especially during the first time of use, and therefore increase acceptance among this age group. The same is true for the influence of “easily understandable feedback” (3.3) and an “intuitive usability” (3.5) (main criterion “usability”). However, these two subcriteria performed worse within our evaluation. The fact that most of the diabetes apps were in English or contained English/foreign language terms (Table 3) diminished the usability especially for non-English-speaking elderly in terms of comprehensibility. This can be seen as one optimization approach for future app development that is comparatively easy to implement.

The evaluation showed moderate results for the main criterion “presentation” (3.5). Our test of three accessibility features indicated a very good operability of the screen readers, especially for Voice over (86.2%) offered by iOS. However, the operability of the features “invert colors” (3.5) and “large type” (21.2%) was rather restricted. Additionally, the minority of diabetes apps (17.8% of the iOS apps) were developed specifically for tablet PCs. However, we assess them as more suitable and user-friendly for elderly diabetes patients due to their larger display and bigger illustrations. With increasing age, cognitive and physical skills are declining, such as eyesight, visual acuity, color vision, contrast detection, and hearing [11]. Especially elderly diabetes patients are often suffering from retinopathy. Shortcomings of diabetes apps concerning the presentation of information (color contrast, size of operating elements, option to flexibly adapt size of operating elements, etc) and the operability of accessibility features are deterring potentially interested patients from using diabetes apps from the outset. Therefore, a barrier-free access is a basic prerequisite for elderly patients to make them use diabetes apps. Against this background, all the other criteria we determined become of secondary importance in terms of usability.

The criterion "fault tolerance" rated worst with a score of 2.8 (Table 6). This means that the available diabetes apps were lacking an efficient fault management (criterion specified in Table 2). Especially inexperienced (elderly) users often have difficulties with inputting data. Some errors are unrecoverable or even cause the application to shut down, as Garcia et al already demonstrated in their analysis [17]. These results have to be viewed with great concern due to the fact that these apps are dealing with medical parameters. This becomes particularly serious if these values provide the basis for further calculations as, for example, the required dose of insulin. Our results demonstrate once again the meaningfulness of an automated transfer of measured values from the blood glucose meter to the mobile device [42]. Additionally, it could be helpful to limit choice to meaningful values, for example, by offering a numeric keyboard to enter blood glucose values.

Our correlation and regression analyses indicated a strong link between usability and the number and kind of functions. In particular, the number of functions and all main usability criteria were significantly negative correlated. These results cast a different light on the aforementioned outcomes of our usability test. Hence, the moderate to good usability scores applied mainly to apps offering a small range of functions. This relation inverted when we looked upon the considerably lower usability scores for multifunctional apps (Multimedia Appendix 6). Considering the special needs of elderly diabetes patients, they would benefit from a comprehensive and easily understandable support as already mentioned above. They are frequently affected by multimorbidity and polypharmacy, particularly after many years of suffering from diabetes [47].

Differed by functions, apps offering a documentation and an analysis function performed worse in terms of usability. This result is surprising as the documentation function is most commonly offered with a share of 53.0%. It can be a valuable support for all diabetes patients measuring and recording blood glucose level regularly. But as interviews with diabetes patients aged 50 or older (conducted in Germany in 2013) have shown, most of them prefer documentation by means of a conventional diary (results not shown). One reason they named was the aforementioned lack of usability and therefore a too complicated and time-consuming handling. Moreover, the use of these two functions is characterized by a higher level of human-technology interaction than, for example, the use of an information function. Of course, this can be accompanied by a wider scope of error sources and usability barriers.

Altogether, the potential of diabetes apps for assisting and supporting diabetes patients aged 50 or older is large. In particular, the target group aged between 50 and 60 years holds great potential as people of this age are already quite familiar with mobile devices and apps [48]. Now, app developers are facing the challenge of taking sufficient account of the usability criteria we examined and addressing those shortcomings. There is no need for a huge number of new app functions. It is more about improving what already exists.

Future Work

The systematic review and the expert-based usability test were conducted within the project "InnoMedTec". In that project, we investigate the question: "How should a mobile application be designed to support an effective self-management for diabetes patients aged 50 or older?" Our market analysis provided the basis for a survey among diabetes patients aged 50 or older and physicians, which we conducted in the second half of 2013. Within guided interviews, we investigated the current use, acceptance promoting/inhibiting factors, potentially needed support, and concrete design features for the development of a diabetes app. Merging the results of the systematic market review and the survey, a user- and needs-oriented prototype app for diabetics aged 50 or older will be developed this year. To guarantee usability and needs orientation, the prospective users and usability experts are involved in the product development process right from the beginning. User- and expert-based usability tests are performed regularly. The results are integrated continuously in the app optimization until its finalization.

Limitations

Systematic Review

The conducted review focused exclusively on apps for the currently leading operating systems, iOS and Android. Currently available diabetes apps for other operating systems, such as Windows Phone, Blackberry OS, or Symbian, were not considered within the analysis. The app publication date was solely available for iOS apps, but not for Android apps. Here, the date of the last update served as reference value. Due to that fact, the results concerning the annually new released diabetes apps were not directly comparable.

The app information was gathered by studying the descriptions in the app stores and within the app itself. More detailed information, such as download statistics, were not available for analysis. Perhaps this information would enable more detailed results concerning the user groups, for example, differentiated by gender, age groups, or type of diabetes.

Expert-Based Usability Evaluation

Within our usability evaluation, we investigated usability criteria exclusively. We evaluated neither the quality of content and functions nor their effectiveness. Furthermore, it has to be mentioned that one usability evaluation cannot claim to cover all possible and critical usage situations that can possibly occur [18-20,49].

We would also stress that we examined a sample of all available diabetes apps, not just a sample of apps developed specifically for the elderly. Hence, many of the apps we evaluated do not claim to be particularly suitable for this age group.

Conclusions

Despite the huge amount of currently available diabetes apps, most of them offer a small number of similar functionalities. Patients and physicians should be directly involved during the app development to tackle the lack of usability and needs-orientation for its main target group diabetics. We think that data forwarding options and automated transmission of measured values to mobile devices will gain more importance in the future.

The usability of diabetes apps for patients aged 50 or older was moderate to good. But this result applied mainly to apps offering a small range of functions. Multifunctional apps performed considerably worse in terms of usability. Differed by functions, the documentation and analysis function indicated significantly lower usability scores. The operability of accessibility features for diabetes apps was quite limited, except for the feature "screen reader".

Acknowledgments

This study is conducted within the InnoMedTec project (Grant no. 100098212), which is funded by the European Social Fund and the Free State of Saxony. We would like to thank Dr Lars Renner, Robert Gurke, Andrea Mathe, Tiffany Evans, and Maike Bellmann for their feedback in the writing process and careful reading. We also would like to express our thanks to our reviewers for their valuable feedback and helpful suggestions that helped improve our paper. Findings from this study have been previously shown in a poster presentation at the Diabetes Congress 2013, Leipzig (Germany), and as an oral presentation at the Telemed 2013, Berlin (Germany).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Dataset diabetes app review.

[XLSX File (Microsoft Excel File), 157KB - jmir_v16i4e104_app1.xlsx]

Multimedia Appendix 2

Screenshots of apps with documentation and analysis function.

[PNG File, 2MB - jmir_v16i4e104_app2.png]

Multimedia Appendix 3

Screenshots of apps with information function.

[PNG File, 1MB - jmir_v16i4e104_app3.png]

Multimedia Appendix 4

Screenshots of apps with communication function.

[PNG File, 568KB - jmir_v16i4e104_app4.png]

Multimedia Appendix 5

Screenshots of a highly reviewed app with various functionalities.

[PNG File, 1MB - jmir_v16i4e104_app5.png]

Multimedia Appendix 6

Usability scores differed by operating systems and apps.

[XLSX File (Microsoft Excel File), 12KB - jmir_v16i4e104_app6.xlsx]

References

1. Ingraham N. Apple: 900,000 apps in the App Store, 375,000 iPad-optimized, 28 million copies of mountain lion sold. 2013. URL: http://www.theverge.com/2013/6/10/4412918/apple-stats-update-wwdc-2013 [accessed 2013-08-30] [WebCite Cache ID 6JG1ngi8P]

2. MHealth Watch. Mobile health care apps growing fast in number. 2013. URL: http://mhealthwatch.com/mobile-health-care-apps-growing-fast-in-number-20052/ [accessed 2013-09-19] [WebCite Cache ID 6JkCsxMq0]
3. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013;15(6):e120 [FREE Full text] [doi: 10.2196/jmir.2600] [Medline: 23770578]
4. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
5. Eng DS, Lee JM. The promise and peril of mobile health applications for diabetes and endocrinology. Pediatr Diabetes 2013 Jun;14(4):231-238. [doi: 10.1111/pedi.12034] [Medline: 23627878]
6. Chomutare T, Fernandez-Luque L, Arsand E, Hartvigsen G. Features of mobile diabetes applications: review of the literature and analysis of current applications compared against evidence-based guidelines. J Med Internet Res 2011;13(3):e65 [FREE Full text] [doi: 10.2196/jmir.1874] [Medline: 21979293]
7. Lehnert H, Wittchen HU, Pittrow D, Bramlage P, Kirch W, Böhler S, et al. Prevalence and pharmacotherapy of diabetes mellitus in primary care. Dtsch Med Wochenschr 2005 Feb 18;130(7):323-328. [doi: 10.1055/s-2005-863050] [Medline: 15712019]
8. Kirch W. Encyclopedia of Public Health: Volume 1: A - H Volume 2: I - Z. New York: Springer; 2008.
9. Du Y, Heidemann C, Scheidt-Nave C, Robert-Koch-Institut. Diabetes mellitus in Deutschland. 2011. URL: http://www.rki.de/DE/Content/Gesundheitsmonitoring/Gesundheitsberichterstattung/GBEDownloadsK/2011_3_diabetes.pdf?__blob=publicationFile [accessed 2013-09-02] [WebCite Cache ID 6JKYGRUrU]
10. International Diabetes Federation. Diabetes Atlas. 2013. URL: http://www.idf.org/diabetesatlas [accessed 2013-08-30] [WebCite Cache ID 6JG2EAtt4]
11. Holzinger A, Searle G, Nischelwitzer A. On some aspects of improving mobile applications for the elderly. In: Stephanidis C, editor. UAHCI'07 Proceedings of the 4th International Conference on Universal Acess in Human Computer Interaction. Coping with Diversity. Berlin, Heidelberg: Springer; 2007:923-932.
12. Lorenz A, Oppermann R, Zahl L, Mielke D. Personalized mobile health monitoring for elderly. In: Cheok AD, Chittaro L, editors. MobileHCI ’07 Proceedings of the 9th International Conference on Human Computer Interaction with Mobile Devices and Services. New York: ACM; 2007:297-304.
13. Mallenius S, Rossi M, Tuunainen VK. Factors affecting the adoption and use of mobile devices and services by elderly people – results from a pilot study. 2010. URL: http://classic.marshall.usc.edu/assets/025/7535.pdf [WebCite Cache ID 6Ofn2S0Ev]
14. Lorenz A, Oppermann R. Mobile health monitoring for the elderly: Designing for diversity. Pervasive and Mobile Computing 2009 Oct;5(5):478-495. [doi: 10.1016/j.pmcj.2008.09.010]
15. Steele R, Lo A, Secombe C, Wong YK. Elderly persons' perception and acceptance of using wireless sensor networks to assist healthcare. Int J Med Inform 2009 Dec;78(12):788-801. [doi: 10.1016/j.ijmedinf.2009.08.001] [Medline: 19717335]
16. Demidowich AP, Lu K, Tamler R, Bloomgarden Z. An evaluation of diabetes self-management applications for Android smartphones. J Telemed Telecare 2012 Jun;18(4):235-238. [doi: 10.1258/jtt.2012.111002] [Medline: 22604278]
17. Garcia E, Martin C, Garcia A, Harrison R, Flood D. Systematic analysis of mobile diabetes management applications on different platforms. In: Holzinger A, Simonic KM, editors. Information Quality in e-Health. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2011:379-396.
18. Sarodnick F, Brau H. Methoden der Usability Evaluation: Wissenschaftliche Grundlagen und praktische Anwendungen. 2nd ed. Bern: Verlag Hans Huber; 2011.
19. Nielsen J. Usability engineering. San Francisco, Calif: Morgan Kaufmann; 1993.
20. Barnum CM. Usability testing essentials: ready, set...test!. San Francisco, Calif: Morgan Kaufmann; 2011.
21. Gondy L. Designing user studies in informatics. London: Springer; 2011.
22. Abascal J, Arrue M, Fajardo I, Garay N. An expert-based usability evaluation of the EvalAccess web service. In: Navarro-Prieto R, Vidal JL, editors. HCI related papers of Interacción 2004. Dordrecht: Springer; 2006:1-18.
23. Lee YS. Older adults’ user experience with mobile phones: identification of user clusters and user requirements. 2007. URL: http://scholar.lib.vt.edu/theses/available/etd-09172007-135013/unrestricted/Dissertation_YSL.pdf [accessed 2014-02-07] [WebCite Cache ID 6ND7upoqj]
24. Deutscher Bundestag. Sechster Bericht zur Lage der älteren Generation in der Bundesrepublik Deutschland - Altersbilder in der Gesellschaft. 2010. URL: http://www.bmfsfj.de/RedaktionBMFSFJ/Abteilung3/Pdf-Anlagen/bt-drucksache-sechster-altenbericht,property=pdf,bereich=bmfsfj,sprache=de,rwb=true.pdf [accessed 2014-02-07] [WebCite Cache ID 6ND865JKN]
25. Caprani N, O'Connor NE, Gurrin C. Touch screens for the older user. 2012. URL: http://www.intechopen.com/books/assistive-technologies/touch-screens-for-the-older-user [accessed 2014-02-07] [WebCite Cache ID 6ND8TYl9d]
26. DIN-Fachbericht 131:2003. Leitlinien für Normungsgremien zur Berücksichtigung der Bedürfnisse von älteren Menschen und von Menschen mit Behinderungen; Deutsche und englische Fassung des CEN/CENELEC-Leitfadens 6. Berlin: Beuth Verlag; 2003.

27. Schmid A, Dörfler I, Dany F, Böpple O. Analyse der Akzeptanzkriterien für mobile Anwendungen im Bereich Gesundheit in der Zielgruppe 50+. In: Shire KA, Leimeister JM, editors. Technologiegestützte Dienstleistungsinnovation der Gesundheitswirtschaft. Wiesbaden: Springer Gabler; 2012:57-82.
28. Renaud K, van Biljon J. Predicting technology acceptance and adoption by the elderly : a qualitative study. In: Navarro-Prieto R, Vidal JL, editors. SAICSIT '08 Proceedings of the 2008 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries: Riding the Wave of Technology. New York: ACM; 2008:210-219.
29. Banse M. Softwareergonomische Optimierung Touchscreen basierter Mensch-Computer-Interaktion. 2008. URL: http://touchscreen-usability.com/ [accessed 2014-02-07] [WebCite Cache ID 6ND4VFGE0]
30. DIN EN ISO 9241-110:2008-09. Ergonomie der Mensch-System-Interaktion - Teil 110: Grundsätze der Dialoggestaltung. Deutsche Fassung EN ISO 9241-110:2006. Berlin: Beuth Verlag; 2008.
31. Tuomilehto J, Lindström J, Eriksson JG, Valle TT, Hämäläinen H, Ilanne-Parikka P, Finnish Diabetes Prevention Study Group. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. N Engl J Med 2001 May 3;344(18):1343-1350. [doi: 10.1056/NEJM200105033441801] [Medline: 11333990]
32. Diabetes Prevention Program (DPP) Research Group. The Diabetes Prevention Program (DPP): description of lifestyle intervention. Diabetes Care 2002 Dec;25(12):2165-2171 [FREE Full text] [Medline: 12453955]
33. Diabetes Prevention Program (DPP) Research Group, Knowler WC, Fowler SE, Hamman RF, Christophi CA, Hoffman HJ, et al. 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. Lancet 2009 Nov 14;374(9702):1677-1686 [FREE Full text] [doi: 10.1016/S0140-6736(09)61457-4] [Medline: 19878986]
34. iMedical Apps. 2013. URL: http://www.imedicalapps.com/ [accessed 2013-11-11] [WebCite Cache ID 6L2vC594j]
35. JMIR mHealth. 2013. URL: https://docs.google.com/spreadsheet/viewform?formkey=dEY0YThtT2lXTXVUQ2VuUHRzSkh5a2c6MQ [WebCite Cache ID 6LT3XC5j633]
36. HealthOn. 2013. URL: http://www.healthon.de/ [accessed 2013-12-10] [WebCite Cache ID 6Ll1QJft6]
37. Klonoff DC. The current status of mHealth for diabetes: will it be the next big thing? J Diabetes Sci Technol 2013 May;7(3):749-758. [Medline: 23759409]
38. Nuffield Council on Bioethics. Medical profiling and online medicine: the ethics of 'personalised healthcare' in a consumer age. London: Nuffield Council on Bioethics; 2010.
39. Ascher M. Telemedicine in diabetes care. 2011. URL: http://next.eller.arizona.edu/courses/BusinessInternationalEnvironments/Fall2011/student_papers/finalmelissaascher.pdf [accessed 2013-09-02] [WebCite Cache ID 6JKQ6Zs1m]
40. VDE Initiative Mikromedizin. VDE-Positionspapier Telemonitoring zur Prävention von Diabetes-Erkrankungen. 2008. URL: https://www.vde.com/de/InfoCenter/Seiten/Details.aspx?eslShopItemID=998e16ea-eb11-42de-8f42-65e672c275dd [accessed 2014-04-01] [WebCite Cache ID 6OVXplBtj]
41. World Health Organization. Telemedicine: opportunities and developments in member states: report on the Second Global Survey on eHealth 2009. Geneva: World Health Organization; 2010.
42. El-Gayar O, Timsina P, Nawar N, Eid W. Mobile applications for diabetes self-management: status and potential. J Diabetes Sci Technol 2013;7(1):247-262 [FREE Full text] [Medline: 23439183]
43. BodyTel Europe GmbH. Glucotel. 2013. URL: http://www.bodytel.com/glucotel.html [accessed 2013-08-30] [WebCite Cache ID 6JG6pfUjY]
44. Sanofi Diabetes. iBGStar. 2013. URL: http://www.ibgstar-shop.de/ [accessed 2013-08-30] [WebCite Cache ID 6JG79fEsf]
45. Holtz B, Lauckner C. Diabetes management via mobile phones: a systematic review. Telemed J E Health 2012 Apr;18(3):175-184. [doi: 10.1089/tmj.2011.0119] [Medline: 22356525]
46. Free C, Phillips G, Felix L, Galli L, Patel V, Edwards P. The effectiveness of M-health technologies for improving health and health services: a systematic review protocol. BMC Res Notes 2010;3:250 [FREE Full text] [doi: 10.1186/1756-0500-3-250] [Medline: 20925916]
47. Wittchen HU, Group DS. Diabetes in Deutschland. DETECT: Eine bundesweite Studie in 3.500 Hausarztpraxen und 55.000 Patienten. 2005. URL: http://www.detect-studie.de/publikationen/DETECT_EXUBERA.pdf [accessed 2014-02-07] [WebCite Cache ID 6ND9AJjEI]
48. Online Publishers Association. A portrait of today’s smartphone user. 2012. URL: http://onlinepubs.ehclients.com/images/pdf/MMF-OPA_--_Portrait_of_Smartphone_User_--_Aug12_%28Public%29.pdf [accessed 2014-02-07] [WebCite Cache ID 6ND9K5a67]
49. Nielsen J. Usability inspection. In: Plaisant C, editor. CHI ’94 Conference Companion, Human Factors in Computing Systems, Celebrating Interdependence. New York, NY: ACM Press; 1994:413-414.

Abbreviations

Apps: mobile applications
HbA1c: glycated hemoglobin

Edited by G Eysenbach; submitted 19.09.13; peer-reviewed by C Lauckner, B Martínez-Pérez, T Chomutare; comments to author 30.10.13; revised version received 24.02.14; accepted 13.03.14; published 09.04.14.

Please cite as:
Arnhold M, Quade M, Kirch W
Mobile Applications for Diabetics: A Systematic Review and Expert-Based Usability Evaluation Considering the Special Requirements of Diabetes Patients Age 50 Years or Older
J Med Internet Res 2014;16(4):e104
URL: http://www.jmir.org/2014/4/e104/
doi:10.2196/jmir.2968
PMID:24718852

Original Paper

Factors Related to Sustained Use of a Free Mobile App for Dietary Self-Monitoring With Photography and Peer Feedback: Retrospective Cohort Study

Elina Helander[1], PhD; Kirsikka Kaipainen[2], MSc; Ilkka Korhonen[1,2], PhD; Brian Wansink[3], PhD

[1]Department of Signal Processing, Tampere University of Technology, Tampere, Finland

[2]VTT Technical Research Centre of Finland, Tampere, Finland

[3]Charles H. Dyson School of Applied Economics and Management, Cornell University, Ithaca, NY, United States

Corresponding Author:
Elina Helander, PhD
Department of Signal Processing
Tampere University of Technology
Korkeakoulunkatu 1
Tampere,
Finland
Phone: 358 4317579
Fax: 358 33641352
Email: elina.helander@tut.fi

Abstract

Background: Healthy eating interventions that use behavior change techniques such as self-monitoring and feedback have been associated with stronger effects. Mobile apps can make dietary self-monitoring easy with photography and potentially reach huge populations.

Objective: The aim of the study was to assess the factors related to sustained use of a free mobile app (“The Eatery”) that promotes healthy eating through photographic dietary self-monitoring and peer feedback.

Methods: A retrospective analysis was conducted on the sample of 189,770 people who had downloaded the app and used it at least once between October 2011 and April 2012. Adherence was defined based on frequency and duration of self-monitoring. People who had taken more than one picture were classified as “Users” and people with one or no pictures as “Dropouts”. Users who had taken at least 10 pictures and used the app for at least one week were classified as “Actives”, Users with 2-9 pictures as “Semi-actives”, and Dropouts with one picture as “Non-actives”. The associations between adherence, registration time, dietary preferences, and peer feedback were examined. Changes in healthiness ratings over time were analyzed among Actives.

Results: Overall adherence was low—only 2.58% (4895/189,770) used the app actively. The day of week and time of day the app was initially used was associated with adherence, where 20.28% (5237/25,820) of Users had started using the app during the daytime on weekdays, in comparison to 15.34% (24,718/161,113) of Dropouts. Users with strict diets were more likely to be Active (14.31%, 900/6291) than those who had not defined any diet (3.99%, 742/18,590), said they ate everything (9.47%, 3040/32,090), or reported some other diet (11.85%, 213/1798) (χ^2_3=826.6, P<.001). The average healthiness rating from peers for the first picture was higher for Active users (0.55) than for Semi-actives (0.52) or Non-actives (0.49) ($F_{2,58167}$=225.9, P<.001). Actives wrote more often a textual description for the first picture than Semi-actives or Non-actives (χ^2_2=3515.1, P<.001). Feedback beyond ratings was relatively infrequent: 3.83% (15,247/398,228) of pictures received comments and 15.39% (61,299/398,228) received “likes” from other users. Actives were more likely to have at least one comment or one “like” for their pictures than Semi-actives or Non-actives (χ^2_2=343.6, P<.001, and χ^2_2=909.6, P<.001, respectively). Only 9.89% (481/4863) of Active users had a positive trend in their average healthiness ratings.

Conclusions: Most people who tried out this free mobile app for dietary self-monitoring did not continue using it actively and those who did may already have been healthy eaters. Hence, the societal impact of such apps may remain small if they fail to reach those who would be most in need of dietary changes. Incorporating additional self-regulation techniques such as goal-setting and intention formation into the app could potentially increase user engagement and promote sustained use.

(J Med Internet Res 2014;16(4):e109) doi:10.2196/jmir.3084

KEYWORDS

healthy eating; food journaling; food photographing; adherence; mobile app; self-monitoring; peer feedback; control theory

Introduction

Dietary Self-Monitoring and Feedback

Despite various efforts to curb the growth of obesity, a significant part of the population still eats unhealthy food in excessive quantities. Knowledge about healthy eating is not sufficient on its own to change eating behavior [1]. On an individual level, one behavioral strategy recommended in weight control and improvement of dietary habits is self-monitoring of food intake [2,3]. Moreover, healthy eating interventions that use self-monitoring combined with other self-regulation techniques from control theory, such as feedback [4], appear to be more effective than interventions that do not include these techniques [5]. Smartphones and photography can be used to make self-monitoring easy and convenient [6-9]. Due to the wide penetration of smartphones in the population, this approach could reach a large number of people with small cost. However, it is not known whether such mobile apps for independent use would engage people and reach those who could benefit the most from dietary monitoring.

Dietary self-monitoring prompts people to reflect on their current behavior and compare it to ideal behavior [4,10]. In weight loss studies, consistent recording of food intake appears to be one of the most effective methods [3,11]. Yet it is not clear how consistent self-monitoring needs to be for the method to be effective. The degree of monitoring is typically reported as the number of food diaries/entries completed per day and/or week, and the duration of the monitoring period has varied from eight weeks to two years in different studies [3]. For example, a six-month intervention study found that the average number of food records per week was 3.7 and greater weight loss was associated with more frequent monitoring [12]. Another study on members of a free online weight loss program found that more frequent weight monitoring was associated with greater weight loss, but no association was found between dietary monitoring frequency and weight loss [13]. Overall, it is not clear whether these results tell more about an individual's engagement to the program or specifically about the effect of monitoring. Studies have also mostly focused on weight loss, not on prevention of weight gain through improvement of eating behavior.

Traditional methods for dietary self-monitoring include more or less detailed food diaries and calorie counting [14]. These methods can be burdensome to people [6,7,15-17] and suffer from underreporting and recall issues [15-19]. Methods that can minimize the temporal distance between eating and recording food intake are likely to improve outcomes; the percentage of food records made within 15 minutes of eating has been found to be associated with weight loss [20]. Recently, smartphone cameras have made just-in-time food journaling possible by taking a photo of food. A pilot study using disposable cameras suggests that recording food before eating can lead to increased consideration of dietary habits and alter food choices better than written diaries [21]. Capturing images of food may improve adherence and accuracy in some groups, such as among adolescents who may be less motivated to keep detailed food diaries [8].

Feedback on performance is a self-regulation technique that either reinforces the current behavior or creates a discrepancy between current and ideal behavior [4]. Individualized feedback has been found to be associated with higher adherence to online interventions promoting healthy lifestyles [22,23]. In terms of the content of feedback, encouraging reflection and self-monitoring may be more important than detailed analysis of nutritional contents when the target is to change eating behavior [10,24,25].

Mobile apps can provide automated feedback on the healthiness of the food based on the photo and also leverage other users to provide feedback through crowdsourcing [6]. Although it is still difficult to estimate food ingredients and portion sizes from a photograph, efforts to develop estimation algorithms based on image processing or crowdsourcing are underway [6,16]. One such app is PlateMate, which crowdsources nutritional assessments from Amazon Mechanical Turk, where individuals assessing the food pictures receive a nominal payment for each picture [6]. Evaluation of the app suggested that these crowd-generated assessments were almost as accurate as those done by professional nutritionists, although pictures containing ambiguous items such as beverages or salad dressing received inaccurate ratings [26]. Beyond nutritional assessments, technology can be used to share advice and feedback on healthy eating between users [27].

Adherence to Mobile Apps

Numerous mobile apps for healthy eating are available in application markets. Although they are easily within anyone's reach, attrition is likely to be a significant challenge, since there is usually very little external pressure or incentive to continue usage [28]. Little research exists about usage behavior of health-promoting apps, but reviews on Web-based interventions have found that adherence is generally lower outside randomized controlled trials and some observational studies have reported adherence rates as low as 1% [29]. One of the few articles published about mobile app usage examined data that was collected from 4125 users between August 2010 and January 2011 [30]. The study found that sessions with apps were short, averaging a little more than a minute, and that different types of apps were used during different times of the day. Communication apps were the most frequently used, 49.50% of app launches, whereas the proportion of health apps was only 0.26% of all app launches [30]. More studies into health app usage behavior are thus warranted. Considering that people's eating patterns and daily routines vary over the week [31,32], analyzing the temporal context of app usage may help identify the best times to start using an app that promotes lifestyle changes.

Study Objectives

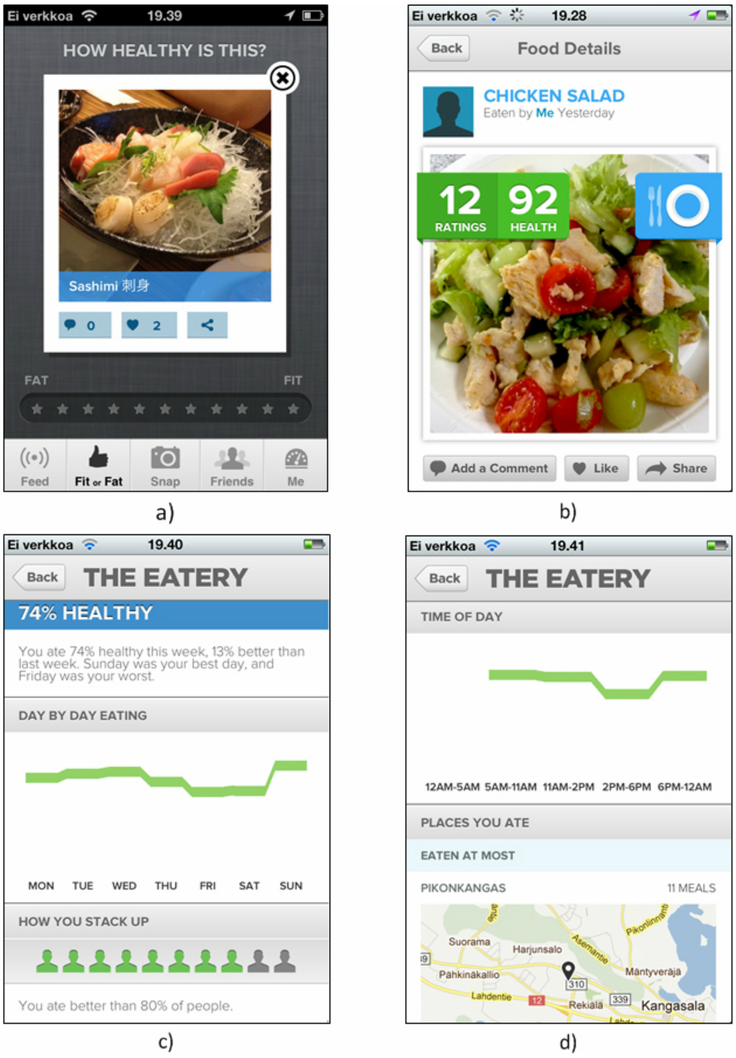
This study assesses the overall usage and reach of a free mobile app for healthy eating ("The Eatery") over a period from October 2011 to April 2012. Specifically, we examine the indicators of sustained use of the app, especially focusing on the initiation of self-monitoring and the influence of peer feedback.

Methods

Mobile App

The Eatery was a free iPhone app developed by the company Massive Health. The app was officially launched on November 1, 2011 in Apple's application market [33]. It was targeted toward English-speaking people and presented as an easy and fun way to eat healthily. Its main functions were photographic food recording, self-evaluation of foods, and crowdsourced peer feedback. Users were asked to take a picture of the foods they were going to eat, rate the picture on an arbitrary healthiness scale from fat (unhealthy) to fit (healthy) (Figure 1), and optionally write a description for the picture. In addition to self-evaluation, users were prompted to rate another user's food pictures every time they opened the app. They could rate as many successive pictures as they wanted (Figure 1a). Each of the user's own food pictures received an average healthiness rating that was calculated from ratings given by other users (Figure 1b) and displayed as a number between 0 ("fat") and 100 ("fit"). Other social support features of the app included the option to follow other users and provide feedback for their pictures in the form of comments and "likes". The app provided automated feedback on the user's past eating behavior by showing the past week's day-to-day healthiness ratings and overall rating (Figure 1c), allowing the user to note the worst times as potential improvement points. Feedback was also given on week-to-week progress in healthiness ratings and on social comparison to other users (Figure 1c). Moreover, the app provided information on past behavior by displaying the most frequent eating locations (Figure 1d) and highlighting the best and worst meals of the week. The only background information asked from the users was their dietary preference on the first launch of the app.

Figure 1. Screenshots of The Eatery app: a) rating other people's food with fat-fit scale, b) feedback received for photographed food, c) weekly summary, and d) summary of user's time-of-day healthiness ratings and places eaten at most.



You ate 74% healthy this week, 13% better than last week. Sunday was your best day, and Friday was your worst.

You ate better than 80% of people.

Study Sample

Altogether 189,770 users downloaded and used the app, The Eatery, at least once between October 15, 2011, and April 3, 2012. During this time, they generated 429,288 pictures and 7,946,447 ratings. In May 2012, Massive Health, the developer of the app, decided to make the anonymized dataset available for research purposes upon contact. The authors obtained the dataset from Massive Health in June 2012.

Data of users, pictures, and ratings included timestamps that represented the local time of the user's mobile phone. The timestamps that were stored when the user first used the app included time zone information for 98.51% (186,933/189,770) of the users: 68.41% (127,884/186,933) of them were from the main US time zones (UTC-8 to UTC-5) and 12.48% (23,335/186,933) were from the main European or African time zones (UTC+0 to UTC+3).

Definition of Variables

Table 1 lists the variables used in the analyses of factors related to usage of the app. As the focus of this study is on dietary self-monitoring, the usage period was defined as the time that elapsed between the first and the last picture taken by the user, even though the users who stopped taking their own food pictures could still continue to rate other users' pictures.

Table 1. Variables related to the usage of the mobile app "The Eatery".

Category	Variable	Description
Usage activity		
	Number of pictures	Total number of pictures taken by the user
	Usage period	Time elapsed between the first picture and the last picture (ie, the duration of self-monitoring)
	Pictures per day	Average number of pictures the user took per day during the usage period
	Ratings given for peers	Total number of ratings the user gave for other users' pictures
Context of use		
	Registration time	Day of week (Sun-Sat) and time of day when the user first used the app
	Dietary preference	The response the user gave to "How do you eat?" question during the first launch of the app. The preference categories are listed in Table 2.
Self-evaluation		
	Own healthiness rating	Healthiness rating the user gave for an own picture (0 to 1)[a]
	Picture description length	Number of characters written in the picture description
Peer feedback		
	Average healthiness rating	Mean peer rating given for the picture (0 to 1)[a]
	Number of ratings	Total number of peer ratings given for a picture
	Number of comments	Total number of comments from peers for a picture
	Number of likes	Total number of peers who "liked" a picture
	Difference to peer ratings	Difference between the user's own healthiness rating and average healthiness rating for a picture

[a] Healthiness ratings were stored as a decimal number from 0 ("fat") to 1 ("fit"), whereas the user saw the ratings as numbers from 0 to 100, as in Figure 1b. The rating that was displayed to the user was a non-linear mapping from peer ratings and user's own rating.

Exclusion Criteria in Analyses

Pictures that did not contain an actual image (3.13%, 13,433/429,288 were such "empty pictures") were removed from the data, resulting in a sample of 415,855 pictures for analysis. The overall quality and content of the pictures was screened by the researchers by examining a random sample of pictures. The examination revealed that, for some users, the first picture served as a test picture (for example, they took a picture of a chair to test out the application). Further examination showed that pictures obtaining a low number of peer ratings were typically something other than food, and therefore should be removed from further analysis. By manual inspection, the threshold of a valid picture was adjusted to 10 ratings: if the first picture taken by a user had received less than 10 ratings, the picture information was excluded and the second picture was used instead. If the user had taken only one picture or the second picture had received less than 10 ratings, the user was excluded from the analyses concerning the peer feedback for the first picture. The total number of pictures for each user was adjusted after the picture validity check of the first two pictures. This decreased the total number of pictures by a user by two pictures at most. Latter pictures were not examined. In total, 398,228 pictures (92.76% of all pictures and 95.76% of non-empty pictures) were classified as valid pictures.

Some users had not rated their own first picture—these users were excluded when analyzing the difference between their own and average peer healthiness ratings.

Adherence Levels

Individual users in the dataset were divided into groups based on their adherence, for the analysis of different indicators of adherence (initiation context of self-monitoring and peer feedback). The level of adherence was defined based on the total number of pictures taken and the length of the usage period of the app.

Two types of adherence classifications were formed for different analyses. In the first case, two user groups were formed: (1) users who had taken no valid pictures or only one valid picture (“Dropouts”, 86.39%, 163,949/189,770), and (2) users who had taken more than one valid picture (“Users”, 13.61%, 25,821/189,770). For users who had taken at least one valid picture, three activity levels were defined: (1) “Actives” who had taken at least 10 pictures and had used the app at least one week (2.58%, 4895/189,770), (2) “Semi-actives” who had taken at least two pictures, but less than 10 pictures or whose usage period was less than one week (11.03%, 20,926/189,770), and (3) “Non-actives” who had taken only one valid picture (17.36%, 32,948/189,770).

The proportion of users who had downloaded the app less than one week before the sampling period ended (on March 28, 2012 or later) was 2.01% (3812/189,770). Hence, they could not be classified as Actives. They were still included in the analyses due to their small number.

Initiation of Self-Monitoring

The association between users’ registration time and adherence level was analyzed to determine whether the temporal context of initial use could have an influence on subsequent usage activity. Registration time was categorized into seven weekdays and each day was divided into five time intervals: time between 0-5 (night), 5-10 (morning), 10-15 (daytime), 15-19 (late afternoon), and 19-24 (evening). These time intervals were chosen to correspond to the natural periods of the day and based on the assumption that most users were from Anglo-American culture, since the app was in English and roughly 68% (127,884/186,933) of the users registered from the main US time zones. The number of Dropouts and Users who had started using the app on each weekday and time of day intervals were calculated. The chi-square (χ^2) test was used to compare whether the proportions of Dropouts and Users in weekday and time of day intervals (35 options) were equal to each other. Bonferroni correction was used to adjust for multiple comparisons and the adjusted significance level was set at P=.0014. Further comparisons were exploratory and were made based on initial results. Registration time was not available for 1.49% of the users (2837/189,770).

Dietary preferences were divided into four categories based on the users’ response to “How do you eat?” question on the first use of the app. Table 2 lists the answer options to the question and the numbers of users in each category: (1) “Not defined” included users who had not given any preference (42.22%, 80,118/189,770), (2) “Everything” included users who chose the option “I eat everything” (46.33%, 87,912/189,770), (3) “Strict” included users who had at least one of the following options chosen: “Low carb, no carb, or paleo”, “Low fat”, or “Vegan/vegetarian”, (8.97%, 17,025/189,770), and (4) “Other” included users who had chosen or written an option that was not included in the first three classes (2.48%, 4715/189,770). For example, the variations of “I eat everything!” response such as “I eat everything except…” were categorized as “Other”.

The associations between dietary preferences and adherence were analyzed by calculating the proportion of (1) Actives out of Users + Non-actives (ie, out of all users who took at least one valid picture), and (2) Users out of Users + Non-actives for each dietary preference category. The chi-square test was used to examine whether the proportions were equal between different dietary preference categories. Tukey’s HSD (honestly significant difference) multiple comparison test among proportions was used to analyze which dietary preference categories differed from each other after obtaining significance value P<.05.

Finally, the existence and length of the textual description given for the first picture taken by the user were compared between Active, Semi-active, and Non-active user groups. One-way ANOVA (analysis of variance) was used for description length and the chi-square test for the existence of the description. This analysis was done to assess the engagement level of the user during the initiation of self-monitoring.

Table 2. Numbers of users according to their dietary preferences based on “How do you eat?” question.[a]

“How do you eat?”	Category	Number of users, n (%) n=189,770
Not defined	Not defined	80,118 (42.22%)
“I eat everything!”	Everything	87,912 (46.33%)
“Low fat”	Strict	7778 (4.10%)
“Low carbs, no carbs, or paleo”	Strict	7146 (3.77%)
“Vegan or vegetarian”	Strict	6223 (3.28%)
“Complex carb diet”	Other	2388 (1.26%)
“Other”	Other	2427 (1.28%)
“Gluten free” or “gluten free”	Other	229 (0.12%)
None of the above	Other	1714 (0.90%)
Total	Strict	17,025 (8.97%)
Total	Other	4715 (2.48%)

[a] Note that some users provided multiple responses to the question.

Peer Feedback

The amount and quality of peer feedback given for the first picture (average healthiness score, number of likes, number of comments, and difference to peer ratings) were compared between Active, Semi-active, and Non-active user groups to determine whether higher level of feedback on the initiation of self-monitoring was connected with adherence. Only those who had at least one valid picture among the first two pictures they had taken were included because the focus was on the initial feedback. For continuous variables, one-way ANOVA was used and for binary variables, the chi-square test was used to compare whether the proportions were equal between user groups. The numbers of ratings given by the users in each dietary preference category were also calculated to determine whether the stated dietary preference would have a connection to the user’s activity in providing peer feedback to others.

Changes in Healthiness Ratings

Changes in healthiness ratings were analyzed only among Active users who had at least one valid picture among the first two pictures they had taken (99.35% of Actives, 4863/4895). Other user groups used the app for such a short time that no trend could reliably be identified. First, a correlation coefficient between the average healthiness rating of the first picture and all subsequent pictures was determined. A change (linear regression coefficient) in healthiness ratings as a function of picture index and corresponding P value was calculated for each Active user. The dependent variable was the healthiness rating of a picture and the independent variable was the picture index $1,2,\ldots,N$ where N was the number of pictures taken by the user. Note that the ordered list of pictures was used instead of real time axis. If a significant ($P<.05$) positive linear coefficient was found, the user was categorized into “Improvers” (improvement in diet), and negative into “Decliners” (deterioration in diet). Student’s t test was used to compare whether usage activity (number of pictures, usage period, and pictures per day) differed between Improvers and other Actives.

Changes in eating behavior among users with different dietary preferences were also examined. One-way ANOVA was used to compare the average healthiness rating for the first picture and the healthiness rating for all pictures between different dietary preference categories. The number of Improvers or Decliners in each dietary category was determined. The chi-square test was used to examine whether there were an equal proportion of Improvers and Decliners in each dietary preference category.

Results

Overall Adherence and Healthiness Ratings

Table 3 shows the numbers of users divided into different adherence levels based on their usage activity. The average number of pictures and usage period in days is also shown for Semi-actives and Actives. Only 2.58% (4895/189,770) of the users became Active users, whereas more than two-thirds of the users did not take any valid pictures. On average, Actives took 1.6 pictures per day and 14.99% (734/4895) of them took more than three pictures per day.

Table 4 summarizes the statistics of valid pictures (92.76% of all pictures, 398,228/429,288) in the dataset from The Eatery. Their average healthiness rating, 0.58, was slightly above the midpoint of the fat-fit scale from 0 to 1.

Table 3. Adherence data for users who downloaded the free dietary self-monitoring app between October 15, 2011 and April 3, 2012 (n=189,770).

User group	Activity level	Description	Count, n (%)	Pictures per user, mean (SD)	Usage period in days, mean (SD)
Dropouts	Non-users	No pictures or no valid pictures	131,001 (69.03%)	-	-
Dropouts	Non-actives	Only 1 valid picture	32,948 (17.36%)	-	-
Users	Semi-actives	At least two valid pictures and less than 10 pictures or usage period shorter than 7 days	20,926 (11.03%)	4.1 (3.7)	9.3 (19.2)
Users	Actives	At least 10 pictures and usage period longer than 7 days	4895 (2.58%)	58.9 (99.5)	46.6 (37.7)

Table 4. Statistics for the 398,228 valid pictures taken by 58,769 users of the dietary self-monitoring app “The Eatery”.

Variable	Description	Value, mean (SD; range) or n (%)
Self-evaluations		
	Number of pictures with textual description	293,692 (73.75%)
	Average length of textual description (if existed) as number of characters	26.1 (18.1; 1-248)
Peer feedback		
	Average healthiness rating	0.581 (0.195; 0.0261-0.986)
	Number of pictures having at least one like	61,299 (15.39%)
	Average number of likes (if existed)	1.3 (0.9; 1-21)
	Number of pictures having at least one comment	15,247 (3.83%)
	Average number of comments (if existed)	1.7 (1.4; 1-28)

Initiation of Self-Monitoring

The associations between users’ registration time and adherence level are presented in Figure 2, which compares the proportions of Users and Dropouts who started using the app on each day of week and time of day interval. A higher proportion of Users started using the app during the daytime on weekdays than Dropouts (20.28%, 5237/25,820 vs 15.34%, 24,718/161,113; χ^2_1=356.3, *P*<.001). Moreover, a higher proportion of Users started using the app especially during Tuesdays than Dropouts (17.66%, 4561/25,820 vs 14.88%, 23,974/161,113; χ^2_1=133.4, *P*<.001).

Most common dietary preferences (see Table 2) reported by users during the first use of the app were “I eat everything” (46.33%) or undefined (42.22%). In total, 8.97% of the users were considered to have Strict diets. Table 5 presents the differences in adherence levels between dietary preference groups among users who took at least one valid picture. Users with Strict diets were the most likely (14.31%, 900/6291) and users who had not defined any diet were the least likely (3.99%, 742/18,590) to use the app actively. A similar trend was observed when Semi-active users were included in the comparisons: half (50.45%, 3174/6291) of the users with Strict diets were Semi-active or Active, a significantly higher proportion than among users who had not defined their diets.

Engagement of the user during the initiation of self-monitoring was also assessed by examining the textual description given for the first picture. A textual description for the first picture was given by 26.09% (15,179/58,170) of users who had at least one valid picture among the first two pictures they had taken. Table 6 presents the comparisons between different adherence groups. Active users had written a description for the first picture more often than Semi-actives or Non-actives, and the average description was also longer in the number of characters.

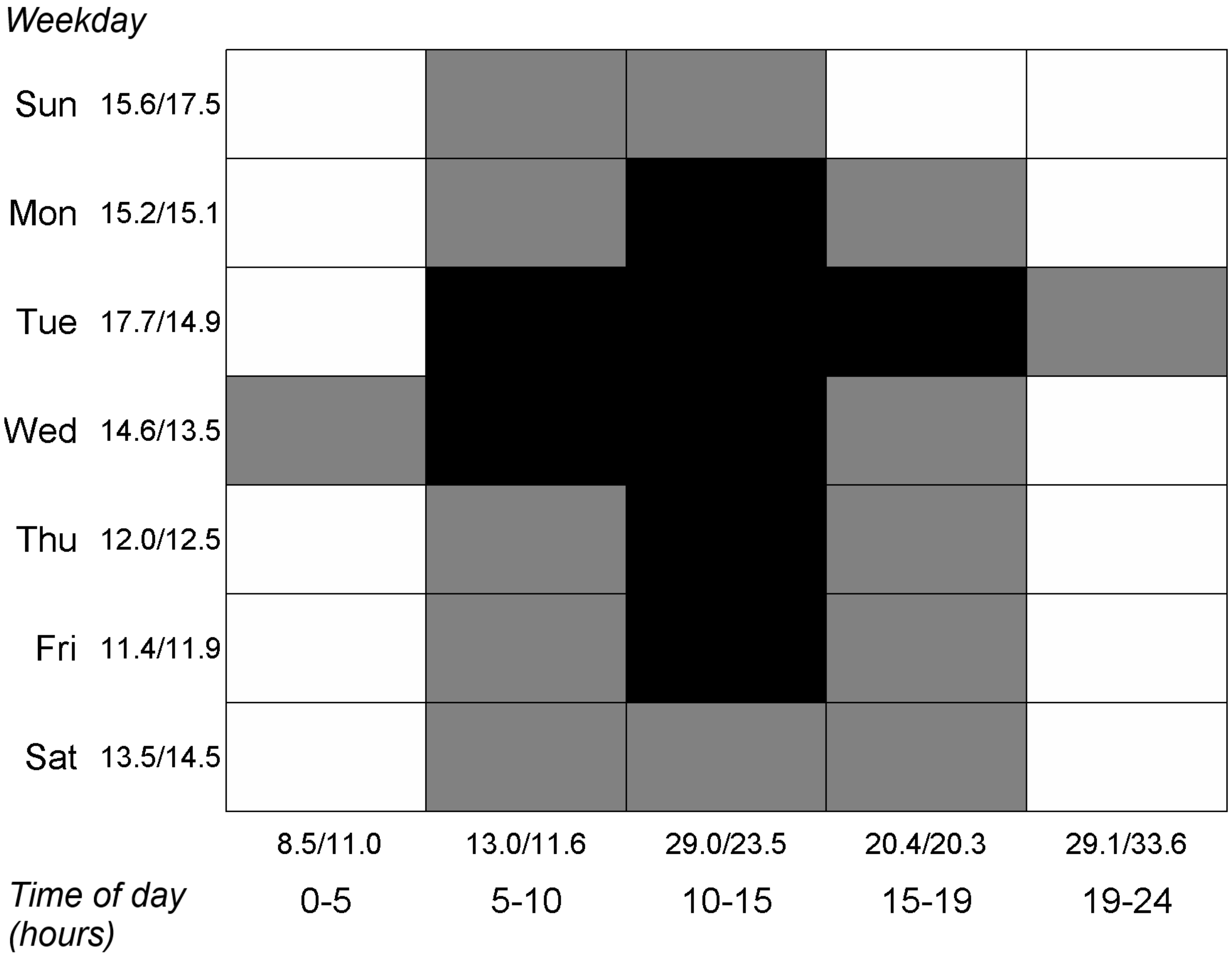
Table 5. Proportions of Semi-active and Active users in each dietary preference category out of all users who took at least one valid picture.

Users	1. Not defined, n (%) n=18,590	2. Everything, n (%) n=32,090	3. Strict, n (%) n=6291	4. Other, n (%) n=1798	Test statistics	Differences in post hoc comparisons
Actives / Users+Non-actives	742 (4.0%)	3040 (9.47%)	900 (14.31%)	213 (11.9%)	χ^2_3=826.6, *P*<.001	All groups
Users / Users+Non-actives	7188 (38.67%)	14,560 (45.37%)	3174 (50.45%)	899 (50.0%)	χ^2_3=371.8, *P*<.001	All but not 3 and 4

Table 6. Comparison of user engagement in the first self-monitoring entry between different adherence groups as measured by the presence and length of textual description for the picture.

First picture characteristics	1. Non-actives (n=32,648)	2. Semi-actives (n=20,659)	3. Actives (n=4863)	Test statistics	Differences in post hoc comparisons
Presence of textual description, n (%)	5783 (17.71%)	6824 (33.03%)	2572 (52.89%)	χ^2_2=3515.1, *P*<.001	All groups
Number of characters in description (if existed), mean (SD)	20.1 (15.8)	23.4 (17.5)	26.8 (19.1)	$F_{2,15176}$=150.1, *P*<.001	All groups

Figure 2. Correlations between users' adherence level and their local registration time. Black=higher proportion of Users (*P*<.0014); White=higher proportion of Dropouts; Grey=no difference. Numbers separated by slashes next to weekday and time of day labels are percentages of Users/Drop-outs for corresponding rows and columns.



Peer Feedback

Feedback received by the users' first pictures was examined to determine whether higher level of feedback on the initiation of self-monitoring was connected with adherence. The first picture had at least one like among 7.68% (4470/58,170) of the users and at least one comment among 3.85% (2240/58,170) of them. Table 7 presents the comparisons of variables related to peer feedback between different adherence groups. Small but significant differences were found between Active and less active users for all variables: the average healthiness rating was higher and the proportion of pictures having comments and likes was higher. Still, comments and likes were relatively rare even among Active users.

Peer feedback was also examined from the perspective of users who gave the ratings to others. Analysis of dietary preferences and rating activity found that users in the "Not defined" diet group gave 21.81% (1,732,976/7,946,447) of all ratings, users in the "Everything" group gave 52.83% (4,198,272/7,946,447), users with Strict diets gave 20.70% (1,645,134/7,946,447), and users with some Other diet gave 4.66% (370,065/7,946,447) of all ratings.

Table 7. Amount and quality of peer feedback for the initial self-monitoring record in the app between different adherence groups.

First picture characteristics	1. Non-actives (n=32,648)	2. Semi-actives (n=20,659)	3. Actives (n=4863)	Test statistics	Differences in post hoc comparisons
Average healthiness rating, mean (SD)	0.49 (0.21)	0.52 (0.20)	0.55 (0.19)	$F_{2,58167}=225.9$, $P<.001$	All groups
Difference to peer ratings, mean (SD)	0.04 (0.22)	0.04 (0.21)	0.05 (0.18)	$F_{2,57738}=5.1$, $P=.006$	1 and 3, 2 and 3
Having at least one like, n (%)	2031 (6.22%)	1792 (8.67%)	647 (13.30%)	$\chi^2_2=343.6$, $P<.001$	All groups
Number of likes (if at least one), mean (SD)	1.1 (0.3)	1.1 (0.4)	1.2 (0.4)	$F_{2,4467}=13.6$, $P<.001$	1 and 3, 2 and 3
Having at least one comment, n (%)	663 (2.03%)	1088 (5.27%)	489 (10.06%)	$\chi^2_2=909.6$, $P<.001$	All groups
Number of comments (if at least one), mean (SD)	1.2 (0.6)	1.3 (0.9)	1.4 (1.1)	$F_{2,2237}=15.1$, $P<.001$	1 and 3, 2 and 3

Changes in Healthiness Ratings

Among the 4863 Active users who had at least one valid picture as their first or second picture, 481 (9.89%) had a significant positive trend in healthiness scores. These "Improvers" differed from other Actives by having a higher total number of pictures (mean 126.68, SD 183.73 vs mean 51.41, SD 82.16; $t_{4861}=16.19$; $P<.001$), a longer usage period in days (mean 68.17, SD 42.95 vs mean 44.16, SD 36.34; $t_{4861}=13.54$; $P<.001$), and a higher number of pictures per day (mean 1.80, SD 1.70 vs mean 1.55, SD 1.58; $t_{4861}=3.25$; $P=.001$). In other words, they used the app for a longer time and did dietary self-monitoring more frequently.

Users with Strict diets had higher healthiness scores than users in other dietary preference categories and they also had the highest proportion of Improvers (Table 8).

Table 8. Average healthiness rating and number of users (Actives) that had a significant linear coefficient in their healthiness rating in each dietary preference category.

Scores/users	1. Not defined (n=732)	2. Everything (n=3023)	3. Strict (n=896)	4. Other (n=212)	Test statistics	Differences in post hoc comparisons
Average healthiness rating (first picture), mean (SD)	0.54 (0.19)	0.53 (0.19)	0.60 (0.18)	0.56 (0.18)	$F_{3,4859}=29.3$, $P<.001$	1 and 3, 2 and 3, 3 and 4
Average healthiness rating (all pictures), mean (SD)	0.56 (0.08)	0.57 (0.09)	0.63 (0.08)	0.60 (0.10)	$F_{3,4859}=149.8$, $P<.001$	All groups
Number of Improvers, n (%)	55 (7.51%)	281 (9.30%)	125 (13.95%)	20 (9.43%)	$\chi^2_3=22.5$, $P<.001$	1 and 3, 2 and 3
Number of Decliners, n (%)	14 (1.91%)	72 (2.38%)	32 (3.57%)	6 (2.83%)	$\chi^2_3=5.4$, $P=.15$	None

Discussion

Overall Adherence and Changes in Healthiness Ratings

Almost 190,000 people downloaded the app, The Eatery, between October 2011 and April 2012, but attrition was very high: less than 3% were active users, that is, used the app for more than a week and took 10 or more food pictures. Most of the users did not take any pictures (69%) or took only one picture (17%), which means that they only downloaded the app and experimented with it once without starting dietary self-monitoring. This is similar to most free apps, which are easy to join and try out even if there is no serious intention or commitment to start using the app [13,25]. Given the short usage period for majority of the users, many probably tried out the application for fun.

The Eatery was not marketed as a weight loss app but instead as a method to eat healthier ("Stop counting calories, start eating better"), and hence may have attracted a large number of users with no real interest in dietary improvements and thus lacking motivation for dietary self-monitoring. However, the few active users used the application on average 1.5 months. Dietary self-monitoring for this amount of time would be enough to lead to increased awareness of eating habits and changes in behavior, if done diligently. The average healthiness rating of all pictures was 0.58, slightly above the midpoint on the scale of 0 (“fat”) to 1 (“fit”). Hence, users did not photograph only healthy foods and there was room for improvement. A positive trend in healthiness ratings was still observed among only 10% of active users (0.3% of all users). Even if we assume that this trend reflects changes in their real-life eating behavior, the impact of the app on eating choices (or choosing which foods to record) appears to have been very small. Active users took less than two pictures per day on average, which means that a large portion of their eating was left unrecorded. Thus, the positive trend among some users could also mean that they started “gaming the system” by selectively photographing their foods to get better ratings and comments.

Initiation of Self-Monitoring

Users who used the app for the first time on weekdays (especially on Tuesdays or Wednesdays) and during morning or daytime became semi-active or active users more often than those who started using the app during evenings or weekends. People’s varying eating patterns that depend on their schedules during workdays and outside work [34] can help explain this finding. Prior studies have found that diet quality is generally poorer during weekends than on weekdays and calorie intake is higher, especially in the form of fat and alcohol [31,32,35], so people do not necessarily want to start tracking their eating at these times. Moreover, people are generally less work-oriented in the evening and during weekends and may try out different apps just for fun. In contrast, someone who downloads a healthy eating app in the middle of the week during daytime probably has the intention to start keeping track of their eating. This time period could also be a fruitful time to suggest initiation of lifestyle changes, although everyone does not have the same work schedules.

The dietary preferences reported on the initial use were also connected to adherence level. Users who reported a “strict” diet (low fat, low/no carbs, or vegan/vegetarian) were most likely to become active users. They also gave 21% of all ratings, although only 9% of all users belonged to the strict group. Hence, it is possible that users with strict diets were already most interested in healthy eating. This is also supported by their healthiness ratings: active users with strict diets had higher average healthiness ratings for their first picture and also higher average healthiness ratings for all pictures than users in other dietary preference groups.

The motivation of sustained users might already be seen on the initiation of self-monitoring by looking at how much time and cognitive capacity they devote to it. This is supported by the finding that more than half of the active users (53%) gave a textual description for their first picture whereas less than one-fifth of the non-active users (18%) did so. In addition to pre-existing intention to start dietary self-monitoring, the initial user experience of the app most likely influenced the users’ intention to continue using it. Positive feedback received from peers for the first picture taken by the user was associated with higher adherence; active users had higher average healthiness ratings for the first picture than less active users. This begs the question—did they happen to take a picture of a healthy food and were encouraged by the good feedback to continue using the application or were they already healthy eaters, thus naturally photographing a healthy food? Because a higher proportion of active users also used the app for the first time during weekdays and daytime, the food that they chose to photograph first was probably their workday lunch, which is often healthier than foods that are eaten during weekends [35]. Hence, the timing of the initial use of the dietary self-monitoring app may be important both in terms of the users’ pre-existing motivation and the type of reinforcing feedback generated by the app.

Peer Feedback

Although active users obtained more comments and likes for their pictures from peers than those who took only few pictures, the total percentage of pictures with comments (4%) and likes (15%) was quite low. Thus, most users had no connection to other users other than receiving and giving anonymous ratings. The social network formed in such a way is very loose: users neither know whose pictures they rate nor have any knowledge of who rates their pictures. The app itself did not offer explicit advice on what to do to improve eating habits or what constitutes a healthy diet. It may be that if users are merely told that their meal is unhealthy but not given any advice on what to do to make it better, they do not get enough value out of the experience and subsequently lack motivation to continue using the app [36]. People may also have very different ideas about what healthy or “fit” food is, but these differences do not seem to have influenced adherence in this study. Although the difference between the user’s own rating and average peer ratings was highest among active users, in practice this difference was very small.

An app like this relies on its users to provide one of its core functions, that is, peer feedback. It would be interesting to study what motivates people to participate in this crowdsourcing activity of giving ratings to others. One explanation is the reciprocity of the action: when a user rates someone else’s pictures, they also get ratings for their own. However, engaging in this activity for a long time might require an existing community or formation of stronger ties between users [27].

Limitations and Challenges

The most significant limitation of the study is the lack of information about user demographics, behavioral outcomes, and initial motives. For example, the association between outcomes and adherence to dietary self-monitoring has been found to differ between race and gender groups in weight loss interventions [12], and it would have been interesting to see if similar patterns had emerged in this context. In this study, the application only provided data about the users’ dietary preferences. The general statistics about smartphone users in early 2012 suggest that iPhone users were slightly older than Android users and downloaded more apps in a month than users of other smartphone systems [37]. Users of The Eatery owned an iPhone so it is possible that their characteristics followed the same pattern. Collecting comprehensive data about users’ background may be challenging in free apps, which aim for fluent user experience, but creative ways to gather data such as asking one question per usage occasion could be devised in further studies.

The reliability of healthiness ratings is questionable because they were entirely crowdsourced. The idea of crowdsourcing is to take an average of many individuals' estimates resulting in an estimate that can be surprisingly close to the truth, although individuals' separate values may lie far from it. Crowdsourced ratings can be biased, resulting from cultural differences [26] or rater's own food preferences. When pictures are rated as in this study, the quality of the picture is also likely to make a difference. Portion size estimation is difficult even when measurement aids are present in the picture [18,19,38]. Moreover, users did not photograph everything they ate so there is no way of knowing how healthily people ate in general and whether the observed positive trend in healthiness ratings among 10% of active users meant anything in practice.

Dietary decisions are often unconscious and affected by environmental factors more than people believe [39]. In this study, the location information of where the pictures were taken was available, but was not exploited, although dietary behaviors are likely to be linked with locations. In a recent study [40], volunteers used Twitter to report their food and were encouraged to add a photo and contextual information, such as company, mood, and reasons for eating. The data was used visualize the relationships between dietary and behavioral factors.

Some updates were released to the app during the six-month timeframe of data collection. These updates consisted of minor modification and fixes in the user interface of the app. They may have had a minor influence on the user experience of the app, but they were not included in the analyses since the main features and functions remained the same.

Finally, the app utilized self-regulation techniques of self-monitoring and feedback, but lacked other techniques derived from control theory [4,5]. The app did not prompt users to set specific goals or review behavioral goals and only implicitly prompted intention formation ("eat healthy"). Implementing these techniques could have given users a more specific purpose for using the app. Practical advice for healthy eating could also have been given by formation of implementation intentions through if-then planning, which has been shown to increase effectiveness of healthy eating interventions [10].

Conclusions

As with most mobile apps, the majority of users tried the dietary self-monitoring app only once. Adherence was higher among users who had diets that were likely to restrict at least some unhealthy foods, and these kinds of users were also more active in rating other users' foods. This could mean that this kind of an application attracts users with special diets and/or those interested in food. Moreover, initiation of self-monitoring in the middle of the week during daytime and the amount of feedback from peers were connected to higher adherence.

Even though the findings show that the app reached a large number of people, its actual impact among users remained small because most did not even start dietary self-monitoring with the app. If people would use the app as intended for dietary self-monitoring on a regular basis, they could experience some benefits through heightened awareness of their eating habits. Still, the app did not implement all self-regulation techniques that could have strengthened its impact and it lacked means to track changes in eating behavior systematically. Reaching those users who could benefit the most from dietary self-monitoring and maintaining their adherence remains a challenge.

Acknowledgments

The authors thank Sylvia Cheng and Max Utter at Jawbone for providing the Massive Health dataset and assisting with preliminary data analyses. Massive Health was acquired by Jawbone in February of 2013 and Jawbone continues to support the Massive Health application and associated research. The authors also thank Aleksi Salo at Tampere University of Technology for carrying out data extraction.

This work was partially supported by the SalWe Research Program for Mind and Body (Tekes-the Finnish Funding Agency for Technology and Innovation grant 1104/10).

Conflicts of Interest

None declared.

References

1. Contento I, Balch G, Bronner Y, Lytle L, Maloney S, Olson C, et al. The effectiveness of nutrition education and implications for nutrition education policy, programs, and research: a review of research. Journal of Nutrition Education 1995;27(6):277-418.
2. National Institute of Health (NIH). Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults: the evidence report. Bethesda, MD: National Heart, Lung, and Blood Institute; 1998. URL: http://www.nhlbi.nih.gov/guidelines/obesity/prctgd_c.pdf [accessed 2013-11-04] [WebCite Cache ID 6KsKK0SU3]
3. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
4. Carver CS, Scheier MF. Control theory: a useful conceptual framework for personality-social, clinical, and health psychology. Psychol Bull 1982 Jul;92(1):111-135. [Medline: 7134324]

5. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
6. Stumbo PJ. New technology in dietary assessment: a review of digital methods in improving food record accuracy. Proc Nutr Soc 2013 Feb;72(1):70-76. [doi: 10.1017/S0029665112002911] [Medline: 23336561]
7. Glanz K, Murphy S, Moylan J, Evensen D, Curb JD. Improving dietary self-monitoring and adherence with hand-held computers: a pilot study. Am J Health Promot 2006;20(3):165-170. [Medline: 16422134]
8. Boushey CJ, Kerr DA, Wright J, Lutes KD, Ebert DS, Delp EJ. Use of technology in children's dietary assessment. Eur J Clin Nutr 2009 Feb;63 Suppl 1:S50-S57 [FREE Full text] [doi: 10.1038/ejcn.2008.65] [Medline: 19190645]
9. Zhu F, Bosch M, Woo I, Kim S, Boushey CJ, Ebert DS, et al. The use of mobile devices in aiding dietary assessment and evaluation. IEEE J Sel Top Signal Process 2010 Aug;4(4):756-766 [FREE Full text] [doi: 10.1109/JSTSP.2010.2051471] [Medline: 20862266]
10. Rothman AJ, Sheeran P, Wood W. Reflective and automatic processes in the initiation and maintenance of dietary change. Ann Behav Med 2009 Dec;38 Suppl 1:S4-17. [doi: 10.1007/s12160-009-9118-3] [Medline: 19787308]
11. Baker RC, Kirschenbaum DS. Self-monitoring may be necessary for successful weight control. Behavior Therapy 1993;24(3):377-394. [doi: 10.1016/S00057894(05)80212-6]
12. Hollis JF, Gullion CM, Stevens VJ, Brantley PJ, Appel LJ, Ard JD, Weight Loss Maintenance Trial Research Group. Weight loss during the intensive intervention phase of the weight-loss maintenance trial. Am J Prev Med 2008 Aug;35(2):118-126 [FREE Full text] [doi: 10.1016/j.amepre.2008.04.013] [Medline: 18617080]
13. Hwang KO, Ning J, Trickey AW, Sciamanna CN. Website usage and weight loss in a free commercial online weight loss program: retrospective cohort study. J Med Internet Res 2013;15(1):e11 [FREE Full text] [doi: 10.2196/jmir.2195] [Medline: 23322819]
14. Thompson FE, Subar AF. Dietary assessment methodology. In: Coulston AM, Boushey CJ, editors. Nutrition in the Prevention and Treatment of Disease. San Diego, CA: Academic Press; 2008:3-39.
15. Thompson FE, Subar AF, Loria CM, Reedy JL, Baranowski T. Need for technological innovation in dietary assessment. J Am Diet Assoc 2010 Jan;110(1):48-51 [FREE Full text] [doi: 10.1016/j.jada.2009.10.008] [Medline: 20102826]
16. Subar AF, Crafts J, Zimmerman TP, Wilson M, Mittl B, Islam NG, et al. Assessment of the accuracy of portion size reports using computer-based food photographs aids in the development of an automated self-administered 24-hour recall. J Am Diet Assoc 2010 Jan;110(1):55-64 [FREE Full text] [doi: 10.1016/j.jada.2009.10.007] [Medline: 20102828]
17. Daugherty BL, Schap TE, Ettienne-Gittens R, Zhu FM, Bosch M, Delp EJ, et al. Novel technologies for assessing dietary intake: evaluating the usability of a mobile telephone food record among adults and adolescents. J Med Internet Res 2012;14(2):e58 [FREE Full text] [doi: 10.2196/jmir.1967] [Medline: 22504018]
18. Frobisher C, Maxwell SM. The estimation of food portion sizes: a comparison between using descriptions of portion sizes and a photographic food atlas by children and adults. J Hum Nutr Diet 2003 Jun;16(3):181-188. [Medline: 12753111]
19. Lee CD, Chae J, Schap TE, Kerr DA, Delp EJ, Ebert DS, et al. Comparison of known food weights with image-based portion-size automated estimation and adolescents' self-reported portion size. J Diabetes Sci Technol 2012 Mar;6(2):428-434 [FREE Full text] [Medline: 22538157]
20. Burke LE, Sereika SM, Music E, Warziski M, Styn MA, Stone A. Using instrumented paper diaries to document self-monitoring patterns in weight loss. Contemp Clin Trials 2008 Mar;29(2):182-193 [FREE Full text] [doi: 10.1016/j.cct.2007.07.004] [Medline: 17702667]
21. Zepeda L, Deal D. Think before you eat: photographic food diaries as intervention tools to change dietary decision making and attitudes. International Journal of Consumer Studies 2008;32(6):692-698. [doi: 10.1111/j.1470-6431.2008.00725.x]
22. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
23. Cugelman B, Thelwall M, Dawes P. Online interventions for social marketing health behavior change campaigns: a meta-analysis of psychological architectures and adherence factors. J Med Internet Res 2011;13(1):e17 [FREE Full text] [doi: 10.2196/jmir.1367] [Medline: 21320854]
24. Helsel DL, Jakicic JM, Otto AD. Comparison of techniques for self-monitoring eating and exercise behaviors on weight loss in a correspondence-based intervention. J Am Diet Assoc 2007 Oct;107(10):1807-1810. [doi: 10.1016/j.jada.2007.07.014] [Medline: 17904942]
25. 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 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]
26. Noronha J, Hysen E, Zhang H, Krzysztof Z. Platemate: crowdsourcing nutritional analysis from food photographs. In: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology. New York, NY: ACM; 2011:1-12.
27. Parker A, Kantroo V, Lee HR, Osornio M, Sharma M, Grinter R. Health promotion as activism: building community capacity to effect social change. In: CHI'12 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY: ACM; 2012:99-108.

28. 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]
29. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
30. Böhmer M, Hecht B, Schöning J, Krüger A, Bauer G. Falling asleep with Angry Birds, Facebook and Kindle: a large scale study on mobile application usage. In: MobileHCI'11 Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services. New York, NY: ACM; 2011:47-56.
31. McCarthy S. Weekly patterns, diet quality and energy balance. Physiol Behav 2014 Mar 1:1 (forthcoming). [doi: 10.1016/j.physbeh.2014.02.046] [Medline: 24589367]
32. Racette SB, Weiss EP, Schechtman KB, Steger-May K, Villareal DT, Obert KA, et al. Influence of weekend lifestyle patterns on body weight. Obesity (Silver Spring) 2008 Aug;16(8):1826-1830 [FREE Full text] [doi: 10.1038/oby.2008.320] [Medline: 18551108]
33. Massive Health. The Eatery website URL: https://eatery.massivehealth.com/ [accessed 2013-11-04] [WebCite Cache ID 6KsJsYqBx]
34. Jastran MM, Bisogni CA, Sobal J, Blake C, Devine CM. Eating routines. Embedded, value based, modifiable, and reflective. Appetite 2009 Feb;52(1):127-136 [FREE Full text] [doi: 10.1016/j.appet.2008.09.003] [Medline: 18835305]
35. Haines PS, Hama MY, Guilkey DK, Popkin BM. Weekend eating in the United States is linked with greater energy, fat, and alcohol intake. Obes Res 2003 Aug;11(8):945-949. [doi: 10.1038/oby.2003.130] [Medline: 12917498]
36. Gummerus J, Pihlström M. Context and mobile services' value-in-use. Journal of Retailing and Consumer Services 2011 Nov;18(6):521-533. [doi: 10.1016/j.jretconser.2011.07.002]
37. Alexander A. Smartphone usage statistics 2012 (infographic) URL: http://ansonalex.com/infographics/smartphone-usage-statistics-2012-infographic/ [accessed 2014-03-31] [WebCite Cache ID 6OTvN95oZ]
38. Hernández T, Wilder L, Kuehn D, Rubotzky K, Moser-Veillon P, Godwin S, et al. Portion size estimation and expectation of accuracy. Journal of Food Composition and Analysis 2006 Aug;19:S14-S21. [doi: 10.1016/j.jfca.2006.02.010]
39. Wansink B. Environmental factors that increase the food intake and consumption volume of unknowing consumers. Annu Rev Nutr 2004;24:455-479. [doi: 10.1146/annurev.nutr.24.012003.132140] [Medline: 15189128]
40. Hingle M, Yoon D, Fowler J, Kobourov S, Schneider ML, Falk D, et al. Collection and visualization of dietary behavior and reasons for eating using Twitter. J Med Internet Res 2013;15(6):e125 [FREE Full text] [doi: 10.2196/jmir.2613] [Medline: 23796439]

Abbreviations

ANOVA: analysis of variance

Edited by G Eysenbach; submitted 05.11.13; peer-reviewed by P Stumbo, L Zepeda; comments to author 05.02.14; revised version received 09.03.14; accepted 30.03.14; published 15.04.14.

Please cite as:
Helander E, Kaipainen K, Korhonen I, Wansink B
Factors Related to Sustained Use of a Free Mobile App for Dietary Self-Monitoring With Photography and Peer Feedback: Retrospective Cohort Study
J Med Internet Res 2014;16(4):e109
URL: http://www.jmir.org/2014/4/e109/
doi:10.2196/jmir.3084
PMID:24735567

Original Paper

Can an Internet-Based Health Risk Assessment Highlight Problems of Heart Disease Risk Factor Awareness? A Cross-Sectional Analysis

Justin B Dickerson[1], MBA, PhD; Catherine J McNeal[2,3], MD, PhD; Ginger Tsai[3]; Cathleen M Rivera[2,3], MD, MS; Matthew Lee Smith[1,4], MPH, PhD; Robert L Ohsfeldt[5], PhD; Marcia G Ory[1], MPH, PhD

[1]School of Public Health, Department of Health Promotion and Community Health Sciences, Texas A&M Health Science Center, College Station, TX, United States

[2]Department of Internal Medicine, Scott & White Healthcare, Temple, TX, United States

[3]College of Medicine, Texas A&M Health Science Center, College Station, TX, United States

[4]College of Public Health, Department of Health Promotion and Behavior, The University of Georgia, Athens, GA, United States

[5]School of Public Health, Department of Health Policy & Management, Texas A&M Health Science Center, College Station, TX, United States

Corresponding Author:
Justin B Dickerson, MBA, PhD
School of Public Health
Department of Health Promotion and Community Health Sciences
Texas A&M Health Science Center
1266 TAMU
College Station, TX, 77843
United States
Phone: 1 979 777 8714
Fax: 1 503 689 8928
Email: hsrhealthservices@gmail.com

Abstract

Background: Health risk assessments are becoming more popular as a tool to conveniently and effectively reach community-dwelling adults who may be at risk for serious chronic conditions such as coronary heart disease (CHD). The use of such instruments to improve adults' risk factor awareness and concordance with clinically measured risk factor values could be an opportunity to advance public health knowledge and build effective interventions.

Objective: The objective of this study was to determine if an Internet-based health risk assessment can highlight important aspects of agreement between respondents' self-reported and clinically measured CHD risk factors for community-dwelling adults who may be at risk for CHD.

Methods: Data from an Internet-based cardiovascular health risk assessment (Heart Aware) administered to community-dwelling adults at 127 clinical sites were analyzed. Respondents were recruited through individual hospital marketing campaigns, such as media advertising and print media, found throughout inpatient and outpatient facilities. CHD risk factors from the Framingham Heart Study were examined. Weighted kappa statistics were calculated to measure interrater agreement between respondents' self-reported and clinically measured CHD risk factors. Weighted kappa statistics were then calculated for each sample by strata of overall 10-year CHD risk. Three samples were drawn based on strategies for treating missing data: a listwise deleted sample, a pairwise deleted sample, and a multiple imputation (MI) sample.

Results: The MI sample (n=16,879) was most appropriate for addressing missing data. No CHD risk factor had better than marginal interrater agreement ($\kappa > .60$). High-density lipoprotein cholesterol (HDL-C) exhibited suboptimal interrater agreement that deteriorated (eg, $\kappa < .30$) as overall CHD risk increased. Conversely, low-density lipoprotein cholesterol (LDL-C) interrater agreement improved (eg, up to $\kappa = .25$) as overall CHD risk increased. Overall CHD risk of the sample was lower than comparative population-based CHD risk (ie, no more than 15% risk of CHD for the sample vs up to a 30% chance of CHD for the population).

Conclusions: Interventions are needed to improve knowledge of CHD risk factors. Specific interventions should address perceptions of HDL-C and LCL-C. Internet-based health risk assessments such as Heart Aware may contribute to public health surveillance, but they must address selection bias of Internet-based recruitment methods.

(J Med Internet Res 2014;16(4):e106) doi:10.2196/jmir.2369

KEYWORDS

health risk assessment; Internet; risk factors; health disease; concordance

Introduction

The Framingham Heart Study defines cardiovascular disease (CVD) as a combination of coronary heart disease (CHD), various types of stroke, peripheral artery disease, and heart failure [1]. CVD is a leading cause of death in both males and females living in the United States [2]. In 2006, CVD was responsible for the death of 1 in 4 Americans [3]. CVD is one of the costliest medical conditions to treat with total economic costs estimated at more than US $300 billion annually [3], a cost that is also predicted to rise substantially over the next 2 decades as a result of technological improvements in care coupled with minimal reduction in the prevalence of CVD [4].

Lifestyle risk factors, including tobacco and alcohol use, poor diet, lack of exercise, and obesity [5], as well as a genetic predisposition to problems such as familial hypercholesterolemia [6] contribute to the high prevalence of CVD. To facilitate prevention of CVD, it is important to measure CVD risk factors on a regular basis. In an effort to reduce CVD mortality, the National Heart Lung and Blood Institute (NHLBI) launched The Heart Truth campaign to raise awareness of CVD risk factors [7]. A year later, the American Heart Association (AHA) adopted the Red Dress symbol and launched its own campaign, Go Red for Women to emphasize the importance of knowing and reducing CVD risk factor values among at-risk females in accordance with clinical guidelines [8].

Patient engagement in appropriate screening and risk factor modification by health care providers is critical in preventing CVD. For example, it has been shown that a key element of dyslipidemia, a low value of high-density lipoprotein cholesterol (HDL-C), is largely an unknown CVD risk factor among the general public [9]. Similar awareness of elements of dyslipidemia has also been shown to vary across different populations [10]. As a result, it is important for patients being screened to be provided with education on CVD risk factors and CVD risk factor modification.

Self-report of risk factors often guides epidemiological studies of disease prevalence [11]. As a result, it is vitally important to establish the accuracy of self-reported values against clinically measured values. Several recent studies have called the accuracy of self-reported data into question, especially for CVD risk factors [11-13]. Some studies have also produced discordant results about the accuracy of self-reported risk factor data for different socioeconomic groups and different geographic locations [14,15]. There has been a general lack of public health investigation of the agreement of self-reported and clinically measured CVD risk factors.

This study has three aims. First, it examines the degree of agreement between self-reported and clinically measured risk factors for CHD (the most prevalent CVD condition representing half of all cardiovascular diseases), including total cholesterol (TC), HDL-C, low-density lipoprotein cholesterol (LDL-C), systolic blood pressure (SBP), diastolic blood pressure (DBP), body mass index (BMI), and diabetes mellitus (DM) status, to understand which risk factors are most accurately reported by community-dwelling adults [16]. Second, it analyzes agreement of self-reported and clinically measured CHD risk factor values according to the Framingham Heart Study's 10-year CHD risk model [17] to determine if those at higher risk of CHD have a greater understanding of their CHD risk factors versus those at lower risk of CHD (ie, the resulting kappa statistics of self-reported and clinically measured agreement are stratified by Framingham 10-year CHD risk). Finally, because self-reported data often present missing data challenges, the study examines whether the method of accounting for missing data influences the results.

Methods

Heart Aware Cardiovascular Health Risk Assessment

Heart Aware is a cardiovascular health risk assessment tool offered by Navigant Consulting, Inc (Navigant Consulting, Inc, Chicago, IL, USA). The data used in this study came from Heart Aware assessments conducted at 127 clinical sites across the United States between January 1, 2006 and June 30, 2010. Research approval for the data was granted by the Institutional Review Board of Texas A&M University. The risk assessment was administered through the Internet. The assessment began by asking respondents, who voluntarily accessed the survey, a series of self-identifying demographic questions about their race, sex, and age. These respondents were recruited to the survey through marketing activities of the hospitals sponsoring the Heart Aware assessment, such as television media advertising campaigns and print media placed throughout both inpatient and outpatient facilities. Respondents were then asked to report their height, weight, whether they used tobacco, level of physical activity, DBP, SBP, TC, HDL-C, and LDL-C. Respondents were then asked to report the last time their health care provider measured their blood pressure, cholesterol, and checked their diabetes status. Finally, respondents were asked a series of questions about their family history, current medications, and current health history, with specific emphasis on CHD symptoms or diagnoses. As the respondents moved through the assessment, 2 unique tools reported data back to the respondents. First, as questions were answered by the respondents, a visual scale indicating the risk for CHD attributable to each risk factor was displayed. Scales for each risk factor were indicated on a color spectrum from green (low risk) to red (critical risk), and were updated as additional information was provided by the respondent. Second, as the respondents moved through different categories of questions, the tool provided the respondents with education about CHD and associated risk factors. This occurred in the form of text boxes on response pages. Definitions of medical terms were also provided to enhance the respondents' knowledge of CHD and related risk factors.

The clinical sites that elected to offer the Heart Aware assessment determined the level of risk that would prompt the clinical site to extend an invitation to the respondent for a free on-site clinical risk assessment in which their self-reported CHD risk factor values would be measured by a clinician for comparison and validation. Each site set its own criteria for inviting participants to the free clinical assessment; Navigant did not document this protocol, including the method of extending the invitation to the participants. Anecdotal evidence suggests respondents were more likely to be invited for a clinical assessment if their self-reported values indicated 2 or more CHD risk factors.

Coronary Heart Disease Risk Factors

The variables analyzed as CHD risk factors were collected from the Heart Aware risk factor assessment. These included self-reported and clinically measured values of TC, HDL-C, LDL-C, SBP, DBP, BMI, as well as DM status and tobacco use. These variables were reported in the clinically assessed dataset as ordinal scales based on their frequencies and clinical guideline ranges. This resulted in ordered categories (referred to as “ranges” below) within each variable as follows: TC (<160 mg/dL, 160-199 mg/dL, 200-239 mg/dL, 240-279 mg/dL, >279 mg/dL), HDL-C (<35 mg/dL, 35-44 mg/dL, 45-49 mg/dL, 50-59 mg/dL, and >59 mg/dL), LDL-C (<100 mg/dL, 100-129 mg/dL, 130-159 mg/dL, 160-189 mg/dL, and >189 mg/dL), SBP (<120 mm Hg, 120-129 mm Hg, 130-139 mm Hg, 140-159 mm Hg, 160-199 mm Hg, and >199 mm Hg), and DBP (<80 mm Hg, 80-84 mm Hg, 85-89 mm Hg, 90-99 mm Hg, 100-114 mm Hg, and >114 mm Hg). DM status and tobacco use were coded as yes or no responses.

Sample Selection Criteria

Respondents who provided a self-reported health assessment and were chosen for and participated in a free clinical assessment of their CHD risk factors were eligible for inclusion in the sample. Those with a prior history of CVD- and CHD-related procedures, such as stroke, acute myocardial infarction, abdominal aortic aneurysm, cardiac arrest, congestive heart failure, angioplasty, catheterization and stent procedures, heart bypass, and carotid procedures, were excluded from the sample to maintain the integrity of the study objective which was to evaluate how community-dwelling adults who had not received a diagnosis of CHD viewed their risk relative to their actual clinically measured risk for CHD.

Depending on the preferred strategy for addressing missing data, 3 samples were available for analysis. To maximize our understanding of the research questions in the study, all 3 samples were made available for analysis. First, analyses were conducted on the original dataset. Given default settings in Stata version 12 (Stata Corp LLP, College Station, TX, USA), this resulted in a listwise deleted sample. Second, analyses were conducted on the original dataset, but with a change in the default settings. Instead of eliminating cases missing any of the variables being analyzed (as was the case with the listwise deleted sample), cases were only removed on a variable-by-variable analysis basis. This resulted in a pairwise deleted sample. Finally, analyses were conducted on the imputed sample. Results of all analyses performed on all 3 samples were then examined.

Self-Assessment Missing Data

Self-reported health status data are known to contain substantial amounts of missing data [18]. Participants often choose not to answer certain questions for a variety of reasons, such as lack of knowledge, time constraints in answering the survey, or a desire not to answer certain questions based on individual preferences. Missing data introduces many analytical challenges, especially relating to biased statistical estimators when making inferences from data [19,20]. To address missing data issues, a 3-step process was used to evaluate the missing data. First, data were analyzed for the degree of missing data, such as the number of missing responses across individual variables and individual cases. Second, the pattern of missing data was analyzed using the “mvpatterns” user-written command [21] in Stata version 12. This command allowed the researcher to determine whether the pattern of missing data was missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR) [20]. Third, based on the result of the first 2 steps noted previously, a methodology was devised to adjust the dataset where appropriate to account for missing data where it was deemed a concern. Several techniques were evaluated for this purpose such as listwise deletion and multiple imputation. Multiple imputation was selected as the preferred method for addressing self-reported missing data because of the large number of missing responses and MAR-identified pattern of missing data discussed later in this paper. Multivariate normal imputation with 5 imputations was used to impute missing values of SBP, DBP, and TC because these variables were most closely aligned with the Framingham 10-year CHD risk model. TC was analyzed because it had more complete self-reported data than HDL-C and LDL-C. These latter variables were key considerations when analyzing missing values in the clinically measured dataset discussed subsequently. Multivariate imputation was chosen because of its ability to take advantage of all variables in the analysis to impute the selected missing variables [22]. Five imputations were selected to balance statistical rigor with processing speed. Independent variables used to impute were selected based on the completeness of data and their theoretical association with the imputed variables. The independent variables used for imputation were age, rural/urban designation, sex, DM status, and BMI range.

The analysis of missing data for the self-reported data was vital to the integrity of the study. Because the self-reported values were used as the basis for choosing individuals for a free clinical assessment, missing data could have profound implications for how participants were selected, possibly resulting in selection bias. Because the clinical sites set their own selection criteria that were not chronicled by Navigant, it was even more important to examine whether the missing self-reported data influenced selection by the clinical sites. This test for selection bias in the clinical assessment process was conducted using the original dataset and the imputed dataset. Using both datasets individually, the analysis was done by examining differences in risk factor variable means between those individuals selected for clinical assessment and those individuals not chosen for clinical assessment. For this purpose, *t* tests with statistical significance determined at the alpha=.05 level were used.

Clinical Assessment Missing Data

Although the statistical challenges of missing data noted for the self-reported data also applied to the clinically measured data, there were additional complexities that necessitated a separate analysis of the clinically measured dataset. First, when analyzing clinically measured data, many of the variables used in the Framingham 10-year CHD risk model were reported as ranges by the clinical sites. These ranges corresponded to the ranges used in the Framingham 10-year CHD risk model. Second, the pattern of missing data in the clinically measured dataset was different than the self-reported dataset. As such, methodologies used to address missing data concerns in the clinically measured dataset had to recognize both unique patterns of missing data and the fact the data were now reflected in ranges unlike individual values found in the self-reported dataset. Given the MAR-identified pattern of missing data discussed later in this paper, missing data in the clinically measured dataset was imputed with an ordered logistic model versus the multivariate normal model discussed previously. The ordered logistic imputation was carried out with 5 imputations and was used to impute missing values for ranges of SBP, DBP, TC, HDL-C, and LDL-C because these variables were most closely aligned with the Framingham 10-year CHD risk model. Note the imputation of the clinically measured dataset imputed HDL-C and LDL-C in addition to TC. This was because these variables provided a more detailed view of dyslipidemia relative to TC. It was possible to do these imputations in the clinically measured dataset because there was not as much missing data relative to the self-reported dataset. Independent variables used to impute were selected based on the completeness of data and their theoretical association with the imputed variables. The independent variables used for imputation were age, rural/urban designation, sex, DM status, and BMI range.

Descriptive Statistics and Kappa Coefficients of Interrater Agreement

Descriptive statistics were conducted on each of the 3 samples. The following variables were dichotomized by sex: age, race/ethnicity, rural/urban designation, SBP, DBP, TC, HDL-C, LDL-C, DM status, and tobacco use. Continuous variables were measured with *t* statistics, and categorical variables were measured with chi-square statistics. Descriptive analyses were performed using Stata version 12. Test statistics were measured for statistical significance at the alpha=.05 level of statistical significance.

Weighted kappa coefficients of interrater agreement between self-reported and clinically measured CHD risk factors were calculated for males and females. A weighted kappa coefficient was used instead of an unweighted kappa coefficient to account for the degree of discordance between each self-reported and clinically measured observation based on the fact the data were categorized as ordinal ranges. The weighting procedure was performed in Stata version 12 using the “wgt(w)” extension of the “kapci” command. The weighted kappa procedure included 100 repetitions. Each coefficient was reported with its standard error and bias-corrected 95% confidence interval. According to established literature [23], the strength of interrater agreement of the kappa coefficient is described as poor ($\kappa<.00$), slight ($\kappa=.00-.20$), fair ($\kappa=.21-.40$), moderate ($\kappa=.41-.60$), substantial ($\kappa=.61-.80$), and almost perfect ($\kappa=.81-1.00$).

Framingham Heart Study’s 10-Year CHD Risk Model

The 10-year CHD risk model from the Framingham Heart Study [17] was used to calculate each respondent’s 10-year CHD risk based on their clinically measured risk factor values. The weighted kappa coefficients were then calculated for each stratum of 10-year CHD risk scores to allow for an evaluation of risk factor agreement by level of CHD risk.

Results

Self-Reported Data, Missing Data, and Participant Selection

The Heart Aware cardiovascular health risk assessment was taken by 373,085 individuals. The clinical sites provided a clinical assessment to 22,346 (5.99%) of these individuals (note the number of individuals offered an assessment but not taking an assessment was not recorded). Among those responding to the self-reported assessment, 238,081 (63.81%) of the respondents did not answer at least 1 of the risk factor variable questions. Based on the robust information presented in the tables subsequently, extended analysis of the results is provided in the discussion section in this paper. Table 1 reports the results of the nonresponse rate for each risk factor variable in the self-reported dataset and in the imputed self-reported dataset.

Table 1. Missing data from the self-reported dataset (n=373,085) and the self-reported imputed dataset (n=373,085).

Variables by dataset	Complete, n	Missing, n	Total, n (n=373,085)	% Missing
Self-reported dataset				
Sex	373,062	23	373,085	0.01%
Age	348,065	25,020	373,085	6.71%
Tobacco use	296,047	77,038	373,085	20.65%
Diabetes status	370,388	2697	373,085	0.72%
Systolic blood pressure	265,693	107,392	373,085	28.78%
Diastolic blood pressure	260,714	112,371	373,085	30.12%
Total cholesterol	189,554	183,531	373,085	49.19%
Complete cases[a]	135,004	238,081	373,085	63.81%
Self-reported imputed dataset				
Sex	373,062	23	373,085	0.01%
Age	348,065	25,020	373,085	6.71%
Tobacco use	296,047	77,038	373,085	20.65%
Diabetes status	370,388	2697	373,085	0.72%
Systolic blood pressure[b]	345,990	27,095	373,085	7.26%
Diastolic blood pressure[b]	345,948	27,137	373,085	7.27%
Total cholesterol[b]	345,569	27,516	373,085	7.38%
Cases with all variables[b]	269,772	103,313	373,085	27.69%

[a] Not missing any variables.

[b] Variables were imputed.

Table 2 reports the results of the test of difference in means for the risk factor variables of those selected for clinical assessment from the self-reported dataset. It should be noted the table compares the clinically measured dataset for each variable unadjusted by imputation methods (ie, each self-reported variable has its own sample size). The sample size and very limited amount of missing data for each variable can be examined in Table 3. Table 3 also reports the same type of information as Table 2, but for those selected for clinical assessment adjusted by imputation methods. It is clear from both these tables that those with clinically measured risk factors had significantly higher values (ie, poorer values) than the overall self-reported population. This confirms the recruitment strategy of the clinics to find those at high risk of CHD and invite them for a free clinical screening of their risk factors.

Table 2. Self-reported and self-reported imputed versus clinically measured dataset: difference in means on critical risk factor values.

Variables by dataset	Sample size	Dataset, mean (SD)		*t* (*df*)[a]	*P*
		Self-reported	Clinically measured		
Self-reported					
Systolic blood pressure	265,693	127.9 (12.1)	134.1 (13.2)	–72.3 (287,545)	<.001
Diastolic blood pressure	260,714	82.3 (6.8)	85.1 (7.0)	–58.2 (282,495)	<.001
Total cholesterol	189,554	193.0 (35.6)	205.6 (30.2)	–49.5 (210,644)	<.001
Self-reported imputed					
Systolic blood pressure	345,990	127.8 (12.2)	132.4 (13.1)	–54.3 (368,243)	<.001
Diastolic blood pressure	345,948	82.4 (6.8)	83.9 (7.0)	–31.8 (368,187)	<.001
Total cholesterol	345,569	192.9 (35.5)	200.9 (33.4)	–32.6 (367,622)	<.001

[a] Pooled degrees of freedom.

Missing Clinical Assessment Data

Clinically measured risk factor values were reported for 22,346 individuals. This dataset consisted of clinically measured data that were expressed as ordinal (ie, ranges of clinical values) and binary data (eg, DM status and tobacco use). Among these individuals, 6423 (28.74%) of the respondents did not answer at least 1 of the risk factor variable questions.

Table 3 reports the results of the nonresponse rate for each risk factor variable in the clinically measured dataset, including the self-reported nonresponse rate for the same respondents. Table 3 also reports the results of the nonresponse rate for each risk factor variable in the imputed clinically measured dataset.

Table 3. Missing data for the clinically measured dataset (n=22,346) and the clinically measured imputed dataset (n=22,346).

Variables in dataset	Clinically measured dataset			Clinical measured imputed dataset		
	Complete, n	Missing, n	% Missing	Complete, n	Missing, n	% Missing
Sex	22,364	0	0.00%	22,364	0	0.00%
Age	22,011	335	1.50%	22,011	335	1.50%
Age (self-reported)	22,011	335	1.50%	22,011	335	1.50%
Tobacco use	22,346	0	0.00%	22,346	0	0.00%
Tobacco use (self-reported)	22,346	0	0.00%	22,346	0	0.00%
Diabetes status	22,346	0	0.00%	22,346	0	0.00%
Diabetes status (self-reported)	22,346	0	0.00%	22,346	0	0.00%
Systolic blood pressure ranges [a]	21,854	492	2.20%	22,255	91	0.41%
Systolic blood pressure ranges (self-reported)[a]	19,465	2881	12.88%	21,981	365	1.63%
Diastolic blood pressure ranges [a]	21,783	563	2.52%	22,241	105	0.47%
Diastolic blood pressure ranges (self-reported)[a]	19,268	3078	13.76%	21,951	395	1.77%
HDL-C ranges [a]	20,105	2241	10.02%	21,850	496	2.22%
HDL-C ranges (self-reported)[a]	9675	12,671	56.66%	20,595	1751	7.83%
LDL-C ranges [a]	16,594	5752	25.72%	21,582	764	3.42%
LDL-C ranges (self-reported)[a]	7829	14,517	64.91%	20,480	1866	8.34%
Total cholesterol ranges [a]	21,092	1254	5.61%	22,055	291	1.30%
Total cholesterol ranges (self-reported)[a]	14,144	8202	36.68%	21,287	1059	4.74%
Cases with all variables [a]	15,923	6423	28.72%	21,241	1105	4.94%
Cases with all variables (self-reported)[a]	7175	15,171	67.84%	20,263	2083	9.31%

[a] These variables in imputed dataset were imputed.

Clinical Assessment Data, Descriptive Statistics: Listwise Deletion, Pairwise Deletion, and Imputed Samples

Table 4 reports the differences in means and proportions of the CHD risk factor variables by sex within the listwise deleted sample (n=5951).

Table 5 reports the differences in means and proportions of the risk factor variables by sex within the pairwise deleted sample (note the sample size varies by each risk factor variable as indicated in the table).

Table 6 reports the differences in means and proportions of the risk factor variables by sex within the imputed sample (n=16,879).

Table 4. Listwise deleted sample (n=5951): differences in means and proportions by sex.

Variables in dataset	Male (n=1938)	Female (n=4013)	Total (n=5951)	*t* (*df*)	χ^2(*df*)	*P* [a]
Age, mean (SD)	55.5 (10.3)	56.3 (11.2)	55.8 (10.6)	–2.6 (5949)		.008
Race/ethnicity, n (%)					32.4 (5)	<.001
Non-Hispanic white	1678 (86.58)	3435 (85.60)	5113 (85.92)			
African-American	114 (5.88)	335 (8.35)	449 (7.54)			
Hispanic	42 (2.17)	112 (2.79)	154 (2.59)			
Asian/Pacific Islander	68 (3.51)	67 (1.67)	135 (2.27)			
American Indian/Alaskan Native	3 (0.15)	8 (0.20)	11 (0.18)			
Other	33 (1.70)	56 (1.40)	89 (1.50)			
Tobacco use, n (%)					33.3 (2)	<.001
Never	1248 (64.40)	2879 (71.74)	4127 (69.35)			
Quit	525 (27.09)	870 (21.68)	1395 (23.44)			
Yes	165 (8.51)	264 (6.58)	429 (7.21)			
Diabetes, n (%)					7.3 (4)	.12
No	1726 (89.06)	3653 (91.03)	5379 (90.39)			
Type 1	16 (0.83)	29 (0.72)	45 (0.76)			
Type 2	184 (9.49)	316 (7.87)	500 (8.40)			
Borderline	7 (0.36)	11 (0.27)	18 (0.30)			
Unsure	5 (0.26)	4 (0.10)	9 (0.15)			
Systolic blood pressure ranges (mm Hg), n (%)					90.8 (5)	<.001
<120	469 (24.20)	1386 (34.54)	1855 (31.17)			
120-129	531 (27.40)	1127 (28.08)	1658 (27.86)			
130-139	461 (23.79)	689 (17.17)	1150 (19.32)			
140-159	394 (20.33)	666 (16.60)	1060 (17.81)			
160-199	79 (4.08)	144 (3.59)	223 (3.75)			
>199	4 (0.21)	1 (0.02)	5 (0.08)			
Self-reported					147.1 (5)	<.001
<120	360 (18.58)	1276 (31.80)	1636 (27.49)			
120-129	750 (38.70)	1514 (37.73)	2264 (38.04)			
130-139	546 (28.17)	750 (18.69)	1296 (21.78)			
140-159	235 (12.13)	406 (10.12)	641 (10.77)			
160-199	45 (2.32)	66 (1.64)	111 (1.87)			
>199	2 (0.10)	1 (0.02)	3 (0.05)			
Diastolic blood pressure ranges (mm Hg), n (%)					107.7 (5)	<.001
<80	829 (42.78)	2206 (54.97)	3035 (51.00)			
80-84	550 (28.38)	1060 (26.41)	1610 (27.05)			
85-89	253 (13.05)	356 (8.87)	609 (10.23)			
90-99	254 (13.11)	323 (8.05)	577 (9.70)			
100-114	47 (2.43)	67 (1.67)	114 (1.92)			
>114	5 (0.26)	1 (0.02)	6 (0.10)			
Self-reported					82.7 (5)	<.001
<80	764 (39.42)	2029 (50.56)	2793 (46.93)			

Variables in dataset	Male (n=1938)	Female (n=4013)	Total (n=5951)	*t* (*df*)	χ^2(*df*)	*P* [a]
80-84	657 (33.90)	1251 (31.17)	1908 (32.06)			
85-89	306 (15.79)	445 (11.09)	751 (12.62)			
90-99	175 (9.03)	240 (5.98)	415 (6.97)			
100-114	31 (1.60)	44 (1.10)	75 (1.26)			
>114	5 (0.26)	4 (0.10)	9 (0.15)			
HDL-C ranges (mg/dL), n (%)					767.1 (4)	<.001
<35	236 (12.18)	1553 (38.70)	1789 (30.06)			
35-44	378 (19.50)	1034 (25.77)	1412 (23.73)			
45-49	323 (16.67)	544 (13.56)	867 (14.57)			
50-59	541 (27.92)	629 (15.67)	1170 (19.66)			
>59	460 (23.74)	253 (6.3)	713 (11.98)			
Self-reported					429.4 (4)	<.001
<35	331 (17.08)	1603 (39.95)	1934 (32.50)			
35-44	514 (26.52)	1112 (27.71)	1626 (27.32)			
45-49	501 (25.85)	693 (17.27)	1194 (20.06)			
50-59	381 (19.66)	448 (11.16)	829 (13.93)			
>59	211 (10.89)	157 (3.91)	368 (6.18)			
LDL-C ranges (mg/dL), n (%)					3.8 (4)	.44
<100	645 (33.28)	1361 (33.91)	2006 (33.71)			
100-129	631 (32.56)	1349 (33.62)	1980 (33.27)			
130-159	439 (22.65)	895 (22.30)	1334 (22.42)			
160-189	162 (8.36)	310 (7.72)	472 (7.93)			
>189	61 (3.15)	98 (2.44)	159 (2.67)			
Self-reported					11.9 (4)	.02
<100	497 (25.64)	1178 (29.35)	1675 (28.15)			
100-129	639 (32.97)	1323 (32.97)	1962 (32.97)			
130-159	544 (30.03)	1013 (25.24)	1557 (26.16)			
160-189	181 (9.34)	365 (9.10)	546 (9.17)			
>189	77 (3.97)	134 (3.34)	211 (3.55)			
Total cholesterol ranges, (mg/dL), n (%)					81.8 (4)	<.001
<160	489 (25.23)	695 (17.32)	1184 (19.90)			
160-199	762 (39.32)	1464 (36.48)	2226 (37.41)			
200-239	497 (25.64)	1307 (32.57)	1804 (30.31)			
240-279	143 (7.38)	438 (10.91)	581 (9.76)			
>279	47 (2.43)	109 (2.72)	156 (2.62)			
Self-reported					53.7 (4)	<.001
<160	420 (21.67)	729 (18.17)	1149 (19.31)			
160-199	763 (39.37)	1351 (33.67)	2114 (35.52)			
200-239	582 (30.03)	1384 (34.49)	1996 (33.04)			
240-279	130 (6.71)	436 (10.86)	566 (9.51)			
>279	43 (2.22)	113 (2.82)	156 (2.62)			

[a] Bonferroni correction was used to determine statistical significance based on 14 comparisons (alpha=.004).

Table 5. Differences in means and proportions by sex in the pairwise deleted sample.

Variables in dataset	N	Male	Female	Total	*t (df)*	χ^2(*df*)	*P* [a]
Age, mean (SD)	18,428	53.7 (12.0)	54.1. (11.5)	54.0 (11.7)	–2.0 (18,426)		.004
Race/ethnicity, n (%)	18,659	5529	13,130	18,659		70.6 (5)	<.001
Non-Hispanic white		4612 (83.41)	10,729 (81.71)	15,341 (82.22)			
African-American		426 (7.70)	1417 (10.79)	1843 (9.88)			
Hispanic		263 (4.76)	610 (4.65)	873 (4.68)			
Asian/Pacific Islander		140 (2.53)	185 (1.41)	325 (1.74)			
American Indian/Alaskan Native		13 (0.24)	44 (0.34)	57 (0.31)			
Other		75 (1.36)	145 (1.10)	220 (1.18)			
Tobacco use, n (%)	18,724	5549	13175	18,724		53.7 (2)	<.001
Never		3545 (63.89)	9138 (69.36)	12683 (67.74)			
Quit		1320 (23.79)	2679 (20.33)	3999 (21.36)			
Yes		684 (12.33)	1358 (10.31)	2042 (10.91)			
Diabetes, n (%)	18,724	5549	13,175	18,724		6.7 (4)	.15
No		5049 (90.99)	12,104 (91.87)	17,153 (91.61)			
Type 1		32 (0.58)	72 (0.55)	104 (0.56)			
Type 2		420 (7.57)	889 (6.75)	1309 (6.99)			
Borderline		15 (0.27)	49 (0.37)	64 (0.34)			
Unsure		33 (0.59)	61 (0.46)	94 (0.50)			
Systolic blood pressure ranges (mm Hg), n (%)	18,270	5407	12,863	18,270		210.9 (5)	<.001
<120		1283 (23.73)	4350 (33.82)	5633 (30.83)			
120-129		1466 (27.11)	3416 (26.56)	4882 (26.72)			
130-139		1221 (22.58)	2270 (17.65)	3491 (19.11)			
140-159		1141 (22.10)	2268 (17.63)	3409 (18.66)			
160-199		284 (5.25)	544 (4.23)	828 (4.53)			
>199		12 (0.22)	15 (0.12)	27 (0.15)			
Self-reported	16,171	4710	11,461	16,171		274.8 (5)	<.001
<120		838 (17.79)	3339 (29.13)	4177 (25.83)			
120-129		1706 (36.22)	4097 (35.75)	5803 (35.89)			
130-139		1250 (26.54)	2253 (19.66)	3503 (21.66)			
140-159		737 (15.65)	1469 (12.82)	2206 (13.64)			
160-199		169 (3.59)	291 (2.54)	460 (2.84)			
>199		10 (0.21)	12 (0.10)	22 (0.14)			
Diastolic blood pressure ranges (mm Hg), n (%)	18,202	5392	12,810	18,202		338.2 (5)	<.001
<80		2161 (40.08)	6833 (53.34)	8994 (49.41)			
80-84		1511 (28.02)	3275 (25.57)	4786 (26.29)			
85-89		701 (13.0)	1184 (9.24)	1885 (10.36)			
90-99		819 (15.19)	1252 (9.77)	2071 (11.38)			
100-114		178 (3.30)	251 (1.96)	429 (2.36)			
>114		22 (0.41)	15 (0.12)	37 (0.20)			
Self-reported	16,013	4685	11,328	16,013		216.5 (5)	<.001
<80		1611 (34.39)	5167 (45.61)	6778 (42.33)			

Variables in dataset	N	Male	Female	Total	*t* (*df*)	χ^2(*df*)	*P* [a]
80-84		1576 (33.64)	3581 (31.61)	5157 (32.21)			
85-89		834 (17.80)	1467 (12.95)	2301 (14.37)			
90-99		536 (11.44)	907 (8.01)	1443 (9.01)			
100-114		104 (2.2)	180 (1.59)	284 (1.77)			
>114		24 (0.51)	26 (0.23)	50 (0.31)			
HDL-C ranges (mg/dL), n (%)	16,779	5048	11,731	16,779		2,200.0 (4)	<.001
<35		539 (10.68)	4162 (35.48)	4701 (28.02)			
35-44		829 (16.42)	2947 (25.12)	3776 (22.50)			
45-49		680 (13.47)	1581 (13.48)	2261 (13.48)			
50-59		1575 (31.20)	2129 (18.15)	3704 (22.8)			
>59		1425 (28.23)	912 (7.7)	2337 (13.93)			
Self-reported	7691	2518	5173	7691		617.5 (4)	<.001
<35		399 (15.85)	2022 (15.3)	2421 (31.48)			
35-44		627 (24.90)	1409 (10.7)	2036 (26.47)			
45-49		626 (24.86)	873 (6.6)	1499 (19.49)			
50-59		509 (20.21)	622 (4.7)	1131 (14.71)			
>59		357 (14.18)	247 (4.77)	604 (7.85)			
LDL-C ranges (mg/dL), n (%)	14,476	4370	10106	14,476		16.8 (4)	<0.001
<100		1336 (30.57)	3334 (32.99)	4670 (32.26)			
100-129		1476 (33.78)	3455 (34.19)	4931 (34.06)			
130-159		1006 (23.02)	2227 (22.04)	3233 (22.33)			
160-189		396 (9.06)	801 (7.93)	1197 (8.27)			
>189		156 (3.57)	289 (2.86)	445 (3.07)			
Self-reported	6769	2196	4573	6769		12.9 (4)	.01
<100		553 (25.18)	1312 (28.69)	1865 (27.55)			
100-129		717 (32.65)	1517 (33.17)	2234 (33.00)			
130-159		610 (27.78)	1141 (24.95)	1751 (25.87)			
160-189		218 (9.93)	418 (9.14)	636 (9.40)			
>189		98 (4.46)	185 (4.05)	283 (4.18)			
Total cholesterol ranges (mg/dL), n (%)	17,627	5265	12,362	17,627		137.9 (4)	<.001
<160		1128 (21.42)	1901 (15.38)	3029 (17.18)			
160-199		2050 (38.94)	4628 (37.44)	6678 (37.89)			
200-239		1467 (27.86)	4100 (33.17)	5567 (31.58)			
240-279		450 (8.55)	1371 (11.09)	1821 (10.33)			
>279		170 (3.23)	362 (2.93)	532 (3.02)			
Self-reported	11,541	3495	8046	11,541		49.8 (4)	<.001
<160		661 (18.91)	1281 (15.92)	1942 (16.83)			
160-199		1292 (36.97)	2687 (33.40)	3979 (34.48)			
200-239		1139 (32.59)	2900 (36.04)	4039 (35.00)			
240-279		291 (8.33)	898 (11.16)	1189 (10.30)			
>279		112 (3.20)	280 (3.48)	392 (3.40)			

[a] Bonferroni correction was used to determine statistical significance based on 14 comparisons (alpha=.004).

Table 6. Differences in means and proportions by sex for imputed sample (n=16,879).

Variables in dataset	Male (n=5011)	Female (n=11,868)	Total (n=16,879)	*t* (*df*)	χ^2(*df*)	*P* [a]
Age, mean (SD)	53.7 (12.1)	54.2 (11.5)	54.0 (11.7)	2.4 (16,877)		.01
Race/ethnicity, n (%)					70.5 (5)	<.001
Non-Hispanic white	4168 (83.18)	9648 (81.29)	13,816 (81.85)			
African-American	385 (7.68)	1321 (11.13)	1706 (10.11)			
Hispanic	245 (4.89)	552 (4.65)	797 (4.72)			
Asian/Pacific Islander	119 (2.37)	161 (1.36)	280 (1.66)			
American Indian/ Alaskan Native	12 (0.24)	43 (0.36)	55 (0.33)			
Other	82 (1.64)	143 (1.20)	225 (1.33)			
Tobacco use, n (%)					56.7 (2)	<.001
Never	3190 (63.66)	8257 (69.57)	11,447 (67.82)			
Quit	1220 (24.35)	2398 (20.21)	3618 (21.43)			
Yes	601 (11.99)	1213 (10.22)	1814 (10.75)			
Diabetes, n (%)					5.4 (4)	.25
No	4557 (90.94)	10883 (91.70)	15,440 (91.47)			
Type 1	27 (0.54)	60 (0.51)	87 (0.52)			
Type 2	383 (7.64)	826 (6.96)	1209 (7.16)			
Borderline	14 (0.28)	46 (0.39)	60 (0.36)			
Unsure	30 (0.60)	53 (0.45)	83 (0.49)			
Systolic blood pressure ranges (mm Hg), n (%)					182.7 (5)	<.001
<120	1220 (24.35)	4049 (34.12)	5269 (31.22)			
120-129	1352 (26.98)	3131 (26.38)	4483 (26.56)			
130-139	1130 (22.55)	2096 (17.66)	3226 (19.11)			
140-159	1043 (20.81)	2087 (17.59)	3130 (18.54)			
160-199	255 (5.09)	493 (4.15)	748 (4.43)			
>199	11 (0.22)	12 (0.10)	23 (0.14)			
Self-reported					226.4 (5)	<.001
<120	931 (18.58)	3428 (28.88)	4359 (25.82)			
120-129	1833 (36.58)	4204 (35.42)	6037 (35.77)			
130-139	1288 (25.70)	2368 (19.95)	3656 (21.66)			
140-159	772 (15.41)	1530 (12.89)	2302 (13.64)			
160-199	176 (3.51)	325 (2.74)	501 (2.97)			
>199	11 (0.22)	13 (0.11)	24 (0.14)			
Diastolic blood pressure ranges (mm Hg), n (%)					294.3 (5)	<.001
<80	2042 (40.75)	6370 (53.67)	8412 (49.84)			
80-84	1401 (27.96)	3008 (25.35)	4409 (26.12)			
85-89	640 (12.77)	1087 (9.16)	1727 (10.23)			
90-99	742 (14.81)	1154 (9.72)	1896 (11.23)			
100-114	165 (3.29)	235 (1.98)	400 (2.37)			
>114	21 (0.42)	14 (0.12)	35 (0.21)			
Self-reported					172.3 (5)	<.001
<80	1784 (35.60)	5376 (45.30)	7160 (42.42)			

Variables in dataset	Male (n=5011)	Female (n=11,868)	Total (n=16,879)	*t* (*df*)	χ^2(*df*)	*P* [a]
80-84	1665 (33.23)	3722 (31.36)	5387 (31.92)			
85-89	875 (17.46)	1568 (13.21)	2443 (14.47)			
90-99	546 (10.90)	987 (8.32)	1533 (9.08)			
100-114	114 (2.27)	189 (1.59)	303 (1.80)			
>114	27 (0.54)	26 (0.22)	53 (0.31)			
HDL-C ranges (mg/dL), n (%)					1800.0 (4)	<.001
<35	594 (11.85)	4100 (34.55)	4694 (27.81)			
35-44	844 (16.84)	2947 (24.83)	3791 (22.46)			
45-49	672 (13.41)	1568 (13.21)	2240 (13.27)			
50-59	1543 (30.79)	2248 (18.94)	3791 (22.46)			
>59	1358 (27.10)	1005 (8.47)	2363 (14.00)			
Self-reported					295.9 (4)	<.001
<35	1101 (21.97)	3994 (33.65)	5095 (30.19)			
35-44	1262 (25.18)	3085 (25.99)	4347 (25.75)			
45-49	1148 (22.91)	2180 (18.37)	3328 (19.72)			
50-59	913 (18.22)	1693 (14.27)	2606 (15.44)			
>59	587 (11.71)	916 (7.72)	1503 (8.90)			
LDL-C ranges (mg/dL), n (%)					9.7 (4)	0.046
<100	1554 (31.01)	3935 (33.16)	5489 (32.52)			
100-129	1713 (34.18)	4007 (33.76)	5720 (33.89)			
130-159	1137 (22.69)	2621 (22.08)	3758 (22.26)			
160-189	444 (8.86)	962 (8.11)	1406 (8.33)			
>189	163 (3.25)	343 (2.89)	506 (3.00)			
Self-reported					9.8 (4)	.04
<100	1296 (25.86)	3307 (27.86)	4603 (27.27)			
100-129	1656 (33.05)	3922 (33.05)	5578 (33.05)			
130-159	1357 (27.08)	3005 (25.32)	4362 (25.84)			
160-189	470 (9.38)	1122 (9.45)	1582 (9.43)			
>189	232 (4.63)	522 (4.40)	754 (4.47)			
Total cholesterol ranges (mg/dL), n (%)					130.8 (4)	<.001
<160	1085 (21.65)	1852 (15.60)	2937 (17.40)			
160-199	1941 (38.73)	4421 (37.25)	6362 (37.69)			
200-239	1389 (27.72)	3920 (33.03)	5309 (31.45)			
240-279	434 (8.66)	1330 (11.21)	1764 (10.45)			
>279	162 (3.23)	345 (2.91)	507 (3.00)			
Self-reported					28.1 (4)	<.001
<160	938 (18.72)	2002 (16.87)	2940 (17.42)			
160-199	1787 (35.66)	3972 (33.47)	5759 (34.12)			
200-239	1657 (33.07)	4150 (34.97)	5807 (34.40)			
240-279	469 (9.36)	1336 (11.26)	1805 (10.69)			
>279	160 (3.19)	408 (3.44)	568 (3.37)			

[a] Bonferroni correction was used to determine statistical significance based on 14 comparisons (alpha=.004)

Weighted Kappa Agreement and Agreement by Risk Stratification

Listwise Deletion

Table 7 reports the results of the weighted kappa interrater agreement analysis by sex for each CHD risk factor variable in the listwise deleted sample. In addition to the weighted kappa statistic, its standard error and bias-corrected 95% confidence interval was reported along with an estimate of the 10-year CHD risk score for each variable's strata of clinical values.

Table 7. Interrater agreement of self-reported and clinically measured Framingham 10-year CHD risk factors by risk score for the listwise deleted sample (n=5951).

Clinically measured risk factor	Males (n=1938)					Females (n=4013)				
	Risk score[a]	n	Weighted kappa statistics			Risk score[a]	n	Weighted kappa statistics		
			κ	SE	95% CI[b]			κ	SE	95% CI[b]
Total cholesterol		1938	.56	.01	.53-.59	4013	4013	.59	.01	.57-.61
	<3%	518	.52	.03	.46-.57	1123	1123	.64	.02	.60-.67
	3-4%	538	.59	.02	.53-.62	1157	1157	.53	.02	.49-.56
	5-6%	524	.54	.03	.48-.60	1178	1178	.55	.02	.51-.59
	>6%	358	.55	.03	.50-.62	555	555	.53	.03	.48-.59
HDL-C		1938	.49	.02	.46-.52	4013	4013	.57	.01	.56-.60
	<3%	518	.56	.03	.50-.60	1123	1123	.60	.02	.57-.64
	3-4%	538	.54	.02	.49-.58	1157	1157	.55	.02	.53-.59
	5-6%	524	.44	.03	.38-.48	1178	1178	.58	.02	.54-.61
	>6%	358	.33	.03	.25-.37	555	555	.45	.03	.41-.50
LDL-C		1938	.58	.02	.55-.62	4013	4013	.58	.01	.55-.60
	<3%	518	.48	.03	.43-.54	1123	1123	.60	.02	.57-.65
	3-4%	538	.58	.03	.52-.63	1157	1157	.49	.02	.45-.53
	5-6%	524	.60	.03	.53-.67	1178	1178	.55	.02	.51-.59
	>6%	358	.57	.03	.50-.63	555	555	.52	.03	.48-.58
Systolic blood pressure		1938	.42	.01	.38-.44	4013	4013	.45	.01	.42-.47
	<3%	518	.47	.03	.40-.52	1123	1123	.51	.02	.48-.56
	3-4%	538	.44	.03	.37-.49	1157	1157	.42	.02	.39-.46
	5-6%	524	.39	.03	.33-.47	1178	1178	.40	.02	.35-.43
	>6%	358	.34	.04	.28-.41	555	555	.36	.03	.31-.43
Diastolic blood pressure		1938	.42	.01	.40-.45	4013	4013	.43	.01	.43-.46
	<3%	518	.49	.03	.43-.55	1123	1123	.50	.02	.46-.54
	3-4%	538	.41	.03	.35-.47	1157	1157	.42	.02	.38-.46
	5-6%	524	.38	.03	.31-.43	1178	1178	.38	.02	.33-.43
	>6%	358	.40	.04	.33-.47	555	555	.38	.03	.31-.43
Tobacco user [c]		165				264				
	≤5%	79				156				
	>5%	86				108				
Diagnosed with diabetes [c]		200				345				
	≤5%	95				179				
	>5%	105				166				

[a] Framingham 10-year CHD risk score.

[b] Bias-corrected 95% CI.

[c] Interrater agreement not measured.

Pairwise Deletion

Table 8 reports the results of the weighted kappa interrater agreement analysis by sex for each CHD risk factor variable in the pairwise deleted sample. In addition to the weighted kappa statistic, its standard error and bias-corrected 95% confidence interval was reported along with an estimate of the 10-year CHD risk score for each variable’s strata of clinical values.

Table 8. Interrater agreement of self-reported and clinically measured Framingham 10-year CHD risk factors by risk score for the pairwise deleted sample.

Clinically measured risk factor	Males (n=varies)					Females (n=varies)				
	Risk score[a]	n	Weighted kappa statistics			Risk score[a]	n	Weighted kappa statistics		
			κ	SE	95% CI[b]			κ	SE	95% CI[b]
Total cholesterol		4524	.61	.01	.59-.63		9547	.58	.01	.57-.60
	<3%	1038	.57	.02	.53-.61	<4%	2410	.62	.01	.59-.63
	3-4%	1176	.61	.02	.57-.64	4-6%	2523	.52	.01	.50-.55
	5-6%	1168	.59	.02	.56-.62	7-8%	3009	.57	.01	.55-.59
	>6%	1142	.64	.02	.61-.69	>8%	1605	.53	.02	.53-.60
HDL-C		3374	.58	.01	.55-.60		6217	.64	.01	.63-.65
	<3%	773	.63	.02	.58-.68	<4%	1600	.65	.01	.62-.68
	3-4%	861	.62	.02	.58-.66	4-6%	1649	.60	.02	.56-.63
	5-6%	893	.54	.02	.50-.57	7-8%	1948	.65	.01	.63-.69
	>6%	847	.47	.03	.43-.52	>8%	1020	.57	.02	.54-.61
LDL-C		2624	.58	.01	.56-.60		4989	.58	.01	.56-.60
	<3%	627	.49	.03	.44-.53	<4%	1318	.58	.02	.54-.61
	3-4%	707	.59	.02	.54-.64	4-6%	1441	.50	.02	.47-.53
	5-6%	697	.61	.03	.56-.66	7-8%	1497	.55	.02	.51-.58
	>6%	593	.55	.03	.49-.60	>8%	733	.54	.02	.50-.58
Systolic blood pressure		5951	.47	.01	.45-.49		13,381	.47	.01	.46-.48
	<3%	1439	.46	.02	.42-.50	<4%	3744	.49	.01	.46-.50
	3-4%	1552	.45	.02	.42-.49	4-6%	3432	.41	.01	.39-.43
	5-6%	1472	.44	.02	.42-.47	7-8%	3967	.44	.01	.43-.46
	>6%	1488	.46	.02	.43-.50	>8%	2238	.43	.01	.40-.45
Diastolic blood pressure		5901	.47	.01	.45-.49		13,199	.43	.01	.42-.45
	<3%	1430	.44	.02	.41-.49	<4%	3709	.46	.01	.44-.48
	3-4%	1539	.45	.02	.42-.49	4-6%	3385	.40	.01	.38-.44
	5-6%	1464	.46	.02	.43-.50	7-8%	3905	.42	.01	.40-.44
	>6%	1468	.51	.02	.48-.54	>8%	2200	.43	.02	.40-.45
Tobacco user [c]		876					1591			
	≤5%	409				≤7%	887			
	>5%	467				>7%	704			
Diagnosed with diabetes [c]		693					1256			
	≤5%	224				≤7%	511			
	>5%	469				>7%	745			

[a] Framingham 10-year CHD risk score.

[b] Bias-corrected 95% CI.

[c] Interrater agreement not measured.

Imputation

Table 9 reports the results of the weighted kappa interrater agreement analysis by sex for each CHD risk factor variable in the imputed sample. In addition to the weighted kappa statistic, its standard error and bias-corrected 95% confidence interval was reported along with an estimate of the 10-year CHD risk score for each variable's strata of clinical values.

Table 9. Imputed sample (n=16,879): interrater agreement of self-reported and clinically measured Framingham 10-year CHD risk factors by risk score.

Clinically measured risk factor	Males (n=5011)					Females (n=11,868)				
	Risk score[a]	n	Weighted kappa statistics			Risk score[a]	n	Weighted kappa statistics		
			κ	SE	95% CI[b]			κ	SE	95% CI[b]
Total cholesterol		5011	.36	.01	.34-.38		11,868	.35	.01	.34-.37
	<3%	1406	.30	.02	.26-.33	<3%	2902	.34	.01	.31-.36
	3-4%	1288	.38	.02	.35-.43	3-5%	3071	.36	.01	.33-.38
	5-6%	1207	.37	.02	.33-.41	6-8%	3237	.34	.01	.32-.38
	>6%	1110	.37	.02	.34-.44	>8%	2658	.33	.01	.30-.36
HDL-C		5011	.25	.01	.23-.27		11,868	.28	.01	.27-.29
	<3%	1406	.28	.02	.25-.31	<3%	2902	.28	.01	.25-.30
	3-4%	1288	.27	.02	.23-.30	3-5%	3071	.36	.01	.33-.38
	5-6%	1207	.24	.02	.21-.27	6-8%	3237	.27	.01	.25-.30
	>6%	1110	.18	.02	.15-.21	>8%	2658	.20	.01	.18-.22
LDL-C		5011	.24	.01	.22-.27		11,868	.21	.01	.20-.23
	<3%	1406	.18	.02	.14-.21	<3%	2902	.19	.01	.16-.21
	3-4%	1288	.25	.02	.21-.28	3-5%	3071	.23	.01	.20-.25
	5-6%	1207	.29	.02	.26-.33	6-8%	3237	.21	.01	.19-.24
	>6%	1110	.25	.02	.19-.28	>8%	2658	.22	.01	.19-.23
Systolic blood pressure		5011	.38	.01	.36-.41		11,868	.39	.01	.37-.40
	<3%	1406	.38	.02	.34-.41	<3%	2902	.33	.01	.30-.35
	3-4%	1288	.35	.02	.31-.39	3-5%	3071	.34	.01	.31-.36
	5-6%	1207	.35	.02	.31-.39	6-8%	3237	.31	.01	.29-.33
	>6%	1110	.33	.02	.31-.37	>8%	2658	.30	.01	.28-.33
Diastolic blood pressure		5011	.37	.01	.34-.39		11,868	.35	.01	.33-.36
	<3%	1406	.35	.02	.32-.38	<3%	2902	.34	.02	.32-.38
	3-4%	1288	.37	.02	.32-.41	3-5%	3071	.35	.01	.33-.38
	5-6%	1207	.37	.02	.32-.41	6-8%	3237	.33	.02	.30-.36
	>6%	1110	.38	.02	.34-.42	>8%	2658	.31	.01	.28-.33
Tobacco user [c]		601					1213			
	≤5%	295				≤6%	629			
	>5%	306				>6%	584			
Diagnosed with diabetes [c]		410					886			
	<5%	153				≤6%	279			
	>5%	257				>6%	607			

[a] Framingham 10-year CHD risk score.

[b] Bias-corrected 95% CI.

[c] Interrater agreement not measured.

When evaluating the trends of interrater agreement between self-reported and clinically measured CHD risk factors, it is important to evaluate both the baseline interrater agreement coefficients for the entire sample and the individual interrater agreement coefficients for the strata based on 10-year CHD risk. Further, it is important to examine the changes in the interrater agreement as 10-year CHD risk increases.

Although there are some noteworthy differences between the listwise deleted and pairwise deleted samples (eg, the deterioration of interrater agreement by strata for SBP and DBP as 10-year CHD risk increases among males in the listwise deleted sample but not the pairwise deleted sample), the main outcome of interest is the difference in baseline interrater agreement coefficients of the imputed sample versus the listwise and pairwise deleted samples. Overall, the baseline interrater agreement coefficient values for each risk factor in the imputed sample were markedly lower than their counterparts in the listwise and pairwise deleted samples. For example, among males in the listwise deleted sample, the interrater agreement coefficient of self-reported and clinically measured ranges of HDL-C was kappa=.49. By comparison, the same coefficient in the imputed sample was .25. This discrepancy was substantial across the risk factor values with the largest amount of missing data (ie, ranges of TC, HDL-C, and LDL-C). By comparison, the differences in interrater agreement coefficients of variables other than ranges of TC, HDL-C, and LDL-C between the listwise deleted and imputed samples were minor. For example, among females in the listwise deleted sample, the interrater agreement coefficient of self-reported and clinically measured SBP was .45. By comparison, the same coefficient in the imputed sample was .39.

As discussed previously, one of the CHD risk factors thought to be less understood by community-dwelling adults is HDL-C. It is noteworthy that both males and females with the highest 10-year risk of CHD in the imputation sample had the lowest level of interrater agreement between self-reported and clinically measured ranges of HDL-C. In fact, the level of agreement can only be characterized as slight, which is a suboptimal level of agreement. Although the difference between a 3% 10-year risk of CHD and an 8% 10-year risk of CHD may seem numerically immaterial, it should be noted these figures are derived from Framingham's clinical risk model [24], which means the difference between 3% and 8% is more than twice the mortality risk of CHD in the next 10 years. Thus, the difference is clinically relevant.

Conversely, interrater agreement of self-reported and clinically measured ranges of LDL-C slightly increased in both sexes as 10-year CHD risk increased. This is consistent with the layperson hypothesis that individuals with higher risk of CHD would be more conscious of LDL-C because it is often referred to as "bad" cholesterol. This finding is also supported with recent evidence suggesting diabetes patients who recall their most recent LDL-C values are more likely to maintain optimal hemoglobin A1C values [25]. LDL-C could simply be the metric noted by community-dwelling adults as the most important metric to gauge in order to avoid CHD and related diseases. This is certainly consistent with how patients have been conditioned to assume LDL-C is bad cholesterol and HDL-C is good cholesterol (a belief that is the subject of rigorous investigation) [26]. If HDL-C is eventually deemed to be just as clinically important as LDL-C, a substantial public health information campaign may be necessary to inculcate this knowledge and its importance among a public much more likely to appreciate CHD risk due to LDL-C.

Sensitivity Missing Data Techniques

Upon examining the differences of interrater agreement coefficients by the approach used to address missing data, 2 things become apparent. First, ranges of SBP, DBP, and both tobacco use and DM status were not substantially different based on the approach employed to account for missing data. This was mostly because of fewer instances of missing data than other variables in the original dataset. As such, it is appropriate to use any of the 3 samples to establish findings about interrater agreement relative to these variables in the study. However, given the significant amount of missing data for ranges of TC, HDL-C, and LDL-C, the multiple imputation strategy resulted in more conservative results of interrater agreement than the listwise and pairwise deleted samples. As such, the researcher is cautioned to use these figures when establishing findings from the study. Because of these facts, the multiple imputation sample was deemed the most appropriate for discussing findings of this study. This is because the imputation sample was conservative on the variables with greatest instances of missing data, but consistent with the other 2 methodologies for the variables with fewer instances of missing data.

Sample Versus Population-Based Coronary Heart Disease Risk Data

Figure 1 illustrates the comparative 10-year CHD risk score as established by the Framingham Heart Study [24] for the general male population by age group. The results from the listwise deleted and imputed samples, respectively, are also shown for comparison. Figure 2 illustrates the comparative 10-year CHD risk score as established by the Framingham Heart Study [24] for the general female population by age group. The results from the listwise deleted and imputed samples, respectively, are also shown for comparison.

Figure 1. Comparison of CHD risk by sample type relative to the overall population for males.

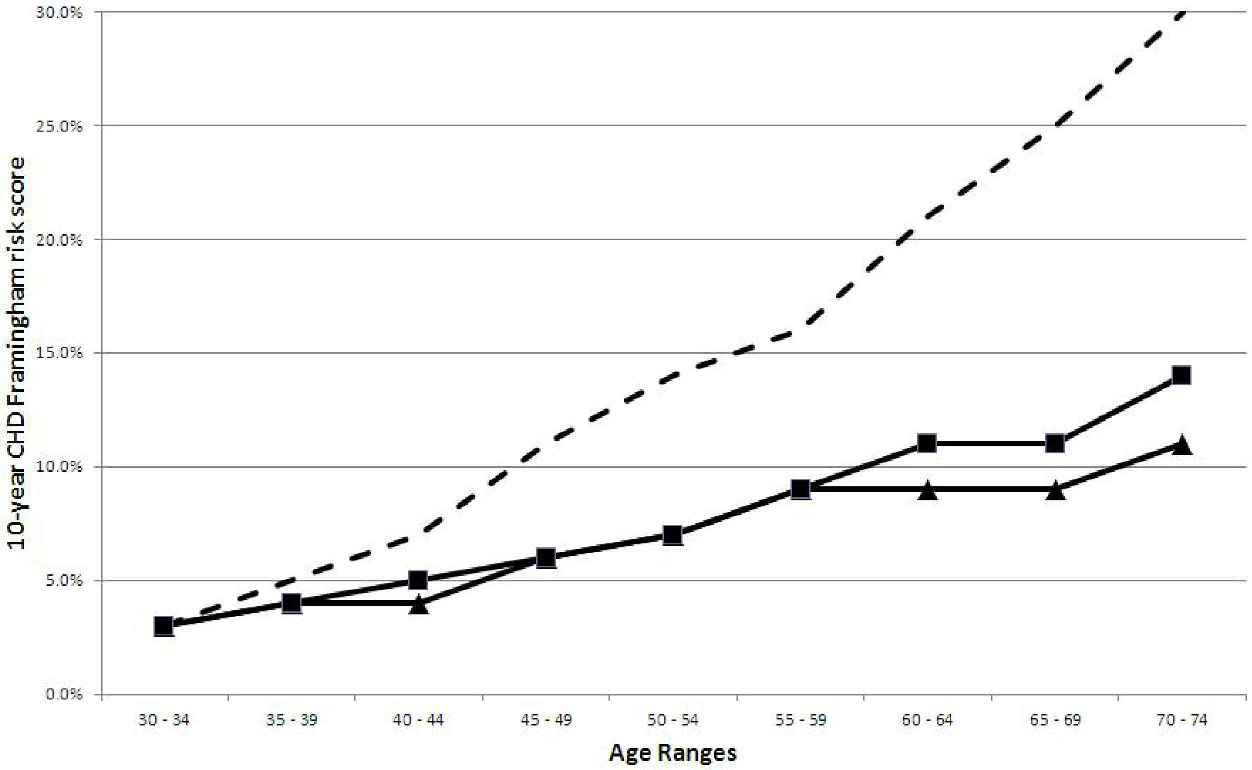
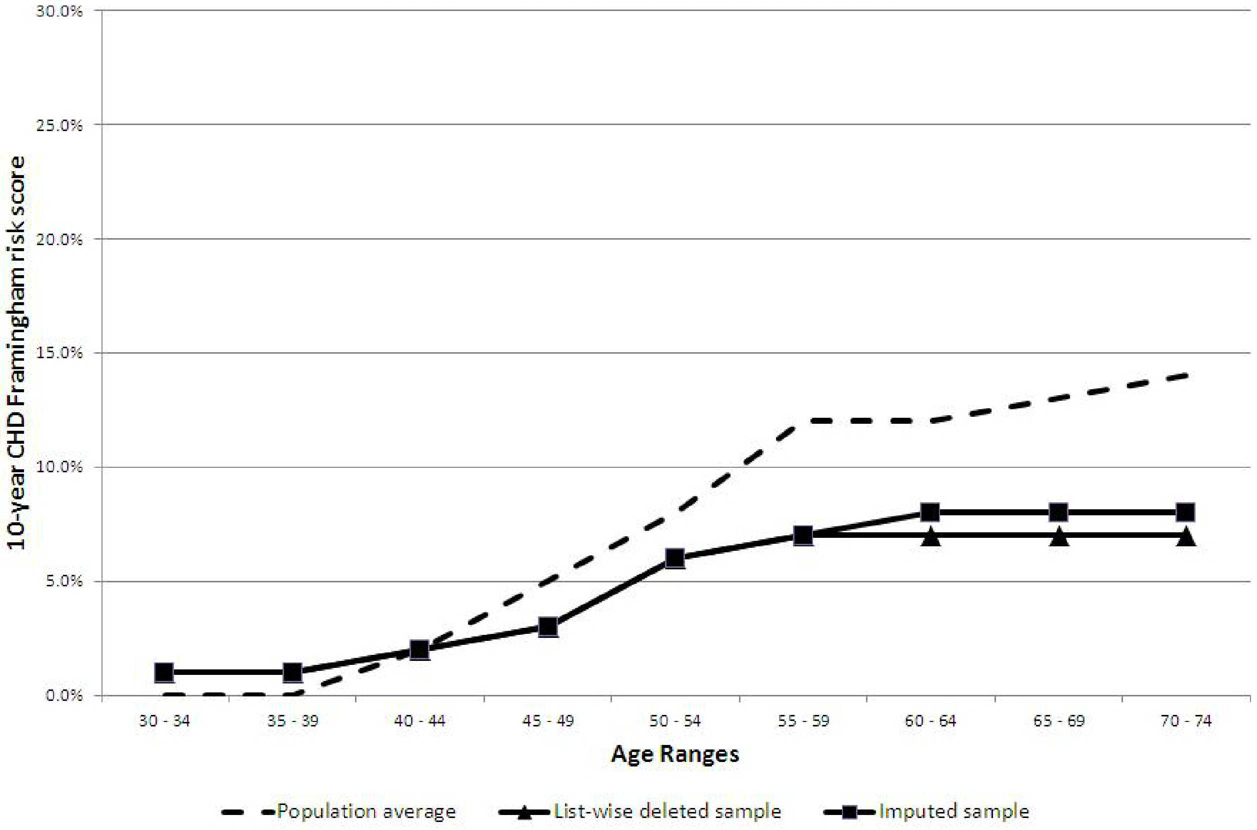


Figure 2. Comparison of CHD risk by sample type relative to the overall population for females.



Discussion

Heart Aware May Underestimate Population Risk for Coronary Heart Disease

The most significant finding from this study is the fact a community-based health risk assessment for heart disease that is delivered via the Internet (Heart Aware) yields a sample with markedly lower risk of CHD than suggested by population health data. For example, Figure 1 demonstrates that despite the method used to account for missing data, those males participating in the Heart Aware assessment had, on average, a 10-year CHD risk that was up to 19 percentage points lower than their counterparts of the same age. Likewise, Figure 2 demonstrates that despite the method used to account for missing data, those females participating in the Health Aware assessment had, on average, a 10-year CHD risk that was up to 7 percentage points lower than their counterparts of the same age.

It should be noted this problem is compounded by the fact participants selected for clinical evaluation in this study were hand-picked by the individual hospitals based on their perceived high risk of CHD (ie, a random selection of community-dwelling adults for clinical measurement of CHD risk factors would likely result in samples with lower CHD risk, thereby exacerbating the differences between the risk of samples established by the Heart Aware assessment versus population health data), recognizing that anyone with diagnosed CHD was excluded from the study.

There are several reasons that could explain these discrepancies. The Heart Aware assessments were offered almost entirely via the Internet. This probably resulted in biased selection of participants because those using the Internet are generally more technically savvy, have higher levels of education and income, and are comparatively healthier than non-Internet users [27]. As such, it is not surprising the tool procured a lower-risk population that is not representative of the general population. This raises a very important issue about health risk assessments such as Heart Aware. Obviously, recruitment cost is greatly reduced by using the Internet, especially through a hospital's existing Internet presence. However, if the data are to be used for public health purposes, how can the data be more representative of the population? One approach could be to expand the methods used to collect the same data, such as using in-clinic kiosks to collect data versus relying on a participant having Internet resources available at home. If financial resources were available, the instrument could be made available through a random-digit dial survey. This method has been shown to improve validity in other CVD-related studies [28]. Finally, a simple and cost-effective method could be to use propensity score matching to create appropriate comparison groups for analysis. This approach is also common in CVD-related studies [29]. However, in the case of the current Heart Aware survey, additional variables need to be collected to account for the underlying demographic differences that are likely associated with the Internet selection bias discussed previously (eg, income, education level, and current health care utilization).

It is difficult to understand the influence of the selection issue on the study results. Although we know the samples in this study were of individuals of much lower CHD risk relative to population health data, and we did exclude those who had already been diagnosed with CHD, the relationship between levels of CHD risk and knowledge of CHD risk factors has yet to be firmly established. In fact, in this study, differences in the interrater agreement of self-reported and clinically measured CHD risk factors varied by sex, individual risk factor, and overall 10-year CHD risk stratum.

Sex-Based Differences Were Not Apparent

As noted previously, the Red Dress symbol and Go Red for Women campaign have been high profile efforts to highlight the fact more females of all ages die of CHD than any other cause of death [30]. Yet, the results of this study indicate a very similar level of awareness of CHD risk factors among the sexes. It is difficult to reconcile this data with the potential success or failure of these high profile campaigns specific to females. Although there may be no marked difference in the interrater agreement between males and females on each measured CHD risk factor in the imputed sample, perhaps awareness would have been worse without the public campaigns focused on females?

What is clear about the differences in the interrater agreement of self-reported and clinically measured CHD risk factors by sex is neither sex has demonstrated a superior understanding of their CHD risk factors. Both sexes demonstrate relatively low levels of agreement on every CHD risk factor. It should also be noted females were generally “healthier” than their male counterparts in this study (see Table 6). It seems rational such a difference would influence CHD risk factor agreement by sex to a greater degree than witnessed in this study. This is an area for continued exploration because it is central to public health policy.

Limitations

In addition to the noteworthy findings of this study, there are several limitations. First, the most substantial limitation is a challenge to internal validity of the results based on a substantial amount of selection bias that was likely the result of the recruitment method (ie, offering the survey to any interested party through media advertising). However, as Guba [31] reminds us in a classic work on naturalistic studies, the process of determining validity is not comparable with rationalistic designs such as randomized controlled trials. Naturalistic trials have a wide array of tools to complement the rationalistic approach to establishing comparable levels of study integrity and quality [31]. Among these methods are techniques such as triangulation of results, replication, and comprehensive descriptive statistics to ensure a thorough understanding of the sample [31]. If naturalistic designs are fundamentally characterized as research conducted in natural settings versus structured environments such as laboratories [32], then Heart Aware should qualify as a tool used in a naturalistic setting. Because of this paradigm, the study does accomplish some of the processes desired by naturalistic researchers, such as the exhaustive approach to examining missing data, the use of multiple imputation to ensure replication of the results displayed in the imputed sample, and the 100 repetitions conducted on each weighted kappa analysis of interrater agreement between self-reported and clinically measured CHD risk factors. Nevertheless, future research in this area should incorporate some of the suggestions made previously to counteract the apparent selection bias of solely using the Internet for recruitment.

The second limitation is the amount of missing data. Although this study has attempted to mitigate this point with multiple approaches, none of these efforts can fully account for the bias that exists in statistical estimation as a result of missing responses. At a very basic level, the latent traits of missing responses remain unknown even with the most sophisticated missing data techniques. However, it should be noted repetition and replication (as noted previously) somewhat mitigate these biases.

The third limitation of the study is the difference between sexes for baseline health behaviors and clinical values. Although some of the clinical value differences are because of normal differences based on sex, some of the discrepancies are very large indicating females are probably healthier than their male counterparts. This influences the ability to fully understand results of the study by sex.

The fourth limitation of the study is the lack of research on how respondents acquired information about their self-reported risk factor values. There could be an element of self-education or access to professional resources that play a role in the findings of the study.

Finally, it was not possible to exclude individuals from the study who had undiagnosed CHD. The sample likely contained some of these individuals and could have contributed to selection bias concerns.

Future Research

The findings from this study have a unique place in the literature based on the large sample size, breadth of heart disease self-reported risk factors collected, and the method of data collection (ie, the Internet). However, this was a cross-sectional study that lacks the internal validity of a stronger design such as a randomized controlled trial. Future efforts in this field would benefit from a prospective randomized study design to ensure some of the self-selection biases and other limitations of this study are appropriately addressed.

Conclusions

This study sought to understand which CHD risk factors were best understood by community-dwelling adults who took an Internet-based CHD risk assessment (ie, Heart Aware). It also sought to examine whether such levels of understanding were associated with varying degrees of 10-year CHD risk for each participant. What the study has shown is although all CHD risk factors had suboptimal levels of interrater agreement between self-reported and clinically measured values, the CHD risk factor with the greatest discordance was HDL-C. This is consistent with the literature noted previously. However, this study provides unique support to this finding by incorporating a thorough review of how interrater agreement coefficients change based on approaches to missing data. Because missing data are a key analytical issue in many surveillance studies [33], the current study provides a robust view that supports the findings of interrater agreement for HDL-C in a variety of methodological settings. Further, these findings were drawn from a very large sample across more than 100 hospitals.

Unlike prior research efforts, this study stratified interrater agreement of self-reported and clinically measured CHD risk factors by 10-year CHD risk as established by the Framingham Heart Study [24]. This allowed the current study to make a very important contribution to the literature, the discovery that interrater agreement for HDL-C deteriorates as 10-year CHD risk increases, whereas interrater agreement for LDL-C improves as 10-year CHD risk increases. This is a powerful finding because it not only supports the literature noted previously regarding the lack of knowledge of HDL-C among community-dwelling adults, but it also shows how the same individuals also view LDL-C. This finding has substantial implications for the health literacy, social and behavioral health, and public health implementation science communities. If the evidence of HDL-C as a protective factor for CHD continues to mature, it will be vital to translate these clinical findings into actionable public health information campaigns in the community.

Several broad themes should be drawn from this study. First, tools such as Heart Aware could be a cost-effective way to collect valuable CHD risk factor data. Researchers should begin to think about leveraging such technology by partnering with private sector firms to improve public health datasets. Such efforts can only improve public health surveillance, which is positive for researchers, policymakers, private sector firms, and community-dwelling adults. However, additional recruitment methodologies should be employed (in addition to the Internet) to reduce selection bias. Second, this research confirms the continuing need to educate community-dwelling adults about the need to understand their CHD risk factors. This is especially true regarding HDL-C and LDL-C. Finally, this research raises questions about how to use stratification of CHD risk factor agreement by 10-year CHD risk as a clinical strategy. Very few differences in interrater agreement for any CHD risk factor by 10-year CHD risk were identified in this study. Clinicians may want to consider additional strategies to improve CHD risk factor knowledge among those who currently exhibit the greatest chance of a CHD event in the next 10 years.

Acknowledgments

We would like to acknowledge Navigant Consulting, Inc for their collaboration in this study.

Conflicts of Interest

Justin B Dickerson, PhD, MBA performed intermittent paid consulting services for Navigant Consulting, Inc on separate data not related to the data used for this research.

References

1. D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation 2008 Feb 12;117(6):743-753 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.107.699579] [Medline: 18212285]
2. Centers for Disease Control and Prevention. 2012. Leading Causes of Death URL: http://www.cdc.gov/nchs/fastats/lcod.htm [accessed 2013-07-09] [WebCite Cache ID 6HzWfaIvm]
3. Centers for Disease Control and Prevention. 2010. Heart Disease Facts URL: http://www.cdc.gov/heartdisease/facts.htm [accessed 2013-07-09] [WebCite Cache ID 6HzX2pUAx]
4. Heidenreich PA, Trogdon JG, Khavjou OA, Butler J, Dracup K, Ezekowitz MD, American Heart Association Advocacy Coordinating Committee, Stroke Council, Council on Cardiovascular RadiologyIntervention, Council on Clinical Cardiology, Council on EpidemiologyPrevention, Council on Arteriosclerosis, ThrombosisVascular Biology, Council on Cardiopulmonary, Critical Care, PerioperativeResuscitation, Council on Cardiovascular Nursing, Council on the Kidney in Cardiovascular Disease, Council on Cardiovascular SurgeryAnesthesia,Interdisciplinary Council on Quality of CareOutcomes Research. Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association. Circulation 2011 Mar 1;123(8):933-944 [FREE Full text] [doi: 10.1161/CIR.0b013e31820a55f5] [Medline: 21262990]
5. Centers for Disease Control and Prevention. 2009. Heart Disease Risk Factors URL: http://www.cdc.gov/heartdisease/risk_factors.htm [accessed 2013-07-09] [WebCite Cache ID 6HzXBxxTO]
6. Broekhuizen K, van Poppel MN, Koppes LL, Kindt I, Brug J, van Mechelen W. No significant improvement of cardiovascular disease risk indicators by a lifestyle intervention in people with familial hypercholesterolemia compared to usual care: results of a randomised controlled trial. BMC Res Notes 2012;5:181 [FREE Full text] [doi: 10.1186/1756-0500-5-181] [Medline: 22490761]
7. National Heart, Lung, and Blood Institute. 2012. The Heart Truth: A Campaign for Women about Heart Disease URL: http://www.nhlbi.nih.gov/educational/hearttruth/ [accessed 2013-07-09] [WebCite Cache ID 6HzXSsil0]
8. American Heart Association. 2012. Go Red For Women URL: http://www.goredforwomen.org/ [accessed 2013-07-09] [WebCite Cache ID 6HzXfcR1E]
9. Dorner T, Fodor JG, Lawrence K, Ludvik B, Rieder A. HDL-knowledge in the lay public: results of a representative population survey. Atherosclerosis 2007 Nov;195(1):195-198. [doi: 10.1016/j.atherosclerosis.2006.09.020] [Medline: 17069819]
10. Tolonen H, Keil U, Ferrario M, Evans A, MONICA Project WHO. Prevalence, awareness and treatment of hypercholesterolaemia in 32 populations: results from the WHO MONICA Project. Int J Epidemiol 2005 Feb;34(1):181-192 [FREE Full text] [doi: 10.1093/ije/dyh298] [Medline: 15333620]
11. Gorber SC, Tremblay M, Campbell N, Hardt J. The accuracy of self-reported hypertension: A systematic review and meta-analysis. CHYR 2008 Feb 01;4(1):36-62. [doi: 10.2174/157340208783497192]
12. Taylor A, Dal Grande E, Gill T, Pickering S, Grant J, Adams R, et al. Comparing self-reported and measured high blood pressure and high cholesterol status using data from a large representative cohort study. Aust N Z J Public Health 2010 Aug;34(4):394-400. [doi: 10.1111/j.1753-6405.2010.00572.x] [Medline: 20649780]
13. Ayala C, Neff LJ, Croft JB, Keenan NL, Malarcher AM, Hyduk A, et al. Prevalence of self-reported high blood pressure awareness, advice received from health professionals, and actions taken to reduce high blood pressure among US adults--Healthstyles 2002. J Clin Hypertens (Greenwich) 2005 Sep;7(9):513-519. [Medline: 16227770]
14. Huang PY, Buring JE, Ridker PM, Glynn RJ. Awareness, accuracy, and predictive validity of self-reported cholesterol in women. J Gen Intern Med 2007 May;22(5):606-613 [FREE Full text] [doi: 10.1007/s11606-007-0144-1] [Medline: 17370032]
15. Ahluwalia IB, Tessaro I, Rye S, Parker L. Self-reported and clinical measurement of three chronic disease risks among low-income women in West Virginia. J Womens Health (Larchmt) 2009 Nov;18(11):1857-1862. [doi: 10.1089/jwh.2009.1360] [Medline: 19951222]
16. Roger VL, Go AS, Lloyd-Jones DM, Adams RJ, Berry JD, Brown TM, American Heart Association Statistics CommitteeStroke Statistics Subcommittee. Heart disease and stroke statistics--2011 update: a report from the American Heart Association. Circulation 2011 Feb 1;123(4):e18-e209 [FREE Full text] [doi: 10.1161/CIR.0b013e3182009701] [Medline: 21160056]
17. Framingham Heart Study. 2011. General Cardiovascular Disease (10-year risk) URL: http://www.framinghamheartstudy.org/risk/gencardio.html [accessed 2013-07-09] [WebCite Cache ID 6HzXyD3Bh]
18. Fox-Wasylyshyn SM, El-Masri MM. Handling missing data in self-report measures. Res Nurs Health 2005 Dec;28(6):488-495. [doi: 10.1002/nur.20100] [Medline: 16287052]
19. Little RJA, Rubin DB. Statistical Analysis with Missing Data. Hoboken, NJ: Wiley; 2002.
20. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: Wiley; 1987.

21. Weesie J. Stata. 2012. Mvpatterns (user-written software program for Stata) URL: http://www.stata.com/stb/stb61/dm91/mvpatterns.hlp [accessed 2014-04-09] [WebCite Cache ID 6OhwWSxJ2]
22. UCLA Institute for Digital Research and Education. 2012. Statistical Computing Seminars: Multiple Imputation in Stata, Part 2 URL: http://www.ats.ucla.edu/stat/stata/seminars/missing_data/mi_in_stata_pt2.htm [accessed 2013-07-09] [WebCite Cache ID 6HzYDC8wf]
23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
24. Framingham Heart Study. 2011. 10-Year Coronary Heart Disease Risk Score URL: http://www.framinghamheartstudy.org/risk/coronary.html [accessed 2013-07-09] [WebCite Cache ID 6HzYOJ11G]
25. Stark Casagrande S, Ríos Burrows N, Geiss LS, Bainbridge KE, Fradkin JE, Cowie CC. Diabetes knowledge and its relationship with achieving treatment recommendations in a national sample of people with type 2 diabetes. Diabetes Care 2012 Jul;35(7):1556-1565 [FREE Full text] [doi: 10.2337/dc11-1943] [Medline: 22498806]
26. Dolgin E. Trial puts niacin--and cholesterol dogma--in the line of fire. Nat Med 2011 Jul;17(7):756. [doi: 10.1038/nm0711-756] [Medline: 21738133]
27. Wangberg SC, Andreassen HK, Prokosch HU, Santana SM, Sørensen T, Chronaki CE. Relations between Internet use, socio-economic status (SES), social support and subjective health. Health Promot Int 2008 Mar;23(1):70-77 [FREE Full text] [doi: 10.1093/heapro/dam039] [Medline: 18083686]
28. Kargman DE, Sacco RL, Boden-Albala B, Paik MC, Hauser WA, Shea S. Validity of telephone interview data for vascular disease risk factors in a racially mixed urban community: the Northern Manhattan Stroke Study. Neuroepidemiology 1999;18(4):174-184. [doi: 10.1159/000026209] [Medline: 10364718]
29. Cao L, Young N, Liu H, Silvestry S, Sun W, Zhao N, et al. Preoperative aspirin use and outcomes in cardiac surgery patients. Ann Surg 2012 Feb;255(2):399-404. [doi: 10.1097/SLA.0b013e318234313b] [Medline: 21997805]
30. Centers for Disease Control and Prevention. 2007. Leading Causes of Death in Females URL: http://www.cdc.gov/women/lcod/ [accessed 2013-07-09] [WebCite Cache ID 6HzYXcV1o]
31. Guba EG. Criteria for assessing the trustworthiness of naturalistic inquiries. Educational Communication and Technology Journal 1981;29(2):75-91. [doi: 10.1007/BF02766777]
32. Bowen GA. Naturalistic inquiry and the saturation concept: a research note. Qualitative Research 2008 Feb 01;8(1):137-152. [doi: 10.1177/1468794107085301]
33. Buhi ER, Goodson P, Neilands TB. Out of sight, not out of mind: strategies for handling missing data. Am J Health Behav 2008;32(1):83-92. [doi: 10.5555/ajhb.2008.32.1.83] [Medline: 18021036]

Abbreviations

AHA: American Heart Association
BMI: body mass index
CHD: coronary heart disease
CVD: cardiovascular disease
DBP: diastolic blood pressure
DM: diabetes mellitus
HDL-C: high-density lipoprotein cholesterol
LDL-C: low-density lipoprotein cholesterol
MI: multiple imputation
NHLBI: National Heart Lung and Blood Institute
SBP: systolic blood pressure
TC: total cholesterol

Edited by G Eysenbach; submitted 26.09.12; peer-reviewed by F Grajales III, E Waters; comments to author 21.01.13; revised version received 18.04.13; accepted 21.08.13; published 18.04.14.

Please cite as:
Dickerson JB, McNeal CJ, Tsai G, Rivera CM, Smith ML, Ohsfeldt RL, Ory MG
Can an Internet-Based Health Risk Assessment Highlight Problems of Heart Disease Risk Factor Awareness? A Cross-Sectional Analysis
J Med Internet Res 2014;16(4):e106
URL: http://www.jmir.org/2014/4/e106/
doi:10.2196/jmir.2369
PMID:24760950

Original Paper

Usage and Effectiveness of a Fully Automated, Open-Access, Spanish Web-Based Smoking Cessation Program: Randomized Controlled Trial

Guillermo Mañanes[1*], PhD(Psych); Miguel A Vallejo[1*], PhD(Psych)

Faculty of Psychology, Department of Clinical Psychology, National Distance Education University (UNED), Madrid, Spain

[*]all authors contributed equally

Corresponding Author:
Miguel A Vallejo, PhD(Psych)
Faculty of Psychology
Department of Clinical Psychology
National Distance Education University (UNED)
Juan del Rosal 10
Madrid, 28040
Spain
Phone: 34 913987936
Fax: 34 913986234
Email: mvallejo@psi.uned.es

Abstract

Background: The Internet is an optimal setting to provide massive access to tobacco treatments. To evaluate open-access Web-based smoking cessation programs in a real-world setting, adherence and retention data should be taken into account as much as abstinence rate.

Objective: The objective was to analyze the usage and effectiveness of a fully automated, open-access, Web-based smoking cessation program by comparing interactive versus noninteractive versions.

Methods: Participants were randomly assigned either to the interactive or noninteractive version of the program, both with identical content divided into 4 interdependent modules. At baseline, we collected demographic, psychological, and smoking characteristics of the smokers self-enrolled in the Web-based program of Universidad Nacional de Educación a Distancia (National Distance Education University; UNED) in Madrid, Spain. The following questionnaires were administered: the anxiety and depression subscales from the Symptom Checklist-90-Revised, the 4-item Perceived Stress Scale, and the Heaviness of Smoking Index. At 3 months, we analyzed dropout rates, module completion, user satisfaction, follow-up response rate, and self-assessed smoking abstinence.

Results: A total of 23,213 smokers were registered, 50.06% (11,620/23,213) women and 49.94% (11,593/23,213) men, with a mean age of 39.5 years (SD 10.3). Of these, 46.10% (10,701/23,213) were married and 34.43% (7992/23,213) were single, 46.03% (10,686/23,213) had university education, and 78.73% (18,275/23,213) were employed. Participants smoked an average of 19.4 cigarettes per day (SD 10.3). Of the 11,861 smokers randomly assigned to the interactive version, 2720 (22.93%) completed the first module, 1052 (8.87%) the second, 624 (5.26%) the third, and 355 (2.99%) the fourth. Completion data was not available for the noninteractive version (no way to record it automatically). The 3-month follow-up questionnaire was completed by 1085 of 23,213 enrolled smokers (4.67%). Among them, 406 (37.42%) self-reported not smoking. No difference between groups was found. Assuming missing respondents continued to smoke, the abstinence rate was 1.74% (406/23,213), in which 22,678 were missing respondents. Among follow-up respondents, completing the 4 modules of the intervention increased the chances of smoking cessation (OR 1.95, 95% CI 1.27-2.97, $P<.001$), as did smoking 30 minutes (OR 1.58, 95% CI 1.04-2.39, $P=.003$) or 1 hour after waking (OR 1.93, 95% CI 1.27-2.93, $P<.001$) compared to smoking within the first 5 minutes after waking.

Conclusions: The findings suggest that the UNED Web-based smoking cessation program was very accessible, but a high level of attrition was confirmed. This could be related to the ease of enrollment, its free character, and the absence of direct contact with professionals. It is concluded that, in practice, the greater the accessibility to the program, the lower the adherence and retention. Professional support from health services and the payment of a reimbursable fee could prevent high rates of attrition.

(J Med Internet Res 2014;16(4):e111) doi:10.2196/jmir.3091

KEYWORDS

smoking cessation; Internet; intervention studies

Introduction

The World Health Organization [1] states that tobacco consumption causes 5.6 million deaths through lung cancer, heart disease, stroke, and other diseases. This number reaches 6 million if passive smokers' deaths are considered. Other data shows that in the European Union [2] 650,000 people die each year because of tobacco consumption, 443,000 deaths each year are attributable to tobacco use in the United States [3], and 55,000 people die every year from tobacco-related diseases in Spain [4,5].

Smokers are aware of the harmful effects of tobacco consumption: 70% would like to quit smoking and half try to quit each year, mostly without professional help [6-11]. Between 3% and 4% successfully quit smoking [12,13].

A problem of this magnitude requires treatments that are both effective and accessible to prevent millions of deaths worldwide [14-16], especially because treatments for tobacco dependence are available for only 14% of the world population [15]. Many smokers do not want to or cannot receive conventional treatment [17]. The Internet may be an effective, accessible, and efficient alternative in such cases [18-24].

Prior work has shown that interactive Web-based interventions for smoking cessation can be more effective than static websites and that there is a relationship between dose of intervention and its effect [23-31]. In relation to smokers' characteristics that affect smoking cessation, some studies have indicated that a higher level of education and a lower number of cigarettes smoked are related to successful smoking cessation [23,32].

Concerning the usage of open-access eHealth programs, some studies have shown massive enrollment of users followed by a high level of dropout in the initial phase of treatment, without second visits to the website. The proportion of dropouts and nonusers decreases during treatment and subsequent follow-up in logarithmic progression. However, users who are registered in these open and public websites can benefit just as much from treatment as the participants in clinical studies, in which there is greater control over the recruitment of subjects and their behavior in the program [31,33-35]. Little is known about the users' characteristics and the effectiveness and usage of fully automated Web-based interventions for smoking cessation in a real-world setting, without control or selection of users and provided free of charge. In Spain, no fully automated and open-access programs to quit smoking have been evaluated. The authors developed the Universidad Nacional de Educación a Distancia (National Distance Education University, UNED) Web-based program to offer an accessible alternative to millions of smokers who wish to quit smoking without attending conventional treatment sessions.

The aim of this paper is to describe the demographic, consumption, and psychological characteristics of 23,213 participants self-enrolled in the UNED Web-based smoking cessation program, to analyze the usage and effectiveness of the program, and the differences between 2 versions: an interactive automated control on the progress of the user and another without control or interactivity. We also examine if the participants' adherence to treatment is related to the effectiveness of the intervention, and which user characteristics predict abstinence.

The hypotheses tested were (1) the interactive and tailored version of the program will yield higher quit rates than the static version, (2) usage of the modules of the intervention will drop drastically from the first module, (3) exposure to the content of the program will improve quit rates, and (4) the participants with higher education or lower physical dependence are more likely to achieve abstinence. Additionally, independent variables predicting module completion (ie, adherence to treatment) and follow-up response (ie, retention) were analyzed.

Methods

Ethical Approval

The study was a service open to all comers. The Bioethical Committee of the UNED approved the study. Registration of this trial was not required. Before starting the experiment, informed consent was obtained from the participants.

Participants and Recruitment

Participants confirmed via the Web the following requirements before they could start the intervention: not undergoing other treatments to quit smoking, being older than 18 years, wishing to cease tobacco consumption in the next 30 days, smoking at least 2 cigarettes per day, having Internet access and an email address, and accepting the treatment conditions. Participants were informed that they had to complete a follow-up questionnaire 3 months after the beginning of the intervention.

At baseline, we collected demographic, psychological, and smoking characteristics data from the self-enrolled participants on the open-access UNED Web-based smoking cessation program [36] from October 2009 to May 2010, through a mandatory 61-item questionnaire. No direct contact with participants was made at any time. No economic incentives were employed. At the home page of the website, the participants were informed about the research nature of the program and about the researchers' UNED affiliations. The launch of the program in October 2009 was announced by the university press office to mass media.

Two days after filling out the baseline questionnaire, the user received a link by email to register in the program. Once the link was activated, the smoker accessed a randomly assigned version of the program and the baseline questionnaire. The obligation to fill out the 61-item self-administered questionnaire and the 2-day delay in treatment access aimed to prevent the enrollment of impulsive smokers who had no intention of following the treatment.

At 3 months, participants were automatically reminded by email to complete a questionnaire about satisfaction with the program along with smoking and psychological variables. There was no live contact with the users to remind them to fill out the follow-up form.

Measures

Participants self-reported their age, sex, nationality, marital status, education, and employment status. Psychological measurement instruments previously used online [26,37,38] were selected to avoid uncontrolled effects because of their use via Internet [39]. Psychological variables were assessed with the following questionnaires: the Symptom Checklist-90-Revised (SCL 90-R) [40], anxiety and depression subscales, and the 4-item Perceived Stress Scale (PSS-4) [41]. The PSS-4 is a reduced version of the 14-item PSS. It measures the degree to which the respondent has perceived stressful situations during the past month. Higher scores are correlated to more stress.

The participants' self-reported on when they started smoking, physician's advice on smoking cessation, motivation to quit smoking, living with smokers, and expectations of treatment success. Physical dependence on nicotine was measured with the Heaviness of Smoking Index (HSI) [42]. The HSI is a reduced 2-item version of the Fagerstrom Tolerance Questionnaire (FTQ) [43]. It measures nicotine dependence by using 2 questions from the FTQ: time of first cigarette in the morning and the number of cigarettes smoked per day. For each item, scores range from 0 to 3. The total score is the sum of the score on these 2 items. Nicotine dependence is then categorized into a 3-category variable: low (0-1), medium (2-4), and high (5-6). The HSI is used when time and resources are scarce [44].

To analyze utilization of the program, at 3 months we studied dropout rates, module completion, satisfaction with the program, and follow-up response rate. The participants could leave the program formally through the program menu option so that they were registered as a dropout. Module completion data was obtained automatically from the module questionnaire. Only when all the questions from each module were correctly answered was the smoker allowed to advance to the next module and module completion registered by the program. That information was only accessible from the interactive version. Satisfaction with the program was rated by the participants on a 5-level scale from 0=not at all satisfied to 4=extremely satisfied.

The follow-up response rate was calculated from the number of participants who completed the 3-month follow-up assessment. There was no contact by phone or by other means to reach the missing respondents.

The effectiveness of the program was measured by self-reported smoking status by using complete case and intention-to-treat analyses (ITT) at 3 months after registration. However, in cases of high dropout rates or very small follow-up response rate, ITT analysis could underestimate the effect of the program on the participants who were using it and who had been exposed to the intervention [34]. Because we were evaluating the effectiveness of an open-access Web-based intervention, exposure to and usage of the program is an essential condition to analyze its effect on abstinence. Otherwise, we could be assessing the follow-up success of the program instead of its effectiveness [45].

Description of the Program

The content of the UNED Web-based smoking cessation program is divided into 4 consecutive modules. The intervention followed the Clinical Guidelines for the Treatment of Smoking [27] and is based on cognitive behavioral therapy methods tested effectively in conventional face-to-face smoking cessation programs [46,47]: education about the quit process, nicotine fading, self-monitoring, self-control, relapse prevention, coping skills, and lifestyle change. Two versions of the same content were implemented: interactive and noninteractive. For both versions, and according to specific algorithms, the output of the program depended on the users' answers to the requirements of the intervention, acceptance of the treatment conditions, and completion of the pretreatment and posttreatment assessments. Both the scores of the psychological scales and the baseline questionnaire report were also automatically obtained through algorithms.

The program randomly assigned the users to either interactive or noninteractive versions of the program, both with the same therapeutic content. In the interactive format, the user had to follow a particular sequence of treatment in such a way that the modules were presented from first to fourth. Each module incorporated an evaluation form, which, if not answered correctly, prevented progress in the treatment sequence in order to guarantee that users had received the contents gradually, according to its own progress, and that users who had completed the last module had been exposed to all the content. An algorithm also allowed or prevented advancing to the next module depending on the time that had elapsed since the start of the module. In addition, another algorithm reminded the users through email that a week had passed and they had not completed the respective module. Greater exposure to the treatment was the aim. The previously completed modules were available to the user for review.

In the noninteractive version, users received identical content to that in the interactive version through a link to a single static Portable Document Format (PDF) file. Participants decided by themselves either to follow the normal sequence of modules or to skip some content. In this version, the program could control neither the behavior nor the progress of the user or the modules' completion.

Statistical Analysis

Overview

An alpha level of .05 was used for the statistical tests. Pearson correlation coefficient (*r*) was used to relate the quantitative variables and the chi-square test (χ^2) for the categorical variables. The relation between categorical and quantitative variables was calculated with the Student *t* test.

Effect Size

Given the large sample size of the study, some very small differences could be statistically significant; therefore, we calculated the strength of relationship or effect size through Cohen's *d* index for *t* test and Cramer's V for chi-square test, according to Cohen's criteria. For Cohen's *d*, the small, medium, and large effect sizes are .20, .50, and .80, respectively. For Cramer's V, the small, medium, and large effect sizes are calculated from the w index according to the number of categories in the variable. For the Pearson correlation coefficient (*r*), small, medium, and large effect sizes are .10, .30, and .50, respectively [48,49].

Statistical Power

With a small sample size, a lack of statistical significance does not necessarily mean the absence of differences, but rather low statistical power of the test. The statistical power depends on the number of subjects, the significance level, and the effect size. In our study, the large number of participants in the sample allowed us to detect small effect sizes in differences of means, correlations, and chi-square tests.

The sample size required to detect small effect sizes with a statistical power of alpha=.8 and an alpha=.5 is 393 for mean difference, 783 for Pearson's *r*, and between 785 and 1362 subjects (depending on the degrees of freedom) in the case of the chi-square test. The sample size required to detect medium effect sizes with a statistical power of alpha=.8 and an alpha=.5 is 64 for mean difference, 85 for Pearson's *r*, and between 87 and 151 subjects (depending on the degrees of freedom) in the case of the chi-square test [49].

Logistic Regression

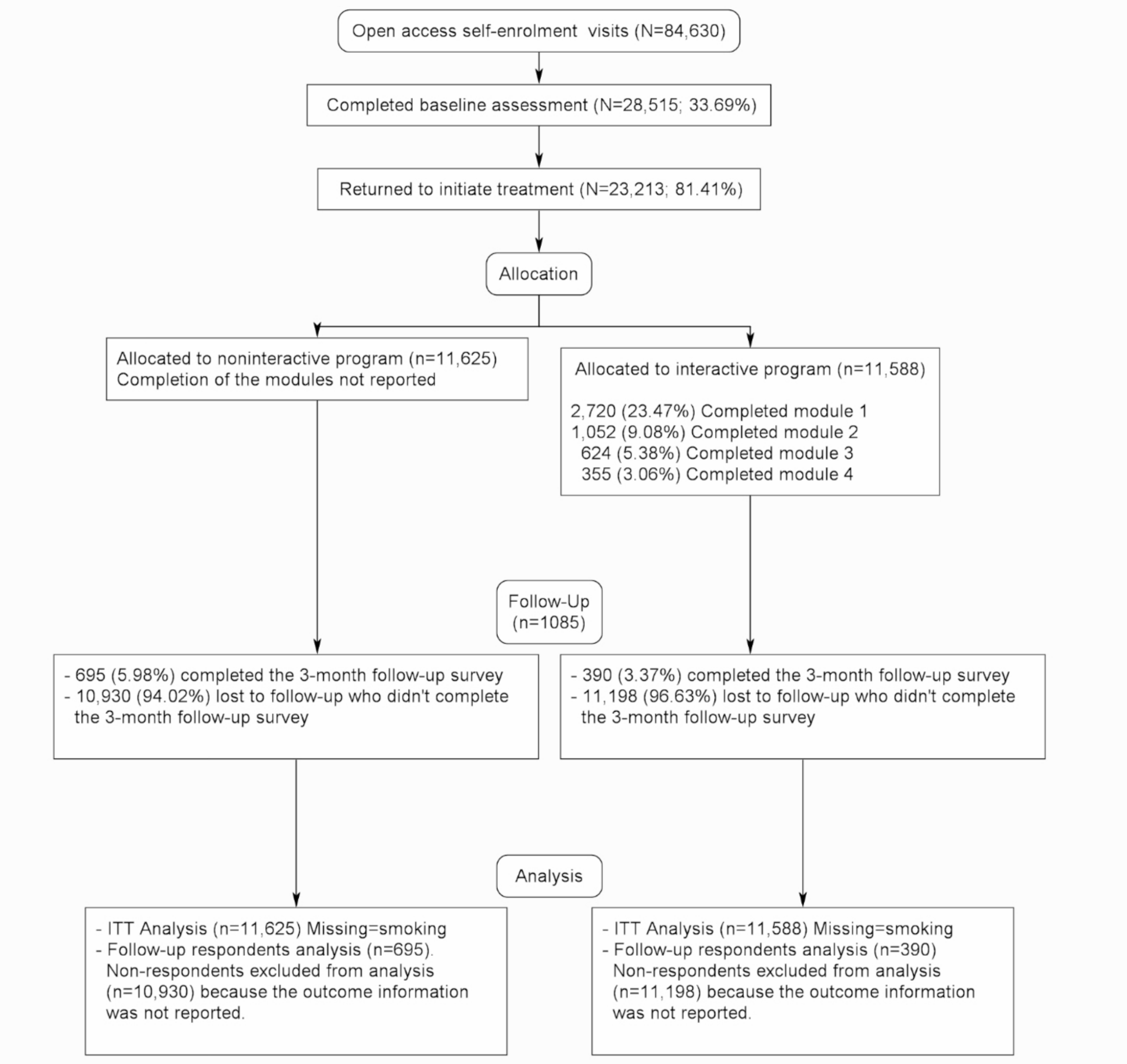
Logistic regression analysis was used to examine the relationship between smoking cessation and predictor variables. Additionally, we examined completion of the program and the 3-month follow-up response as dependent variables. We calculated adjusted odds ratio (OR) for every predictor variable to understand its effect on the dependent variable. Following the recommendations of Hosmer and Lemeshow [50], the initial model incorporated predictor variables with significant value of less than .25 in the preliminary bivariate analysis using chi-square and Student *t* tests. The forward stepwise method was used to test variables for entry into the model one by one. After each entry, variables that were already in the model were tested for possible removal based on the significance of likelihood ratio.

Results

Overview

The Consolidated Standards of Reporting Trials (CONSORT) flow diagram with the number of participants randomly assigned to each group, participants who completed each module of the treatment in the interactive group, and participants analyzed for the outcome is shown in Figure 1.

Figure 1. CONSORT flow diagram of visits to the website, participants who completed assessments, started treatment, and their allocation, as well as completers, follow-up, and analysis.



Participant Demographic Characteristics

Demographic characteristics at baseline are presented in Table 1. Randomization was successful, with neither group differing significantly on number of enrollees (11,902 in the noninteractive and 11,861 in the interactive, *P*=.79) or any baseline variable. The mean age of participants was 39.5 years (SD 10.3) and most were of Spanish nationality (93.57%, 21,721/23,213), married (46.10%, 10,701/23,213), with university education (46.03%, 10,686/23,213), and employed (78.73%, 18,275/23,213).

Table 1. Demographic characteristics of the program users.

Variable	Total	Version		*P*	Cramer's V
		Noninteractive	Interactive		
Users registered, n (%)	23, 213 (100)	11,625 (50.07)	11,588 (49.93)	.79	
Sex, n (%)				.14	
Male	11,593 (49.94)	5862 (50.42)	5731 (49.46)		
Female	11,620 (50.06)	5763 (49.58)	5857 (50.54)		
Age, mean (SD)	39.50 (10.3)	39.56 (10.33)	39.43 (10.30)	.36	
Nationality, n (%)				.47	
Spanish	21,721 (93.57)	10,867 (93.48)	10,854 (93.66)		
Other EU[a]	336 (1.48)	162 (1.39)	174 (1.50)		
Non-EU[a]	1156 (4.98)	596 (5.13)	560 (4.84)		
Marital status, n (%)				.62	
Single	7992 (34.43)	4009 (34.48)	3983 (34.37)		
Married	10,701 (46.10)	5328 (45.83)	5373 (46.36)		
Separated	2099 (9.04)	1066 (9.16)	1033 (8.91)		
Living as a couple	2192 (9.44)	1097 (9.44)	1095 (9.45)		
Widowed	229 (0.99)	125 (1.09)	104 (0.91)		
Education, n (%)				.54	
Primary	3064 (13.20)	1502 (12.92)	1562 (13.48)		
High school	5002 (21.55)	2535 (21.81)	2467 (21.29)		
Professional training	4461 (19.22)	2242 (19.28)	2219 (19.15)		
University	10,686 (46.03)	5346 (45.99)	5340 (46.08)		
Employment status, n (%)				.07	0.012
Working	18,275 (78.73)	9096 (78.24)	9179 (79.21)		
Unemployed	4938 (21.27)	2529 (21.76)	2409 (20.79)		

[a] EU=European Union.

Participant Smoking Variables

Program users' mean daily tobacco consumption was 19.3 cigarettes (SD 10.3). Most participants smoked within the first half hour upon waking up. According to the HSI [44], nicotine dependence was medium in 48.16% (11,179/23,213) of participants and low in 34.85% (8090/23,213). Average age at onset of smoking was 17.3 years (SD 3.6). Of the participants, 64.27% (14,921/23,213) had received medical advice to quit smoking. Desire for abstinence was high (mean 7.9/10, SD 1.9), and expectations of success were positive (56.85%, 13,197/23,213). Both groups differed significantly on desire for abstinence and expectations of success, but with a very small effect size (Table 2).

Additional analysis showed that the number of cigarettes smoked per day increased in males, with age, with lower educational level, and in the case of separated people and widowers, with medium effect sizes for age and education and small effect sizes for sex and marital status (Table 3).

Table 2. Smoking characteristics of the program users.

Variable	Total	Version		*P*	Cohen's *d*	Cramer's V
		Noninteractive	Interactive			
Cigarettes per day, mean (SD)	19.3 (10.3)	19.4 (10.3)	19.3 (10.3)	.50		
First cigarette of the day, n (%)				.22		
≤5 min	5431 (24.48)	2760 (50.81)	2671 (49.19)			
6-30 min	9891 (44.59)	4987 (50.42)	4904 (49.58)			
31-60 min	3687 (16.62)	1803 (48.90)	1884 (51.10)			
>60 min	3171 (14.31)	1564 (52.16)	1607 (47.84)			
HSI, [a] n (%)				.32		
Low	7647 (34.84)	3780 (49.43)	3867 (50.57)			
Medium	10,569 (48.16)	5305 (50.19)	5264 (49.81)			
High	3729 (17.00)	1897 (50.87)	1832 (49.13)			
Age at onset, mean (SD)	17.3 (3.6)	17.4 (3.7)	17.4 (3.6)	.73		
Physician's advice, n (%)				.30		
Yes	14,921 (64.27)	7510 (50.33)	7411 (49.67)			
No	8292 (35.73)	4115 (49.62)	4177 (50.38)			
Desire for abstinence (0-10), mean (SD)	7.9 (1.9)	8.0 (1.9)	7.9 (2.0)	.01	.00	
Lives with smokers, n (%)				.94		
Yes	14,267 (61.46)	7142 (50.05)	7125 (49.95)			
No	8946 (38.54)	4483 (50.11)	4463 (49.89)			
Expectations of success, n (%)				.02		0.021
Not at all	1030 (4.44)	488 (47.38)	542 (52.62)			
Some	8986 (38.71)	4426 (49.25)	4560 (50.75)			
Pretty much	9014 (38.83)	4553 (50.51)	4461 (49.94)			
Completely	4183 (18.02)	2158 (51.59)	2025 (48.41)			

[a] HSI=Heaviness of Smoking Index.

Table 3. Users according to number of cigarettes smoked and demographic variables.

Variable	Number of cigarettes, %						*P*	Cramer's V
	≤10 n=5026	11-20 n=11,688	21-30 n=4292	31-40 n=1799	41-50 n=267	≥51 n=139		
Sex							<.001	0.191
Male	17.32	46.81	22.36	10.74	1.66	1.11		
Female	26.08	53.94	14.51	4.81	0.56	0.10		
Age							<.001	0.150
≤31	31.62	54.52	11.22	2.32	0.21	0.11		
32-39	22.66	52.81	17.72	5.81	0.58	0.42		
40-47	16.30	49.22	22.31	10.22	1.42	0.53		
≥48	15.19	44.31	23.25	13.22	2.32	1.71		
Education							<.001	0.153
Primary	11.97	49.01	24.02	11.78	2.22	1.00		
High school	18.65	49.15	20.82	9.31	1.25	0.82		
Professional training	20.71	52.52	18.10	7.81	0.72	0.41		
University	26.15	50.42	16.11	5.74	0.92	0.66		
Marital status							<.001	0.076
Single	27.42	52.04	14.86	4.72	0.68	0.28		
Married	18.61	49.64	20.49	9.06	1.38	0.82		
Separated	16.24	47.12	22.02	12.11	1.49	1.02		
Living as a couple	21.81	50.75	18.12	7.74	1.01	0.57		
Widowed	14.38	45.38	26.09	10.99	2.72	0.44		

Participant Psychological Variables

Mean scores on anxiety, depression, and stress were not clinically significant. Groups did not differ significantly on any score. Additional analysis showed a significant relation between cigarette consumption and the psychological variables, but with a very small effect size (Table 4). The relationship between the depression subscale and the number of cigarettes smoked in men showed a small effect size (Cohen's *d*=.11); in women, small effect sizes were observed in the depression subscale and the PSS-4 in relation to cigarettes smoked (Cohen's *d*=.12 and .11, respectively).

Table 4. Pearson correlations between psychological variables and number of cigarettes smoked.

Scales	Anxiety[a]		Depression[b]		PSS[c]		Cigarettes	
	r	n	*r*	n	*r*	n	*r*	n
Anxiety[a]	-	23,212						
Depression[b]	.67[d]	23,212	-	23,213				
PSS-4[c]	.48[d]	23,212	.67[d]	23,213	-	23,213		
Cigarettes	.05[d]	22,885	.08[d]	22,886	.08[d]	22,886	-	22,886

[a] Anxiety subscale of the SCL-90-R.

[b] Depression subscale of the SCL-90-R.

[c] PSS-4: 4-item Perceived Stress Scale.

[d] *P*<.001

Dropout Rates and Module Completion

Of the 23,213 self-enrolled smokers, 1326 (5.71%) formally dropped out of the program. There were statistically significant differences between the interactive and noninteractive groups. The interactive group had higher dropout rates than the noninteractive (6.87%, 816/11861 vs 4.28%, 510/11902; χ^2_1=75.9, *P*<.001; Cramer's V=0.057, *P*<.001), with very small effect size.

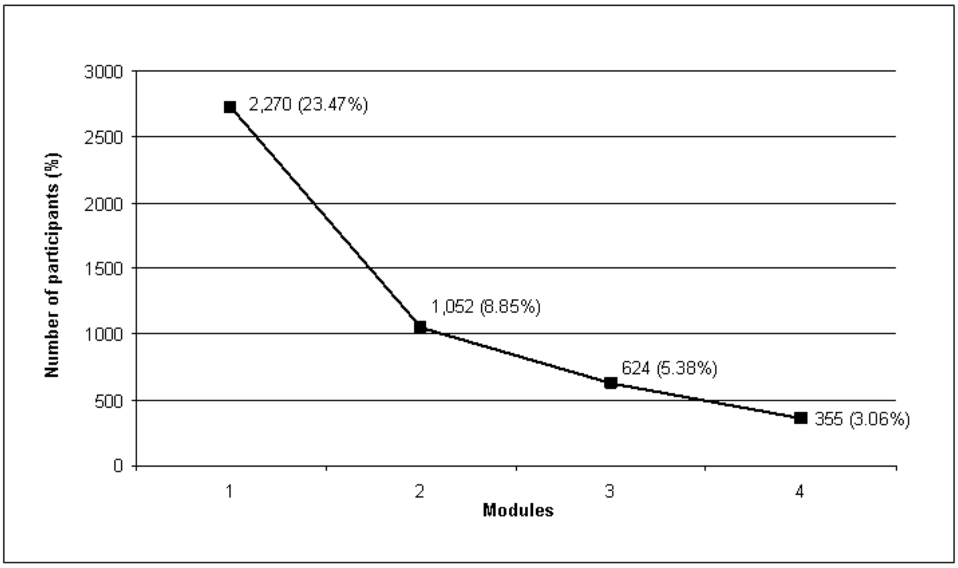
Of the 11,861 participants randomly assigned to the interactive version, 2720 (22.93%) completed the first module, 1052 (8.87%) the second, 624 (5.26%) the third, and 355 (2.99%) the fourth (Figure 2). Completion data was not available for the static PDF version because there was no way to record it automatically.

Logistic regression analyses revealed that being a Spaniard (OR 4.32, 95% CI 1.77-10.51, *P*<.001), aged between 40 and 47 years (OR 1.90, 95% CI 1.38-2.61, *P*<.001) or older than 48 years (OR 1.61, 95% CI 1.14-2.27, *P*=.01), having university education (OR 1.76, 95% CI 1.16-2.67, *P*=.01), and "some," "pretty much," and "complete" expectations of success (OR 3.72, 95% CI 1.17-11.79, *P*=.05; OR 6.03, 95% CI 1.91-18.98, *P*<.001; OR 5.49, 95% CI 1.72-17.57, *P*<.001, respectively) increased the chances of respondents completing all the intervention modules. Being male (OR 0.76, 95% CI 0.61-0.95, *P*=.05), smoking between 16 and 20 (OR 0.74, 95% CI 0.58-0.96, *P*=.05) or more than 20 cigarettes per day (OR 0.65, 95% CI 0.48-0.87, *P*<.001), and a higher level of stress (OR 0.94, 95% CI 0.90-0.98, *P*<.001) decreased chances (Table 5). There was no significant effect (95% CI for the OR includes 1) for the other variables included in the initial model, HSI and anxiety.

Table 5. Logistic regression predicting completion of the intervention in the interactive version (n=10,856).

Variable	B	SE	Wald	*df*	*P*	OR	95% CI
Sex (male)	–0.28	0.12	5.63	1	.02	0.76	0.61-0.95
Age (years)							
≤31	Ref		21.28	3	<.001		
32-39	0.11	0.18	0.37	1	.55	1.11	0.79-1.57
40-47	0.64	0.16	15.60	1	<.001	1.90	1.38-2.61
≥48	0.48	0.18	7.41	1	.01	1.61	1.14-2.27
Nationality (Spanish)	1.46	0.45	10.38	1	<.001	4.32	1.77-10.51
Education							
Primary	Ref		8.20	2	.02		
Secondary	0.37	0.22	2.90	1	.09	1.44	0.95-2.20
University	0.57	0.21	7.12	1	.01	1.76	1.16-2.67
Cigarettes per day							
≤15	Ref		9.83	2	.01		
16-20	–0.30	0.13	5.14	1	.02	0.74	0.58-0.96
≥21	–0.43	0.15	8.22	1	<.001	0.65	0.48-0.87
Expectation of success							
Not at all	Ref		21.89	3	<.001		
Some	1.31	0.59	4.99	1	.03	3.72	1.17-11.79
Pretty much	1.80	0.59	9.42	1	<.001	6.03	1.91-18.98
Completely	1.70	0.59	8.25	1	<.001	5.49	1.72-17.57
Stress	–0.07	0.02	10.10	1	<.001	0.94	0.90-0.98
Constant	–6.27	0.80	62.28	1	<.001	0.00	

Figure 2. Attrition diagram of the modules completed against number of participants in the interactive group (n=11,588).



Follow-Up Response Rate

At 3 months, 1085 (4.67%) of the 23,213 smokers enrolled in the program completed a follow-up questionnaire about smoking status and satisfaction with the program. The 2 groups differed in the proportion of assessments: in the interactive group, 390 (3.29%) of 11,902 participants completed the follow-up questionnaire, whereas 695 of 11,861 (5.84%) did so in the noninteractive (χ^2_1=88.7, *P*=.01; Cramer’s V=0.061, *P*=.01) with very small effect size.

Logistic regression analyses showed that being Spanish (OR 2.70, 95% CI 1.82-4.01, *P*<.001); aged between 32 and 39 years (OR 1.27, 95% CI 1.05-1.54, *P*=.01), between 40 and 47 years (OR 1.52, 95% CI 1.26-1.83, *P*<.001), or older than 47 years (OR 1.57, 95% CI 1.29-1.91, *P*<.001); having a higher level of education (secondary level: OR 1.36, 95% CI 1.07-1.74, *P*=.01; university: OR 1.96, 95% CI 1.54-2.48, *P*<.001); smoking between 16 and 20 (OR 1.43, 95% CI 1.20-1.69, *P*<.001) or more than 20 cigarettes per day (OR 1.20, 95% CI 1.01-1.43, *P*=.04); and having “pretty much” (OR 1.91, 95% CI 1.27-2.88, *P*<.001) or “complete” expectations (OR 1.99, 95% CI 1.31-3.04, *P*<.001) of success in the and noninteractive group (OR 1.78, 95% CI 1.56-2.03, *P*<.001) increased the chances of completing the follow-up questionnaire. Being male decreased chances (OR 0.85, 95% CI 0.75-0.97, *P*=.02) (Table 6). There was no significant effect (95% CI for the OR includes 1) of the other variables included in the initial model: marital status, first cigarette of the day, and level of dependence on nicotine.

Table 6. Logistic regression predicting 3-month follow-up response (n=21,385).

Variable	B	SE	Wald	*df*	*P*	OR	95% CI
Sex (male)	–0.16	0.07	5.66	1	.02	0.85	0.75-0.97
Age (years)							
≤31	Ref		25.97	3	<.001		
32-39	0.24	0.10	6.21	1	.01	1.27	1.05-1.54
40-47	0.42	0.10	19.09	1	<.001	1.52	1.26-1.83
≥48	0.45	0.10	20.65	1	<.001	1.57	1.29-1.91
Nationality (Spanish)	0.99	0.20	24.07	1	<.001	2.70	1.82-4.01
Education							
Primary	Ref		46.59	2	<.001		
Secondary	0.31	0.13	6.10	1	.01	1.36	1.07-1.74
University	0.67	0.12	30.51	1	<.001	1.96	1.54-2.48
Cigarettes per day							
≤15	Ref		16.91	2	<.001		
16-20	0.35	0.09	16.55	1	<.001	1.43	1.20-1.69
≥21	0.19	0.09	4.37	1	.04	1.20	1.01-1.43
Expectation of success							
Not at all	Ref		25.44	3	<.001		
Some	0.37	0.21	3.01	1	.08	1.44	0.95-2.18
Pretty much	0.65	0.21	9.57	1	<.001	1.91	1.27-2.88
Complete	0.69	0.22	10.23	1	<.001	1.99	1.31-3.04
Version of the program (noninteractive)	0.58	0.07	73.91	1	<.001	1.78	1.56-2.03
Constant	–5.69	0.32	308.90	1	<.001	0.00	

Satisfaction With the Program

User satisfaction was reported at 3-month follow-up (n=1085). The participants rated their satisfaction from not at all satisfied to extremely satisfied. Among them, 11.15% (121/1085) were not at all satisfied, 19.26% (209/1085) were slightly satisfied, 34.10% (370/1085) were somewhat satisfied, 25.81% (280/1085) were very satisfied, and 9.67% (105/1085) were extremely satisfied. There were differences in satisfaction in terms of the version of the program (χ^2_4=25.4, *P*<.001; Cramer's V=0.153, *P*<.001) such that users of the interactive version showed higher proportions of very satisfied and extremely satisfied. Older people were more satisfied (χ^2_{12}=39.8, *P*<.001; Cramer's V=0.111, *P*<.001), as were completers (χ^2_4=47.9, *P*<.001; Cramer's V=0.350, *P*<.001), abstainers (χ^2_4=97.2, *P*<.001; Cramer's V=0.299, *P*<.001), and participants with positive expectations (χ^2_{12}=73.1, *P*<.001; Cramer's V=0.260, *P*<.001).

Tobacco Cessation

At 3 months, 1085 users reported their smoking status. Because 97.69% (22,678/23,213) of the participants did not report outcome information, we conducted both ITT analysis and follow-up respondent analysis.

Intention-to-Treat Analysis

Assuming the missing respondents continued to smoke, the abstinence rate was 1.74% (406/23,213), in which 22,678 were missing respondents. There was a significant difference in abstinence rates between groups (χ^2_1=26.7, *P*<.001; Cramer's V=0.034, *P*<.001) with a very small effect size.

Logistic regression analyses revealed that being married or living as a couple (OR 1.26, 95% CI 1.02-1.55, *P*=.03) and complete expectations of success (OR 2.23, 95% CI 1.12-4.46, *P*=.02) increased the chances of quitting smoking. Smoking between 16 and 20 cigarettes per day (OR 0.70, 95% CI 0.56-0.89, *P*<.001) or more than 20 cigarettes per day (OR 0.55, 95% CI 0.42-0.72, *P*<.001), and using the interactive version of the program (OR 0.60, 95% CI 0.49-0.74, *P*<.001) decreased chances of quitting (Table 7). There was no significant effect of the other variables included in the initial model, time to first cigarette of the day, HSI, physician's advice, and depression.

Table 7. Logistic regression predicting smoking cessation at 3 months (n=21,707).

Variable	B	SE	Wald	*df*	*P*	OR	95% CI
Marital status (married/couple)	0.23	0.11	4.70	1	.03	1.26	1.02-1.55
Cigarettes per day							
≤15	Ref		21.15	2	<.001		
16-20	–0.35	0.12	8.71	1	<.001	0.70	0.56-0.89
≥21	–0.59	0.14	18.70	1	<.001	0.55	0.42-0.72
Expectation of success							
Not at all	Ref		10.97	3	.01		
Some	0.43	0.35	1.55	1	.21	1.54	0.78-3.04
Pretty much	0.68	0.35	3.85	1	.05	1.97	1.00-3.86
Complete	0.80	0.35	5.19	1	.02	2.23	1.12-4.46
Version of the program (interactive)	–0.51	0.11	22.55	1	<.001	0.60	0.49-0.74
Constant	–4.27	0.35	150.76	1	<.001	0.01	

Follow-Up Respondent Analysis

Considering only the participants who responded to the follow-up questionnaire (n=1085), the abstinence rate was 37.42% (406/1085), whereas 62.58% (679/1085) self-reported as continuing to smoke. There was no significant difference in self-reported abstinence rates between groups (χ^2_1=0.4, *P*=.50; Cramer's V=0.02, *P*=.50).

Logistic regression analyses revealed that smokers who had their first cigarette of the day after 30 minutes (OR 1.58, 95% CI 1.04-2.39, *P*=.003) or 1 hour (OR 1.93, 95% CI 1.27-2.93, *P*<.001) of waking increased their chances of quitting compared to those who had their first cigarette earlier. There was no significant effect of the other variables included in the regression model, version of the program, age, sex, first cigarette of the day, number of cigarettes smoked, level of dependence on nicotine, expectations, physician advice, depression, and marital status.

The abstinence rate among the participants who completed the intervention was 46.47% (165/355) versus 35.49% (3727/10,500) among noncompleters (*P*=.01). To study the effect of module completion on abstinence (exposure to treatment), we conducted a secondary analysis with the participants who answered the follow-up questionnaire in the interactive version (n=375). Only module completion had a significant effect on smoking cessation (OR 1.95, 95% CI 1.27-2.97, *P*<.001). There was no significant effect of other variables included in the regression model, age, sex, first cigarette of the day, number of cigarettes smoked, level of dependence on nicotine, expectations, physician advice, depression, and marital status.

Continuing to Smoke Analysis

Participants who self-reported they continued to smoke (676/23,213, 2.91%) significantly reduced both the average number of cigarettes smoked daily (from 17.9 to14) and the level of physical dependence on nicotine with medium effect sizes of .45 and .53, respectively (Table 8).

Table 8. Number of cigarettes smoked (n=628) and level of nicotine dependence (n=614) at baseline and after 3 months among participants who self-reported as continuing to smoke.

Variables	At baseline, mean (SD)	At 3 months, mean (SD)	Diff	*r*	*P*	*t (df)*	*P*	Cohen's *d*
Cigarettes	17.90 (9.83)	14.00 (8.17)	3.90	.55	<.001	11.31 (627)	<.001	.45
HSI[a]	2.84 (1.54)	2.20 (1.51)	0.64	.68	<.001	13.04 (613)	<.001	.53

[a] HSI=Heaviness of Smoking Index.

Discussion

Principal Findings

The aim of this paper was to describe the characteristics of participants self-enrolled in the UNED Web-based smoking cessation program and to analyze the program's usage and effectiveness and the differences between its 2 versions: interactive and noninteractive. We also examined if participants' exposure to treatment was related to its effectiveness and which user characteristics predicted abstinence. Additionally, independent variables predicting module completion (ie, adherence to treatment and follow-up response, such as retention) were analyzed.

From October 2009 to May 2010, 23,213 users self-enrolled in the UNED Web-based smoking cessation program, which was free of charge and fully automated. There were no differences in the demographic, psychological, and smoking characteristics between the 2 versions of the program.

Participant Characteristics

The demographic and smoking characteristics of the study participants are similar to those reported in other Internet smoking cessation programs [23,32,51-55] except for sex: in the UNED program, the number of males and females was not significantly different, whereas in most other studies, females were the majority. This may be because the sample sizes were smaller in the mentioned studies, ranging from 351 [55] to 17,159 participants [51]. With larger samples, as in the case of Barrera and colleagues [51], the proportions of men and women were almost the same (49.3% men vs 50.7% women). In contrast, in the UNED, men's participation among non-European users was almost twice that of women's. Except for Spaniards, we did not register the nationality of each user, but rather European or non-European nationality. Because the program was provided only in the Spanish language and announced through national and regional mass media, we can assume that non-European users (n=1156) were primarily people from Latin American countries living in Spain. Despite the fact that there has been an increase in female tobacco consumption in recent years in Latin American countries, the proportion of male smokers is still triple or quadruple that of women in countries such as Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Paraguay, and Peru [56].

The proportion of users with university education (46%) versus other categories is noteworthy, confirming once again the existence of a digital divide in the specific field of eHealth. In Spain, the percentage of people with higher education in 2009 was 16.29% versus 20.22% with primary education. However, that same year, the percentage of Internet users with higher education was 28.34% versus 7.85% with primary education according to the Instituto Nacional de Estadística (National Institute of Statistics) [57,58]. Therefore, we can confirm that smoking cessation programs are similar to other eHealth programs in that individuals with a lower educational level are underrepresented [59,60]. Moreover, a higher ability to understand, organize, and analyze written information is required for the use of Web-based programs compared to a traditional face-to-face intervention [20,61]. All this has a dissuasive effect on participants with a lower educational level. It has also been suggested that the excessive length of the written material employed in self-help formats may impede their use, especially by smokers who are not accustomed to dealing with this type of information [62].

Concerning scores of psychological scales, the results do not differ from those obtained in other studies in the general population [37,63]. In spite of the fact that cigarette consumption has been associated with higher levels of anxiety, depression, and stress [64,65], the correlation was nonsignificant in the present study.

Usage of the Program

The registration of smokers in the program was high, with 23,213 participants self-enrolled in 8 months. This may be because of the characteristics of Web-based programs, which do not require transportation or schedules and, in the case of the UNED, were free of charge and open-access. In addition, the university press office announced the launch of the program to the mass media.

The number of users who completed the 4 modules in the program confirmed the law of attrition proposed by Eysenbach [34], according to which a high rate of noncompleters and those lost to follow-up are a typical feature of Web-based programs. Adherence to the intervention decreased dramatically over time: 22.93% of users completed the first module, 8.87% the second, 5.26% the third, and 2.99% the fourth. Despite the high motivation of the users and their positive expectations at baseline, we found only 2.99% of participants completed the intervention and 4.57% completed the 3-month follow-up even with the 2 measures implemented to prevent impulsive enrollment (ie, delayed access to the program and a long mandatory baseline questionnaire). These results are similar to those shown in other eHealth programs [33,35,66]. In the case of smoking cessation programs, comparison of adherence results between studies is difficult because either the usage measures employed do not include the proportion of users who complete the interventions or they are not structured into successive modules or sessions. For example, Etter [26] reported that 16% of users completed at least 2 of 3 interactive surveys of an intervention that consisted of providing therapeutic content according to the answers to each survey. McKay et al [32] reported an average of 1 visit to the website in a cognitive behavioral program. Seidman et al [23] reported an average of 8 completed sessions of a 32-session cognitive behavioral program, and Rabius et al [53] confirmed no more than 2 visits in a study with random allocation to 5 different websites to quit smoking. Although the unit of adherence measure differs between studies, there is a general underutilization of Web-based interventions, especially in open-access programs. Our hypothesis predicting a high rate of nonusers from the first module was confirmed.

Because a high number of participants enrolled over the course of 8 months (23,213), the UNED Web-based smoking cessation program was shown to be very accessible. No extra measures were used, except email reminders, to persuade the participants to continue using the program and respond to the follow-up questionnaire. The study confirmed that the easier it is to enroll or to leave the program, the higher the nonusage rate and the rate of those lost to follow-up will be, as Eysenbach [34] proposed.

In addition, the regression model showed that the strongest predictors of completing the program were positive expectations of success and being older than 40 years. This could be explained by older participants having an increased perception of the harmful effects of smoking on health compared to younger people, and by positive expectations being related to a high level of self-efficacy, both causing greater use of the application. Similar results were obtained by Zbikowski et al [31] and Wangberg et al [67].

Effectiveness of the Program

The ITT abstinence rate was 1.71%. This result is difficult to interpret. ITT analysis is feasible when we know the outcome of every participant, regardless of whether they have completed the program or not. If the lack of data on the outcome is very high, ITT analysis is not feasible because considering every participant lost to follow-up as one without successful treatment will produce results that rely on unverifiable assumptions [68,69]. That the missing respondents continue to smoke is an untestable assumption about the missing data for 95.44% of the participants in the study, and this could bias results about the real effectiveness of the program [45]. Furthermore, Tomson et al [69] found that nonresponders to a Quitline trial were even more likely to be abstinent than responders.

If we consider the data from the participants who actually reported their smoking status at follow-up (ie, those we know the outcome), the cessation rate was 37.42%. This result is in the range of other studies that analyzed cessation rates in participants who answered a 3-month follow-up by phone or through the Web. Cobb et al [70] found a 44.6% abstinence rate and Etter [52] reported 37.3% at 11 weeks. In the study by McKay et al [32], the proportion was considerably lower, at 19.7%.

Despite the results of previous work [23,24,27,29,30,52] and in contrast to our first hypothesis, in the follow-up respondent analysis we found no differences between the 2 versions of the program—interactive and noninteractive—whereas in the ITT analysis the interactive version decreased chances of quitting. In the present study, both groups were exposed to the same content. The only difference was interactivity; consequently, the comparison was strictly interactivity versus no interactivity rather than interactivity versus another different intervention. To our knowledge, this is the first study that evaluated interactivity versus noninteractivity while maintaining the same content in both groups. However, we should confirm whether the absence of differences found in the study is maintained in the long term.

We consider the interactivity because this is the main experimental manipulation. Nevertheless, the interactive version includes several components: the regular programmed progression across the modules, the evaluation at the end of each module, and the necessity of interaction with the Web application to follow the programmed steps. Knowing the effect of the interactivity will be necessary to explore these variables.

The regression analysis showed that completion of the treatment (ie, exposure to the intervention) increased the chances of smoking cessation, which confirms our third hypothesis. However, contrary to our fourth hypothesis, we could not confirm the effect of education and physical dependence on smoking abstinence. There was no significant effect of any demographic variables on abstinence. Education affected the adherence to the program, but did not improve smoking cessation. Anxiety, depression, and stress did not affect abstinence either. This result is consistent with prior studies that have failed to demonstrate the effect of incorporating mood management components in the intervention [24,71].

Limitations

First, the information obtained at baseline was not confirmed by other means. Second, self-reported abstinence was not validated chemically. In a program with more than 23,000 participants from any part of the world, data verification is not viable. Some studies confirmed the validity of self-reported abstinence when compared with chemical verification, becoming the standard method in Web-based studies [27,72,73]. Another limitation of the study is related to the 3-month follow-up. This period is not long enough to establish clear conclusions about the effectiveness of the program.

Conclusions

The large number of enrollees in the program is encouraging in terms of accessibility and confirms the feasibility of delivering Web-based programs for smoking cessation to smokers that do not have access to conventional face-to-face programs. It would be very difficult to reach 23,213 smokers through a face-to-face intervention setting. This study showed that completion of the program (ie, adherence) improves the chances of cessation. However, high rates of nonusage and loss to follow-up were confirmed. To increase the program’s usage and follow-up completion, some professional involvement could be implemented in such a way as to ensure a self-obligation toward both the program and the professional [74]. For example, professional support from primary health care could prevent early dropout and reinforce adherence. It has been found that users of a Web-based smoking cessation program referred by general practitioners showed lower dropout rates compared to those who accessed the program directly [75]. Another therapeutic measure for increasing adherence to free-of-charge programs could be the payment of a fee to achieve higher commitment to the program, which would be reimbursable if completion is achieved.

To summarize, the main findings of this research are to disconfirm the general belief that anxiety, depression, and stress are associated with cigarette consumption [64,65] and confirmation that it is not necessary to incorporate mood management components in an intervention [24,71]. This outcome allows for configuration of an intervention less associated with clinical terms and more related to health behaviors. Secondly, the utility of a simple noninteractive intervention program versus a more complex interactive one. Finally, this study has shown the potential efficacy of an intervention for people with diverse levels of physical dependence and different education levels.

The results of our study suggest that future research is needed to determine (1) the factors that will reduce the high rates of attrition in public Web-based programs for smoking cessation, (2) the real proportion of nonsmokers among those lost to follow-up, and (3) the long-term abstinence rate in the UNED Web-based smoking cessation program.

Acknowledgments

This research was funded by a grant from the Spanish Ministry of Science and Technology (grant ID# I+D+I SEJ2004-03392/PSI). We thank Ramón Rodríguez and Patricia Valle (in memoriam) for work on the Web application used in this research. We would also like to thank the Department of Communication of the UNED for aid in the Web dissemination of this research.

Conflicts of Interest

None declared.

References

1. WHO report on the global tobacco epidemic, 2011: Warning about the dangers of tobacco. Geneva: World Health Organization; 2011. URL: http://whqlibdoc.who.int/publications/2011/9789240687813_eng.pdf [accessed 2013-10-29] [WebCite Cache ID 6KjEbyYd2]
2. Commission of the European Communities. Green paper Towards a Europe free from tobacco smoke: Policy options at EU level. Brussels: Commission of the European Communities; 2007. URL: http://www.europarl.europa.eu/registre/docs_autres_institutions/commission_europeenne/com/2007/0027/COM_COM(2007)0027_EN.pdf [accessed 2013-10-29] [WebCite Cache ID 6KjGn709C]
3. Centers for Disease ControlPrevention (CDC). Smoking-attributable mortality, years of potential life lost, and productivity losses--United States, 2000-2004. MMWR Morb Mortal Wkly Rep 2008 Nov 14;57(45):1226-1228 [FREE Full text] [Medline: 19008791]
4. Banegas JR, Díez L, González J, Villar F, Rodriguez-Artalejo F. La mortalidad atribuible al tabaquismo comienza a descender en España. Med Clin (Barc) 2005;124(20):769-771. [doi: 10.1157/13075847]
5. Montes A, Pérez M, Gestal JJ. Impacto del tabaquismo sobre la mortalidad en España. Adicciones 2004;16(2):75-82.
6. Centers for Disease ControlPrevention (CDC). Quitting smoking among adults--United States, 2001-2010. MMWR Morb Mortal Wkly Rep 2011 Nov 11;60(44):1513-1519 [FREE Full text] [Medline: 22071589]
7. Fiore MC, Novotny TE, Pierce JP, Giovino GA, Hatziandreu EJ, Newcomb PA, et al. Methods used to quit smoking in the United States. Do cessation programs help? JAMA 1990;263(20):2760-2765. [Medline: 2271019]
8. Friend K, Levy DT. Reductions in smoking prevalence and cigarette consumption associated with mass-media campaigns. Health Educ Res 2002 Feb;17(1):85-98 [FREE Full text] [Medline: 11888047]
9. Shiffman S, Brockwell SE, Pillitteri JL, Gitchell JG. Use of smoking-cessation treatments in the United States. Am J Prev Med 2008 Feb;34(2):102-111. [doi: 10.1016/j.amepre.2007.09.033] [Medline: 18201639]
10. Taylor T, Lader D, Bryant A, Keyse L, Joloza MT. Office for National Statistics. 2006. Smoking-related behaviour and attitudes 2005 URL: http://www.ons.gov.uk/ons/rel/lifestyles/smoking-related-behaviour-and-attitudes/2005/index.html [accessed 2013-10-30] [WebCite Cache ID 6KlKwkZah]
11. Zhu S, Melcer T, Sun J, Rosbrook B, Pierce JP. Smoking cessation with and without assistance: a population-based analysis. Am J Prev Med 2000 May;18(4):305-311. [Medline: 10788733]
12. Hughes JR, Keely J, Naud S. Shape of the relapse curve and long-term abstinence among untreated smokers. Addiction 2004 Jan;99(1):29-38. [Medline: 14678060]
13. West R. Smoking in England. 2006. Background smoking cessation rates in England URL: http://www.smokinginengland.info/downloadfile/?type=sts-documents&src=6 [WebCite Cache ID 6KjLVPO7C]
14. Fiore MC, Croyle RT, Curry SJ, Cutler CM, Davis RM, Gordon C, et al. Preventing 3 million premature deaths and helping 5 million smokers quit: a national action plan for tobacco cessation. Am J Public Health 2004 Feb;94(2):205-210. [Medline: 14759928]
15. World Health Organization. 2011. Tobacco Fact sheet #339 URL: http://www.who.int/mediacentre/factsheets/fs339/en/ [accessed 2013-10-29] [WebCite Cache ID 6KjVBoFrK]
16. Task Force on Community Preventive Services. Recommendations regarding interventions to reduce tobacco use and exposure to environmental tobacco smoke. Am J Prev Med 2001 Feb;20(2 Suppl):10-15. [Medline: 11173214]
17. Curry SJ. ASPO Joseph W. Cullen Memorial Award Lecture. Bridging the clinical and public health perspectives in tobacco treatment research: scenes from a tobacco treatment research career. Cancer Epidemiol Biomarkers Prev 2001 Apr;10(4):281-285 [FREE Full text] [Medline: 11319167]
18. 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]
19. Civljak M, Sheikh A, Stead LF, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2010(9):CD007078. [doi: 10.1002/14651858.CD007078.pub3] [Medline: 20824856]
20. Etter JF. Using new information technology to treat tobacco dependence. Respiration 2002;69(2):111-114 [FREE Full text] [doi: 10.1159/000056311] [Medline: 11961422]
21. Lancaster T, Stead LF. Self-help interventions for smoking cessation. Cochrane Database Syst Rev 2005(3):CD001118. [doi: 10.1002/14651858.CD001118.pub2] [Medline: 16034855]

22. Myung SK, McDonnell DD, Kazinets G, Seo HG, Moskowitz JM. Effects of Web- and computer-based smoking cessation programs: meta-analysis of randomized controlled trials. Arch Intern Med 2009 May 25;169(10):929-937. [doi: 10.1001/archinternmed.2009.109] [Medline: 19468084]
23. Seidman DF, Westmaas JL, Goldband S, Rabius V, Katkin ES, Pike KJ, et al. Randomized controlled trial of an interactive internet smoking cessation program with long-term follow-up. Ann Behav Med 2010 Feb;39(1):48-60. [doi: 10.1007/s12160-010-9167-7] [Medline: 20177844]
24. Shahab L, McEwen A. Online support for smoking cessation: a systematic review of the literature. Addiction 2009 Nov;104(11):1792-1804. [doi: 10.1111/j.1360-0443.2009.02710.x] [Medline: 19832783]
25. Cobb NK, Graham AL, Bock BC, Papandonatos G, Abrams DB. Initial evaluation of a real-world Internet smoking cessation system. Nicotine Tob Res 2005 Apr;7(2):207-216 [FREE Full text] [doi: 10.1080/14622200500055319] [Medline: 16036277]
26. Etter JF. Comparing the efficacy of two Internet-based, computer-tailored smoking cessation programs: a randomized trial. J Med Internet Res 2005;7(1):e2 [FREE Full text] [doi: 10.2196/jmir.7.1.e2] [Medline: 15829474]
27. Fiore MC, Jaen CR, Baker TB, Bailey WC, Benowitz NL, Curry SJ. Treating tobacco use and dependence: 2008 Update. Washington, DC: US Department of Health and Human Services, Public Health Service; 2008. URL: http://www.ahrq.gov/professionals/clinicians-providers/guidelines-recommendations/tobacco/clinicians/treating_tobacco_use08.pdf [accessed 2013-10-29] [WebCite Cache ID 6KjYBoFeA]
28. Japuntich SJ, Zehner ME, Smith SS, Jorenby DE, Valdez JA, Fiore MC, et al. Smoking cessation via the internet: a randomized clinical trial of an internet intervention as adjuvant treatment in a smoking cessation intervention. Nicotine Tob Res 2006 Dec;8 Suppl 1:S59-S67. [Medline: 17491172]
29. Lenert L, Muñoz RF, Perez JE, Bansod A. Automated e-mail messaging as a tool for improving quit rates in an internet smoking cessation intervention. J Am Med Inform Assoc 2004 Aug;11(4):235-240 [FREE Full text] [doi: 10.1197/jamia.M1464] [Medline: 15064291]
30. Strecher VJ, Shiffman S, West R. Randomized controlled trial of a web-based computer-tailored smoking cessation program as a supplement to nicotine patch therapy. Addiction 2005 May;100(5):682-688. [doi: 10.1111/j.1360-0443.2005.01093.x] [Medline: 15847626]
31. Zbikowski SM, Jack LM, McClure JB, Deprey M, Javitz HS, McAfee TA, et al. Utilization of services in a randomized trial testing phone- and web-based interventions for smoking cessation. Nicotine Tob Res 2011 May;13(5):319-327 [FREE Full text] [doi: 10.1093/ntr/ntq257] [Medline: 21330267]
32. McKay HG, Danaher BG, Seeley JR, Lichtenstein E, Gau JM. Comparing two web-based smoking cessation programs: randomized controlled trial. J Med Internet Res 2008;10(5):e40 [FREE Full text] [doi: 10.2196/jmir.993] [Medline: 19017582]
33. Christensen H, Griffiths KM, Korten AE, Brittliffe K, Groves C. A comparison of changes in anxiety and depression symptoms of spontaneous users and trial participants of a cognitive behavior therapy website. J Med Internet Res 2004 Dec 22;6(4):e46 [FREE Full text] [doi: 10.2196/jmir.6.4.e46] [Medline: 15631970]
34. 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]
35. Farvolden P, Denisoff E, Selby P, Bagby RM, Rudy L. Usage and longitudinal effectiveness of a Web-based self-help cognitive behavioral therapy program for panic disorder. J Med Internet Res 2005;7(1):e7 [FREE Full text] [doi: 10.2196/jmir.7.1.e7] [Medline: 15829479]
36. Dejar de Fumar. Tratamiento para dejar de fumar URL: http://www.apsiol.uned.es/dejardefumar/ [accessed 2014-04-14] [WebCite Cache ID 6OpPigFor]
37. Herrero J, Meneses J. Short web-based versions of the Perceived Stress (PSS) and Center for Epidemiological Studies - Depression (CESD) scales: A comparison to pencil and paper responses among Internet users. Comput Human Behav 2006;22. [doi: 10.1016/j.chb.2004.03.007]
38. Vallejo MA, Mañanes G, Isabel Comeche MA, Díaz MI. Comparison between administration via Internet and paper-and-pencil administration of two clinical instruments: SCL-90-R and GHQ-28. J Behav Ther Exp Psychiatry 2008 Sep;39(3):201-208. [doi: 10.1016/j.jbtep.2007.04.001] [Medline: 17573039]
39. Buchanan T. Internet-based questionnaire assessment: appropriate use in clinical contexts. Cogn Behav Ther 2003;32(3):100-109. [doi: 10.1080/16506070310000957] [Medline: 16291542]
40. Derogatis LR. SCL-90-R: Administration, scoring and procedures Manual I for the revised version of the SCL-90. Baltimore, MD: Johns Hopkins University Press; 1977.
41. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
42. Heatherton TF, Kozlowski LT, Frecker RC, Rickert W, Robinson J. Measuring the heaviness of smoking: using self-reported time to the first cigarette of the day and number of cigarettes smoked per day. Br J Addict 1989 Jul;84(7):791-799. [Medline: 2758152]
43. Fagerström KO. Measuring degree of physical dependence to tobacco smoking with reference to individualization of treatment. Addict Behav 1978;3(3-4):235-241. [Medline: 735910]

44. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. Br J Addict 1991 Sep;86(9):1119-1127. [Medline: 1932883]
45. Leykin Y, Aguilera A, Torres LD, Pérez-Stable EJ, Muñoz RF. Interpreting the outcomes of automated internet-based randomized trials: example of an International Smoking Cessation Study. J Med Internet Res 2012;14(1):e5 [FREE Full text] [doi: 10.2196/jmir.1829] [Medline: 22314016]
46. Becoña E, Vázquez FL. Tratamiento del tabaquismo. Madrid: Dykinson; 1998:978-984.
47. Abrams DB, Niaura R, Brown RA, Emmons KM, Goldestein MG, Monti PM. The Tobacco Dependence Treatment Handbook: A Guide to Best Practices. New York: Guilford Press; 2003.
48. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [Medline: 19565683]
49. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Hillsdale, NJ: Erlbaum; 1988.
50. Hosmer DW, Lemeshow S. Applied Logistic Regression, 2nd ed. New York: John Wiley and Sons; 2000.
51. Barrera AZ, Pérez-Stable EJ, Delucchi KL, Muñoz RF. Global reach of an Internet smoking cessation intervention among Spanish- and English-speaking smokers from 157 countries. Int J Environ Res Public Health 2009 Mar;6(3):927-940 [FREE Full text] [doi: 10.3390/ijerph6030927] [Medline: 19440423]
52. Etter JF. Comparing the efficacy of two Internet-based, computer-tailored smoking cessation programs: a randomized trial. J Med Internet Res 2005;7(1):e2 [FREE Full text] [doi: 10.2196/jmir.7.1.e2] [Medline: 15829474]
53. Rabius V, Pike KJ, Wiatrek D, McAlister AL. Comparing internet assistance for smoking cessation: 13-month follow-up of a six-arm randomized controlled trial. J Med Internet Res 2008;10(5):e45 [FREE Full text] [doi: 10.2196/jmir.1008] [Medline: 19033146]
54. Saul JE, Schillo BA, Evered S, Luxenberg MG, Kavanaugh A, Cobb N, et al. Impact of a statewide Internet-based tobacco cessation intervention. J Med Internet Res 2007;9(3):e28 [FREE Full text] [doi: 10.2196/jmir.9.4.e28] [Medline: 17942390]
55. Swartz LH, Noell JW, Schroeder SW, Ary DV. A randomised control study of a fully automated internet based smoking cessation programme. Tob Control 2006 Feb;15(1):7-12 [FREE Full text] [doi: 10.1136/tc.2003.006189] [Medline: 16436397]
56. Müller F, Wehbe L. Smoking and smoking cessation in Latin America: a review of the current situation and available treatments. Int J Chron Obstruct Pulmon Dis 2008;3(2):285-293 [FREE Full text] [Medline: 18686737]
57. Instituto Nacional de Estadística. 2009. Encuesta sobre equipamiento y uso de tecnologías de información y comunicación en los hogares 2009 [Survey of facilities and use of information and communication technologies at home 2009] URL: http://www.ine.es/metodologia/t25/t25304506609.pdf [accessed 2013-10-30] [WebCite Cache ID 6Kklu1i7C]
58. Instituto Nacional de Estadística. 2011. Indicadores Sociales 2011 [Social indicators 2011] URL: http://www.ine.es/daco/daco42/sociales11/sociales.htm [accessed 2013-10-30] [WebCite Cache ID 6KkmCQbie]
59. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
60. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
61. Meade CD, Byrd JC, Lee M. Improving patient comprehension of literature on smoking. Am J Public Health 1989 Oct;79(10):1411-1412. [Medline: 2782514]
62. Strecher VJ, Kreuter M, Den Boer DJ, Kobrin S, Hospers HJ, Skinner CS. The effects of computer-tailored smoking cessation messages in family practice settings. J Fam Pract 1994 Sep;39(3):262-270. [Medline: 8077905]
63. González-Rivera JL, De-las-Cuevas C, Rodríguez-Abuin M. El cuestionario de 90 síntomas. Adaptación española del SCL-90-R [The 90-symptoms questionnaire. Spanish adaptation]. Madrid: Publicaciones de Psicología Aplicada, TEA Ediciones; 2002:978-984.
64. Jamal M, Willem Van der Does AJ, Cuijpers P, Penninx BW. Association of smoking and nicotine dependence with severity and course of symptoms in patients with depressive or anxiety disorder. Drug Alcohol Depend 2012 Nov 1;126(1-2):138-146. [doi: 10.1016/j.drugalcdep.2012.05.001] [Medline: 22633368]
65. Jorm AF, Rodgers B, Jacomb PA, Christensen H, Henderson S, Korten AE. Smoking and mental health: results from a community survey. Med J Aust 1999 Jan 18;170(2):74-77. [Medline: 10026688]
66. Verheijden MW, Jans MP, Hildebrandt VH, Hopman-Rock M. Rates and determinants of repeated participation in a web-based behavior change program for healthy body weight and healthy lifestyle. J Med Internet Res 2007;9(1):e1 [FREE Full text] [doi: 10.2196/jmir.9.1.e1] [Medline: 17478410]
67. Wangberg SC, Bergmo TS, Johnsen JA. Adherence in Internet-based interventions. Patient Prefer Adherence 2008;2:57-65 [FREE Full text] [Medline: 19920945]
68. Lewis JA, Machin D. Intention to treat--who should use ITT? Br J Cancer 1993 Oct;68(4):647-650 [FREE Full text] [Medline: 8398686]
69. Tomson T, Björnström C, Gilljam H, Helgason A. Are non-responders in a quitline evaluation more likely to be smokers? BMC Public Health 2005 May 23;5:52 [FREE Full text] [doi: 10.1186/1471-2458-5-52] [Medline: 15910682]
70. Cobb NK, Graham AL, Bock BC, Papandonatos G, Abrams DB. Initial evaluation of a real-world Internet smoking cessation system. Nicotine Tob Res 2005 Apr;7(2):207-216 [FREE Full text] [doi: 10.1080/14622200500055319] [Medline: 16036277]

71. Muñoz RF, Barrera AZ, Delucchi K, Penilla C, Torres LD, Pérez-Stable EJ. International Spanish/English Internet smoking cessation trial yields 20% abstinence rates at 1 year. Nicotine Tob Res 2009 Sep;11(9):1025-1034 [FREE Full text] [doi: 10.1093/ntr/ntp090] [Medline: 19640833]
72. Patrick DL, Cheadle A, Thompson DC, Diehr P, Koepsell T, Kinne S. The validity of self-reported smoking: a review and meta-analysis. Am J Public Health 1994 Jul;84(7):1086-1093. [Medline: 8017530]
73. Barrueco M, Jiménez Ruiz C, Palomo L, Torrecilla M, Romero P, Riesco JA. Veracidad de la respuesta de los fumadores sobre su abstinencia en las consultas de deshabituación tabáquica. Archivos de Bronconeumología 2005;41(3):135-140. [doi: 10.1157/13071584]
74. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
75. Smit ES, Hoving C, Cox VC, de Vries H. Influence of recruitment strategy on the reach and effect of a web-based multiple tailored smoking cessation intervention among Dutch adult smokers. Health Educ Res 2012 Apr;27(2):191-199 [FREE Full text] [doi: 10.1093/her/cyr099] [Medline: 22058288]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
FTQ: Fagerstrom Tolerance Questionnaire
HSI: Heaviness of Smoking Index
ITT: intention-to-treat
PDF: Portable Document Format
PSS-4: 4-item Perceived Stress Scale
SCL 90-R: Symptom Checklist-90-Revised
UNED: Universidad Nacional de Educación a Distancia (National Distance Education University)

Edited by G Eysenbach; submitted 07.11.13; peer-reviewed by R Munoz, R Whittaker; comments to author 05.02.14; revised version received 09.03.14; accepted 24.03.14; published 23.04.14.

Please cite as:
Mañanes G, Vallejo MA
Usage and Effectiveness of a Fully Automated, Open-Access, Spanish Web-Based Smoking Cessation Program: Randomized Controlled Trial
J Med Internet Res 2014;16(4):e111
URL: http://www.jmir.org/2014/4/e111/
doi:10.2196/jmir.3091
PMID:24760951

Original Paper

Adherence to Self-Monitoring via Interactive Voice Response Technology in an eHealth Intervention Targeting Weight Gain Prevention Among Black Women: Randomized Controlled Trial

Dori M Steinberg[1*], RD, PhD; Erica L Levine[1*], MPH; Ilana Lane[1,2*], MA; Sandy Askew[1*], MPH; Perry B Foley[1*], MPH, MSW; Elaine Puleo[3*], PhD; Gary G Bennett[1,2*], PhD

[1]Duke Obesity Prevention Program, Duke Global Health Institute, Duke University, Durham, NC, United States

[2]Department of Psychology and Neuroscience, Duke University, Durham, NC, United States

[3]School of Public Health and Health Sciences, Department of Biostatistics, University of Massachusetts Amherst, Amherst, MA, United States

*all authors contributed equally

Corresponding Author:
Dori M Steinberg, RD, PhD
Duke Obesity Prevention Program
Duke Global Health Institute
Duke University
Room #134 Trent Hall
310 Trent Drive
Durham, NC, 27708
United States
Phone: 1 919 613 5453
Fax: 1 919 249 2146
Email: dori.steinberg@duke.edu

Abstract

Background: eHealth interventions are effective for weight control and have the potential for broad reach. Little is known about the use of interactive voice response (IVR) technology for self-monitoring in weight control interventions, particularly among populations disproportionately affected by obesity.

Objective: This analysis sought to examine patterns and predictors of IVR self-monitoring adherence and the association between adherence and weight change among low-income black women enrolled in a weight gain prevention intervention.

Methods: The Shape Program was a randomized controlled trial comparing a 12-month eHealth behavioral weight gain prevention intervention to usual care among overweight and obese black women in the primary care setting. Intervention participants (n=91) used IVR technology to self-monitor behavior change goals (eg, no sugary drinks, 10,000 steps per day) via weekly IVR calls. Weight data were collected in clinic at baseline, 6, and 12 months. Self-monitoring data was stored in a study database and adherence was operationalized as the percent of weeks with a successful IVR call.

Results: Over 12 months, the average IVR completion rate was 71.6% (SD 28.1) and 52% (47/91) had an IVR completion rate ≥80%. At 12 months, IVR call completion was significantly correlated with weight loss ($r=-.22$; $P=.04$) and participants with an IVR completion rate ≥80% had significantly greater weight loss compared to those with an IVR completion rate <80% (−1.97 kg, SE 0.67 vs 0.48 kg, SE 0.69; $P=.01$). Similar outcomes were found for change in body mass index (BMI; mean difference −0.94 kg, 95% CI −1.64 to −0.24; $P=.009$). Older, more educated participants were more likely to achieve high IVR call completion. Participants reported positive attitudes toward IVR self-monitoring.

Conclusions: Adherence to IVR self-monitoring was high among socioeconomically disadvantaged black women enrolled in a weight gain prevention intervention. Higher adherence to IVR self-monitoring was also associated with greater weight change. IVR is an effective and useful tool to promote self-monitoring and has the potential for widespread use and long-term sustainability.

Trial Registration: Clinicaltrials.gov NCT00938535; http://www.clinicaltrials.gov/ct2/show/NCT00938535 (Archived by WebCite at http://www.webcitation.org/6P1FFNJRs).

(J Med Internet Res 2014;16(4):e114) doi:10.2196/jmir.2996

KEYWORDS

eHealth; obesity; interactive voice response; self-monitoring

Introduction

During the past decade, a growing body of evidence has demonstrated the efficacy of electronic health (eHealth) interventions for weight management [1,2]. These interventions have been tested on a range of digital platforms (eg, Web, text messaging, mobile applications) and show promise in overcoming some of the challenges inherent with traditional weight loss interventions (eg, cost, reach, scale, expertise) [3]. Although intervention designs vary considerably, evidence suggests that eHealth interventions can produce clinically meaningful weight loss outcomes [2]. Such findings are promising, particularly among racial/ethnic minority groups who have high utilization of mobile and Web-based technologies [4,5] and bear the most burden from obesity [6].

Most weight control interventions promote some form of self-monitoring, usually recommending that participants provide detailed reports of diet, physical activity, weight, or obesity-related risk behaviors (eg, sugar sweetened beverages) using paper diaries. Indeed, evidence indicates that self-monitoring is highly predictive of weight loss success [7,8]. Despite its effectiveness, adherence to traditional paper-based approaches declines over time [7,9].

eHealth approaches offer unique features that may help abate the usual decline in self-monitoring adherence. eHealth self-monitoring strategies (eg, Web-based dietary monitoring, mobile applications with food diaries, activity trackers) are often more portable, allow for more proximal reporting, can prompt individuals based on timing, context, or participant progress, and have the ability to provide immediate and tailored feedback [10]. Furthermore, qualitative evidence indicates that participants are more receptive to eHealth approaches compared to paper-based methods [11,12]. Identifying effective eHealth self-monitoring strategies that can further enhance adherence is important because evidence consistently demonstrates that the magnitude of behavior change in eHealth interventions is largely dependent on the level of participant adherence or engagement with the intervention [1].

Interactive voice response (IVR) is one such eHealth self-monitoring approach. IVR allows participants to interact with a computer system via outbound or inbound telephone calls using the keypad or speech. Use of IVR is ubiquitous in the wider consumer market (eg, used with telephone banking, checking airline flight status, automated appointment reminders with health systems, etc) and, given its widespread familiarity, might be an effective way to collect self-monitoring data within health interventions.

IVR has been used in a variety of clinical contexts as a means of both delivering intervention content and collecting data [13-17]. IVR may have a number of distinct advantages over the use of other eHealth modalities [18]. It can be particularly useful for low literacy populations [19]: the task of listening to a voice prompt and responding with a simple numerical answer may be far less cognitively and numerically demanding than producing detailed reports of self-monitoring data (eg, caloric intake or fat intake). IVR calls may also be less time-consuming than other modalities for self-monitoring that might require participants to keep a paper food diary, log on to an online system, or conduct an extensive search for required numerical data (eg, calorie intake) [20,21]. Additionally, IVR systems can be used to provide dynamic and immediate feedback in response to self-monitoring data. Indeed, a review of the evidence suggests improvements in health outcomes with the use of IVR for self-monitoring [17].

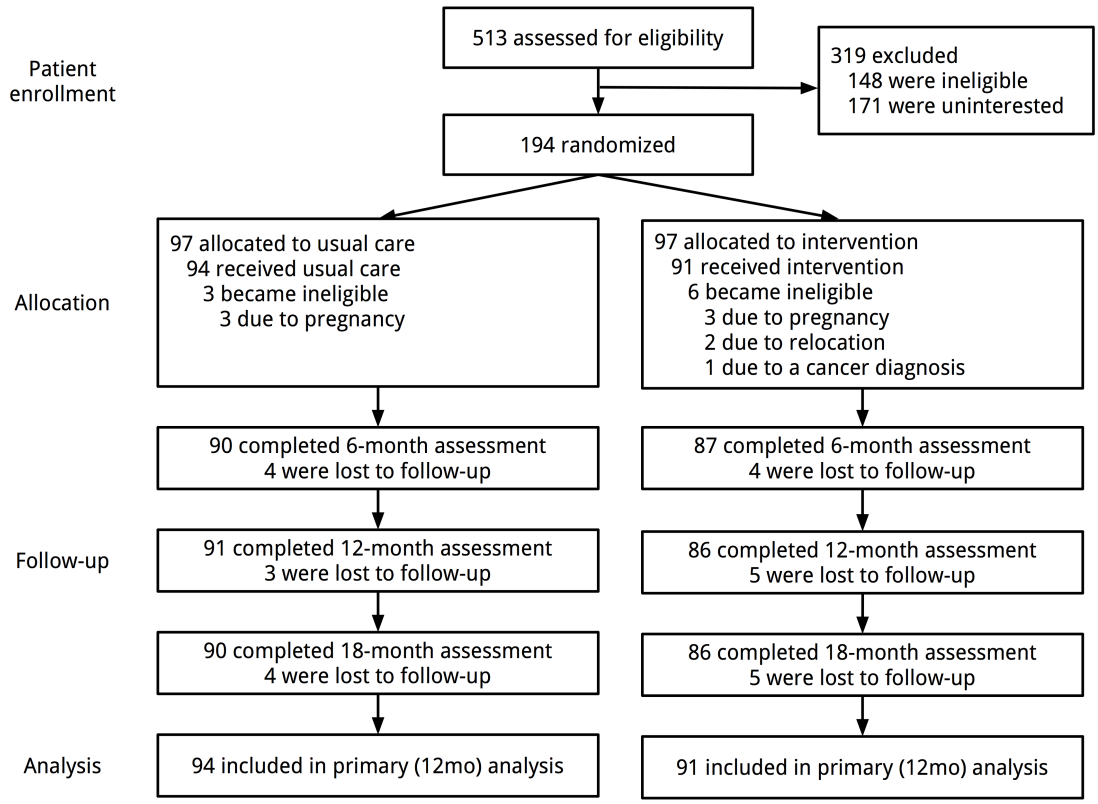
Despite the growing literature surrounding the use of IVR technologies, limited evidence exists on the use and effectiveness of IVR for weight control [22,23]. We sought to examine the association between IVR self-monitoring and weight change among socioeconomically disadvantaged black women enrolled in the Shape Program (“Shape”). Shape was an 18-month randomized controlled trial comparing an eHealth weight gain prevention intervention to usual care among overweight and class 1 obese black female primary care patients [24]. Intervention participants self-monitored their behavior change goals via weekly IVR phone calls. Findings from Shape indicate that the intervention was successful in preventing weight gain among intervention participants relative to those receiving standard primary care [25]. At 12 months, a larger proportion of intervention participants (62%) were at or below their baseline weight, compared to those in usual care (45%; *P*=.02). Similar findings were observed at 18 months [25]. In the present analysis, we describe patterns of IVR self-monitoring adherence over time, examine relevant predictors of adherence, and explore the association between adherence and weight change.

Methods

Study Design

The Shape Program design and methods have been detailed elsewhere [24,25]. Starting in December 2009, participants were recruited via mail from five community health centers operated by Piedmont Health (PHS) in central North Carolina. Participants were black women, aged 25 to 44 years, with a body mass index (BMI) of 25-34.9 kg/m^2. Following eligibility screening, informed consent, and baseline measures, we randomized participants (n=194) to either the Shape intervention or usual care arm. All participants were re-evaluated at 6 and 12 months, with additional follow-up at 18 months post randomization (Figure 1). Final assessments were completed in October 2012. The relevant university and health system review boards approved all study procedures.

Figure 1. Participant enrollment and retention (CONSORT).



Intervention Description

The Shape intervention included five main components: (1) behavior change goals known to promote weight change, (2) self-monitoring of these goals via weekly IVR phone calls, (3) tailored skills training materials, (4) monthly interpersonal counseling calls with a PHS registered dietitian (“Shape coach”), and (5) a 12-month YMCA membership. Usual care group participants received routine standard of care from their providers at PHS.

The intervention utilized the interactive obesity treatment approach (iOTA), which has been extensively tested in previous studies [23,26]. iOTA uses an algorithm to prescribe tailored, evidence-based behavior change goals in order to create a sufficient caloric deficit to produce weight change (eg, five or more fruits and vegetables/day, no fast food, no sugar sweetened beverages, walking 7000 steps/day). Participants do not self-select which goals to track. Rather, participants are assigned through the algorithm three behavior change goals from a library of 21 goals based on their self-efficacy and readiness, and the potential for the goal to produce a caloric deficit. For novelty and to ensure that participants changed multiple behaviors, goals were updated every two months based on the output from the original algorithm.

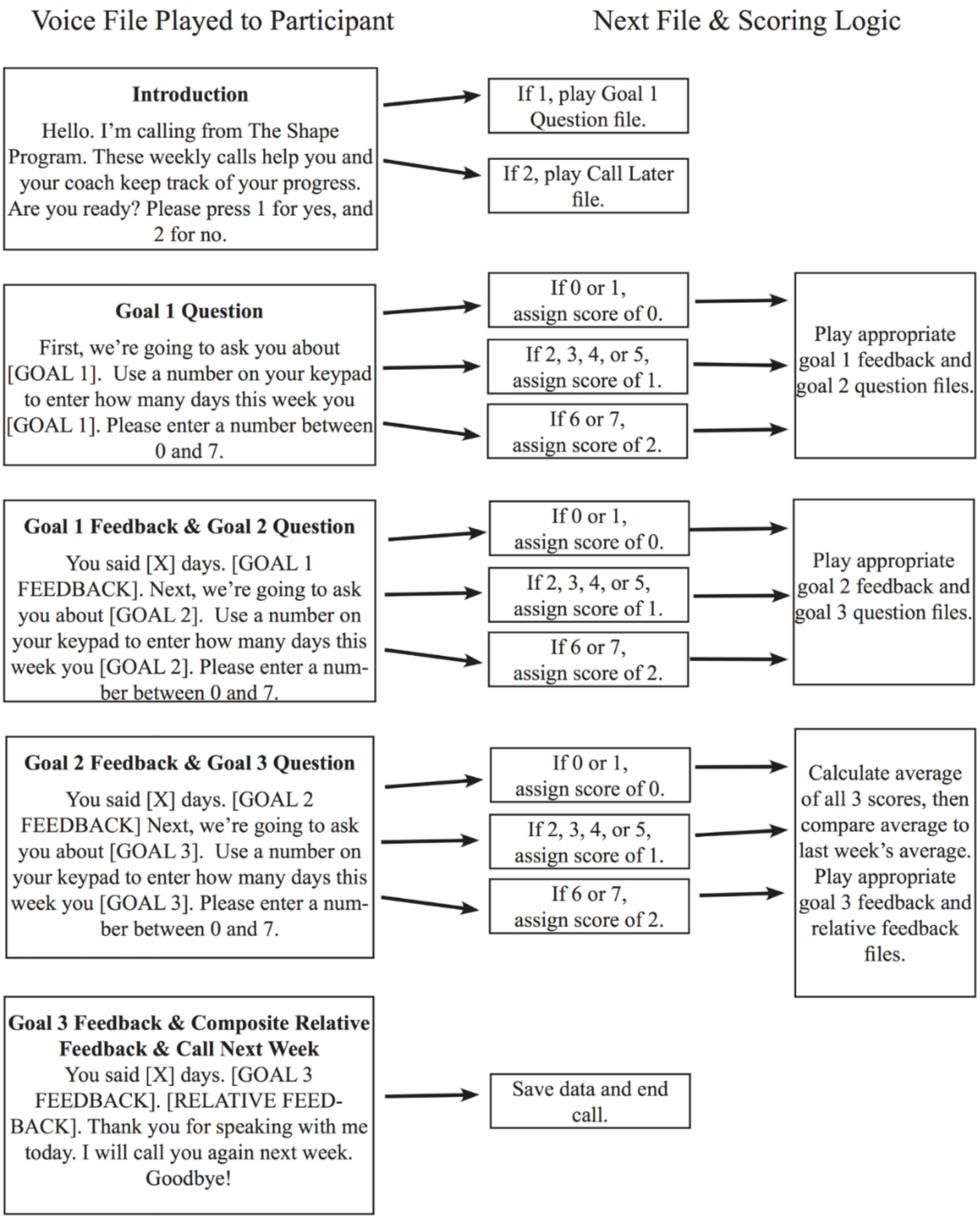
Participants self-monitored these goals throughout the 12-month intervention via weekly IVR phone calls. The IVR calls were on average 2-4 minutes in duration. The IVR system called each participant once a week at a predetermined time. If a participant was not reached on the initial attempt, an extensive retry protocol was put into place, with a maximum of 16 attempts over two days. As shown in Figure 2, once the IVR system made contact with the participant, it asked how many days this past week they achieved each of their assigned behavioral goals (eg, “this past week, how many days did you drink sugary drinks?”). When relaying the self-monitoring data, participants had the option to reference a paper-based goal-tracking sheet that included daily reporting of these goals. For example, if a participant’s goal was “no sugary drinks”, she might report in her paper log each day whether or not she drank any sugary drinks. At the end of the week, she summed the days she drank any sugary drinks. The IVR questions were phrased in such a way as to be concrete and dichotomous, making it easier to remember goal achievement during the weekly calls (eg, “How many days this week did you drink sugary drinks?”). Thus, the paper log was considered optional. Participants were encouraged to use the paper logs only as a means to help them relay the data through the IVR system. However, the IVR system was designed to be simple enough to use without the paper logs

After self-monitoring data was collected and stored in a study database, brief tailored feedback and short skills training tips were immediately provided. Based on goal performance, a score was assigned to each goal (eg, if a participant reported drinking sugary drinks zero days last week, she received a high score of

2). An average goal attainment score across all behaviors determined feedback messages, which were pulled from a pre-determined set of feedback messages that were transferred to voice files (eg, “Looking at all of your goals together,…”). The algorithm included a component to prevent repeated feedback messages within a set number of weeks.

Feedback messages (Figure 2) included a relative comparison to the previous week’s self-monitoring data (eg, “...you did better than last time—great job!” or “...you’re not doing quite as well as you did last week. Let’s turn this slip around”) and specific tips on how to boost performance (eg, “Think about the things that you did on the day you met your goal and how you can you do those things more often next week” or “You’re doing great! Stick with it and you can get to 7 days next week”). A sample call can be heard in Multimedia Appendix 1. Messaging content was selected from a large library of feedback messages from previous studies using the iOTA approach in similar populations [23,26,27]. All IVR call logic was rigorously tested and continuous quality control protocols were performed to ensure fidelity to protocol. Adherence to IVR tracking was also encouraged during the monthly coaching calls and participants were aware that IVR data were relayed to the coaches. The coaches had access to IVR completion rates for each participant and provided feedback and counseling on strategies to maintain weekly tracking.

Figure 2. IVR call logic.



Measures

Self-Monitoring

The IVR system collected and stored response data to each of the three goals, as well as data on call time and call length in a study database. Self-monitoring adherence was calculated as the proportion of intervention participants who successfully completed IVR calls over the number of expected to complete a call by study week. Calls were deemed successful once data on each of the three goals were received. Some participants (n=9) requested to suspend or stop intervention activities and/or experienced technical problems with the IVR system. We will assess self-monitoring adherence with and without these participants. Although IVR was the primary self-monitoring mode, participants were given the option to use paper logs daily. At 12 months, participants self-reported via an online questionnaire the average number of days per week they used the paper log (ie, 5-7 days per week, 3-4 days per week, 1-2 days per week, or not at all). We also assessed perceptions about IVR self-monitoring at 12 months using an online questionnaire. Participants reported agreement via a 6-point Likert scale that ranged from “strongly agree” to “strongly disagree” on various statements assessing perceptions about IVR self-monitoring (eg, the tracking calls made it easy for me to keep track of my behaviors, the tracking calls were difficult to use, or I enjoyed receiving the tracking calls).

Determinants of Adherence

We selected several relevant baseline sociodemographic variables and psychosocial constructs that might predict self-monitoring adherence based on behavior change theories [28] and previous literature suggesting psychosocial variables such as stress [29], perceived social support [30], and self-efficacy [31] may impact engagement and outcomes in weight control trials. Demographic variables such as age, marital status, educational attainment (five categories: less than high school, high school, vocational school, some college, or college degree or more), employment, and income (US $30,000/year or more vs less than US $30,000/year) were collected. The 8-item Patient Health Questionnaire assessed the presence of depressive symptoms [32]. The 19-item subscale from the Medical Outcomes Study (MOS-SSS) was utilized to assess availability of social support. Four subscales were included: emotional/informational, tangible, affectionate, and positive social interaction [33]. A 16-item questionnaire that was used in the CARDIA study was administered to measure frequency of stressful life events [34]. The Marcus self-efficacy for exercise questionnaire was used to assess confidence in one’s ability to exercise when tired, in a bad mood, don’t have time, on vacation, or when it is raining/snowing [35]. All measures were administered via online questionnaires.

Given that the Shape coaches were able to view IVR call patterns, coaching call completion was assessed as another potential predictor. The monthly coaching calls were delivered via a similar software system as the IVR calls. As a result, we were able to capture start and end time of each call. We used call duration data, along with coach documentation of topics covered, as a proxy of completion. Coaching call completion was operationalized as the actual number completed over the number expected.

Anthropometrics

Study staff collected weight and height data at baseline, 6, and 12 months within study offices. Body weights were measured to the nearest 0.1 kg using a portable electronic scale (Seca Model 876) and heights were measured using a calibrated wall-mounted stadiometer (Seca 214) [36].

Statistical Analyses

All analyses were conducted within the intervention group only (n=91). Descriptive statistics were conducted to characterize the sample and examine average IVR completion rate over the 12-month period. IVR adherence was dichotomized using a median split (80% or more) to examine differences in outcomes among high completers compared to those below the median. Adherence was also analyzed as tertiles of successful weekly IVR calls. We conducted bivariate analyses using *t* tests and chi-square to examine potential predictors of average IVR completion rate and categories of IVR completion. Pearson correlations examined the relationship between weight and BMI change and IVR call completion rate. ANOVA (analysis of variance) analyzed differences in weight change and BMI change among high and low IVR completers, tertiles of IVR completion, and categories of IVR and paper log completion. Last, given that participants were nested within health centers, we tested the intraclass correlation (ICC) and did not find a meaningful effect of the nested design (ICC=.07; 95% CI 0.01-0.38). Therefore, no further adjustment was required.

Results

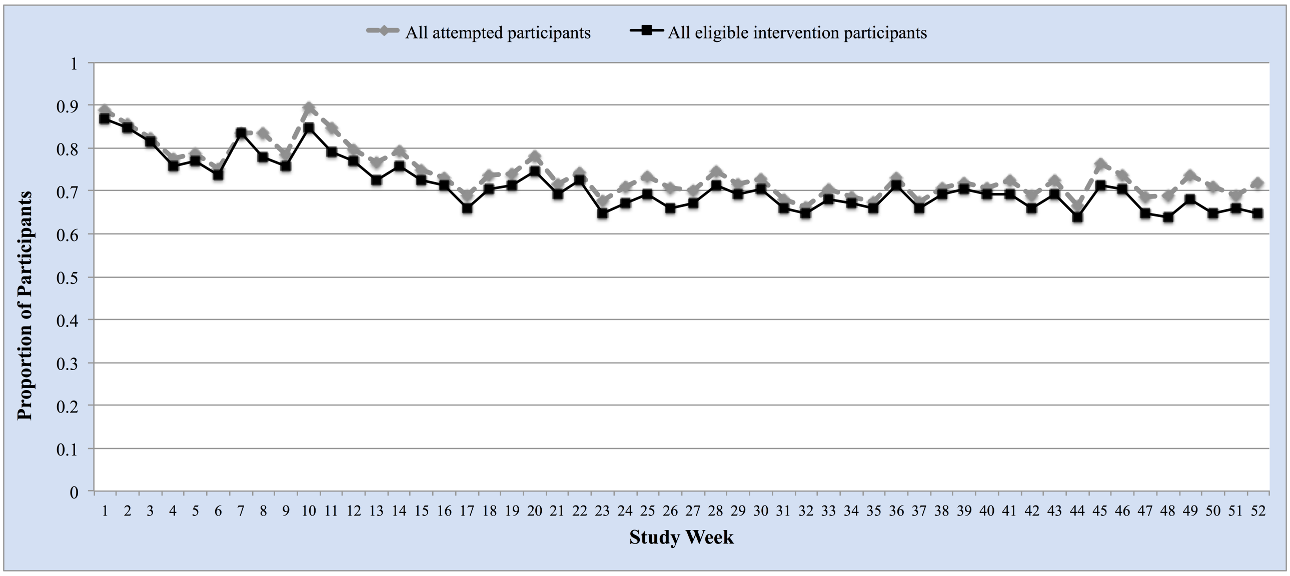
Baseline Characteristics and Retention

Baseline characteristics and main outcomes have been reported in detail elsewhere [24,25]. Briefly, participants reported at baseline a mean age of 35.4 years (SD 5.5) and a mean BMI of 30.2 kg/m^2 (SD 2.5). Most (71.4%, 130/182) were currently employed with an annual income <US $30,000/year (74.3%, 136/183). The majority (79.7%, 145/182) had less than a college degree. We retained 95.7% (177/185) of participants at 12- and 18-months post randomization (Figure 1) and there were no statistically significant differences in attrition between groups.

Self-Monitoring Adherence

Figure 3 shows IVR adherence rates over time by study week. Among all intervention participants (n=91), the average IVR completion rate over 12 months was 71.6% (SD 28.1) with a weekly range from 52% to 96%. Similar results are seen among all attempted participants; this rate excludes participants at each study week that may have requested to suspend or stop intervention activities and/or experienced technical problems with the IVR system (82/91 at week 52). A total of 52% (47/91) of intervention participants had an IVR completion rate of 80% or more, and two-thirds (66%, 60/91) completed at least 60% of IVR calls. Throughout the 12-month period, 39% (20/52) of IVR calls were completed on the first attempt and 78% (40/52) of calls were completed by the third attempt. About half of participants (49%, 41/83) self-reported using the paper tracking log at least 5 days each week in order to relay behavioral goal attainment to the weekly IVR calls.

Figure 3. Proportion of participants who completed IVR calls by study week.



Predictors of Adherence

Age and education were the only sociodemographic or psychosocial variables that significantly predicted IVR adherence. High IVR completers (≥80%) were older (*P*=.03) and more educated (*P*=.003) than participants who completed <80% of IVR calls. Similar findings were found when assessing IVR call completion as a continuous variable (data not shown). Intervention participants completed 82% (10/12) of counseling calls during the 12-month intervention period. Coaching call completion was highly correlated with IVR completion rate (*r*=.77; *P*<.001). The percent of coaching calls completed was also predictive of the odds of completing 80% of IVR calls. Each 10% increase in coaching calls completed was associated with almost three times the odds of completing 80% or more of IVR calls (OR 2.57, 95% CI 1.59-4.26; *P*<.001).

Self-Monitoring Adherence and Weight Change

IVR call completion was significantly correlated with 12-month weight loss (Spearman's *r*=−.22; *P*=.04). Tables 1 and 2 show the associations between specific thresholds of IVR adherence and weight outcomes. At 12 months, participants with an IVR completion rate of at least 80% had greater weight loss outcomes compared to those with an IVR completion rate of less than 80% (mean difference −2.45 kg, 95% CI −4.37 to −0.54; *P*=.01). We observed similar findings for 12-month change in BMI (mean difference −0.94 kg/m^2, 95% CI −1.64 to −0.24; *P*=.009). Although differences in weight change by tertiles of IVR call completion did not reach significance (Table 2), there were significant differences in BMI change between those who completed less than 60% of IVR calls and those who completed at least 93% of IVR calls (mean difference −0.87 kg/m^2, 95% CI −1.73 to −0.01; *P*=.047). Similar results were seen when comparing those who completed less than 60% to 60-93% call completion (mean difference −0.88 kg/m^2, 95% CI −1.76 to −0.002; *P*=.049).

As determined by self-reported paper log use at 12 months, the use of the paper logs did not enhance weight outcomes beyond what was achieved by the use of IVR. Although it did not reach statistical significance, at 12 months, participants with high IVR completion (≥80%) and high self-reported paper log use (≥ 5 days per week) lost 1.94 kg (SE 1.2), while those with lower IVR completion, but high self-reported paper log use gained 1.38 kg (SE 1.3) (mean difference −3.32 kg/m^2, 95% CI −7.53 to 0.89; *P*=.17). Similar findings were seen when comparing participants with high IVR completion, but low self-reported paper log use [−2.04 kg (SE 0.76)] to those with lower IVR completion and low tracking log use [0.30 kg (SE 0.86); mean difference −2.34 kg/m^2, 95% CI −6.35 to 1.67; *P*=.42].

Table 1. Change in weight and body mass index by IVR call completion (n=91).[a]

Anthropometric Changes	Time	IVR<80%, mean (SE) (n=44)	IVR ≥80%, mean (SE) (n=47)	Difference, mean (95% CI)[c]
Change in weight, kg				
	Month 6	−0.56 (0.57)	−1.16 (0.56)	−0.60 (−2.19 to 0.99)
	Month 12	0.48 (0.69)	−1.97 (0.67)	−2.45 (−4.37 to −0.54)
Change in body mass index [b]				
	Month 6	−0.17 (0.21)	−0.34 (0.21)	−0.17 (−0.76 to 0.42)
	Month 12	0.25 (0.25)	−0.70 (0.25)	−0.94 (−1.64 to −0.24)

[a] Denominators vary because of missing data.

[b] Calculated as weight in kilograms divided by height in meters squared.

[c] Confidence intervals that do not contain zero have a *P* value <.05.

Table 2. Change in weight and body mass index by tertiles of IVR call completion (n=91).[a]

Anthropometric changes	Time				Difference between tertiles, mean (95% CI)[c]		
		Tertile 1 IVR<60% (n=30)	Tertile 2 IVR 60-92% (n=29)	Tertile 3 IVR ≥93% (n=32)	Between 1st and 2nd	Between 1st and 3rd	Between 2nd and 3rd
Change in weight, kg, mean (SE)							
	Month 6	−0.46 (0.70)	−1.50 (0.71)	−0.69 (0.67)	−1.04 (−3.01 to 0.94)	−0.23 (−2.16 to 1.69)	0.81 (−1.14 to 2.75)
	Month 12	0.78 (0.85)	−1.60 (0.87)	−1.51 (0.82)	−2.39 (−4.80 to 0.02)	−2.29 (−4.64 to 0.06)	−0.09 (−2.28 to 2.46)
Change in body mass index [b] **, mean (SE)**							
	Month 6	−0.18 (0.26)	−0.44 (0.26)	−0.16 (0.25)	−0.27 (−1.00 to 0.7)	0.02 (−0.70 to 0.73)	0.28 (−0.44 to 1.01)
	Month 12	0.35 (0.31)	−0.54 (0.32)	−0.52 (0.30)	−0.88 (−1.76 to −0.002)	−0.87 (−1.73 to −0.01)	0.01 (−0.85 to 0.88)

[a] Denominators vary because of missing data.

[b] Calculated as weight in kilograms divided by height in meters squared.

[c] Confidence intervals that do not contain zero have a *P* value <.05.

Perceptions of IVR Self-Monitoring

Generally, Shape participants perceived IVR self-monitoring positively. Most (89%, 73/82) agreed that the IVR calls made it easy to self-monitor behavioral goals and 72% (59/82) strongly disagreed that using IVR-based self-monitoring was difficult. A majority (62%, 50/81) reported that the IVR calls were enjoyable and more than half (56%, 46/82) reported that it was easy to fit self-monitoring via IVR into their daily routine. Most (83%, 67/81) reported that IVR self-monitoring made it easy to share information with their Shape coaches; however, only 7% (6/83) said that they answered the IVR calls because they knew the coaches would see their data. Rather, a majority of participants (66%, 55/83) reported that the motivation for answering the IVR calls was to stay on track with their behavioral goals. Most (84%, 68/81) said weekly self-monitoring via IVR was the appropriate frequency and 91% (73/80) reported that the duration of the calls was just right.

Discussion

Principal Findings

We found high adherence to weekly IVR self-monitoring calls among low-income black women enrolled in a weight gain prevention intervention. Over the 12-month intervention, nearly three-quarters were adherent to the self-monitoring protocol and more than half of the women completed at least 80% of the 52 IVR calls. Adherence was higher for older, more educated women. Although weight loss was unintended in this trial, we found a positive relation between self-monitoring adherence and weight change; those who completed at least 80% of calls lost almost 2.5 kg more than those with lower adherence. We provided daily paper-based logs as an additional self-monitoring option, but use of the paper-based approach did not enhance IVR adherence or weight loss outcomes. Most participants reported that IVR self-monitoring was easy, helpful, and fit into their daily routine. Compliance with the monthly coaching calls also helped enhance adherence to IVR self-monitoring. We conclude that IVR self-monitoring is effective, produces high adherence rates, and has the potential for greater sustainability in a socioeconomically disadvantaged patient population.

Self-monitoring adherence is one of the strongest predictors of weight outcomes [7]. eHealth approaches can enhance self-monitoring adherence by reducing some of the barriers typically associated with paper-based approaches; however, adherence rates do vary by eHealth modality. Burke and colleagues tested the use of a personal digital assistant (PDA) with or without feedback compared to paper-based self-monitoring within the context of a weight loss intervention and found that less than half of participants remained adherent to self-monitoring using PDAs, and only 30% remained adherent to the paper-based approach at 6 months [37]. By 18 months, rates of adherence decreased below 20% for all groups [9]. Tate and colleagues found that adherence to a Web-based self-monitoring food and exercise diary averaged around 50% by the end of a 12-month intervention [38]. Despite the portability and convenience of text messaging for self-monitoring, adherence rates to text message self-monitoring also average around 50-60% [39,40]. We found that IVR self-monitoring produced higher adherence compared to these other eHealth self-monitoring approaches and remains significantly higher than paper-based methods. Given that high adherence equates to better behavior change outcomes, IVR may be more effective than other eHealth approaches; however, comparative effectiveness studies are needed.

Our high adherence rates may be a result of the type of self-monitoring and frequency of self-monitoring required in addition to the mode through which participants' monitored. Most weight loss trials ask participants to keep a detailed daily diary of complex aspects of dietary intake and exercise. This requires participants to measure food and perform mathematical functions such as counting calories or grams of fat. This can be difficult to sustain for extended periods of time. Indeed, adherence to these approaches declines over time [21]. Our study, in contrast, asked participants to track weekly a limited number (3) of discrete, simple behavior change goals associated with weight loss (eg, eat 5 or more fruits and vegetables per day, no sugary drinks, 10,000 steps/day) in order to achieve the desired caloric deficit. Self-monitoring of specific behaviors rather than detailed dietary records may be less burdensome and therefore easier to continue over long periods of time. It may also be easier to recall during brief weekly IVR calls. Given our study design, it is not possible determine whether the mode, frequency, or type of self-monitoring was driving the high adherence rates.

This study is among the first to provide detailed evidence for the utility of IVR self-monitoring for weight gain prevention. To our knowledge, only two other studies [22,23] have utilized IVR in a weight loss trial. Estabrooks and colleagues [22] similarly found a 75% adherence rate in their 3-month weight loss trial; although participants were only required to complete 12 calls, while our trial included 52 calls. Additionally, participants in the Estabrooks et al study were predominantly white (69%; 5% black) and more socioeconomically advantaged than Shape participants [22]. Bennett and colleagues [23] also utilized IVR as an intervention tool within a 24-month randomized-controlled effectiveness trial among older, obese, hypertensive patients with similar sociodemographic characteristics to our study, although both males and females were included. Intervention participants could choose to self-monitor their progress on behavior change goals using either IVR or a study website, with the majority of participants (61%) choosing IVR. In terms of adherence, the IVR call completion rates were similar to other eHealth modalities; on average, 57% of IVR calls were completed at 12 months and 48% by 24 months [23]. Our higher adherence rates may be a result of a number of methodological and sociodemographic differences, such as age, comfort with technology, and the fact that our study focused on weight gain prevention while the Bennett et al study focused on achieving weight loss.

IVR has particular promise for socioeconomically disadvantaged populations because it is telephone-based and does not rely on Internet connectivity as is required for other eHealth approaches (eg, Web-based, mobile tracking). These findings are important as these populations, particularly black women, have the highest prevalence of obesity compared to any other group [6], and achieving clinically meaningful weight loss among low-income black women, in particular, has been challenging [41]. Across numerous studies, black women achieve smaller weight losses compared to other groups [42]. It is not clear whether self-monitoring adherence varies by race/ethnicity, as there appear to be no studies in the literature that have examined racial/ethnic disparities in adherence rates. One might suspect that the poorer weight loss outcomes might be, in part, driven by poorer adherence rates. We found that 50% of the sample achieved greater than 80% IVR call completion, which was associated with greater weight loss. Given these high rates of adherence, IVR would appear indicated in future weight control interventions that target these high-risk populations.

Strengths and Limitations

Our study is one of the first to examine the utility of IVR technology for self-monitoring within a weight management intervention. We examined self-monitoring adherence among a population typically underrepresented in weight control research and for whom obesity treatment is of clinical importance. We had a long follow-up period and maintained high adherence and retention throughout the 52-week intervention. Furthermore, we tested a unique goal-oriented self-monitoring approach for self-monitoring that is less cumbersome compared to more traditional detailed monitoring. This approach may be more effective and sustainable, particularly for high-risk populations. Although we sustained high adherence at 12 months, longer-term follow-up would help determine the true sustainability of an IVR-based approach. Our findings are conservative as we chose to report adherence rates among all eligible intervention participants and not disaggregate participants who experienced technical problems from those who chose to stop intervention activities. Future research would benefit from a more detailed account of the potential causes of low adherence. With the current study design, it is not clear whether IVR self-monitoring is more effective than other eHealth modes. Comparative effective studies are necessary to determine the most effective approach for self-monitoring. Last, this study examined the utility of IVR self-monitoring within the context of a weight maintenance intervention among black women in the primary care setting; thus, we cannot infer whether IVR as the main self-monitoring strategy would be similarly effective within the context of a weight loss intervention in different populations and settings.

Conclusions

IVR technology is a promising goal-oriented self-monitoring tool within weight control interventions, particularly for high-risk populations. Using this technology produced adherence rates that were higher than other eHealth approaches to self-monitoring. It was also more favorably received than other approaches to self-monitoring. Given the ubiquity of mobile phones, particularly among racial/ethnic minority populations [5,43], IVR can be a useful tool to promote self-monitoring and facilitate intervention delivery.

Acknowledgments

We express deep gratitude to the administration and staff of Piedmont Health for their continued collaboration and participation in the Shape Program. Most importantly, we would especially like to thank the women participating in Shape.

This trial is funded by grant R01DK078798 from the National Institute for Diabetes and Digestive and Kidney Diseases. Dr Bennett was supported by K22CA126992. The National Institute for Diabetes and Digestive and Kidney Diseases had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; or preparation, review, or approval of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Sample IVR call.

[WAV File, 7MB - jmir_v16i4e114_app1.wav]

Multimedia Appendix 2

CONSORT-EHEALTH checklist V1.6.2 [44].

[PDF File (Adobe PDF File), 989KB - jmir_v16i4e114_app2.pdf]

References

1. 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]
2. 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]
3. Ritterband LM, Tate DF. The science of internet interventions. Ann Behav Med 2009 Aug;38(1):1-3. [doi: 10.1007/s12160-009-9132-5] [Medline: 19816750]
4. Smith A. Pew Internet and the American Life Project. Washington, DC: Pew Research Center; 2010 Sep 17. Technology trends among people of color URL: http://www.pewinternet.org/2010/09/17/technology-trends-among-people-of-color/ [accessed 2014-04-11] [WebCite Cache ID 6OlJMXndV]
5. Fox S, Duggan M. Internet and American Life Project. Washington, DC: Pew Research Center; 2012 Nov 08. Mobile Health 2012 URL: http://www.pewinternet.org/2012/11/08/mobile-health-2012/ [accessed 2014-04-11] [WebCite Cache ID 6OlJTDGxZ]
6. 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]
7. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
8. Sarwer DB, von Sydow Green A, Vetter ML, Wadden TA. Behavior therapy for obesity: where are we now? Curr Opin Endocrinol Diabetes Obes 2009 Oct;16(5):347-352. [doi: 10.1097/MED.0b013e32832f5a79] [Medline: 19623061]
9. Burke LE, Styn MA, Sereika SM, Conroy MB, Ye L, Glanz K, et al. Using mHealth technology to enhance self-monitoring for weight loss: a randomized trial. Am J Prev Med 2012 Jul;43(1):20-26. [doi: 10.1016/j.amepre.2012.03.016] [Medline: 22704741]

10. Coons MJ, Roehrig M, Spring B. The potential of virtual reality technologies to improve adherence to weight loss behaviors. J Diabetes Sci Technol 2011 Mar;5(2):340-344 [FREE Full text] [Medline: 21527103]
11. Glanz K, Murphy S, Moylan J, Evensen D, Curb JD. Improving dietary self-monitoring and adherence with hand-held computers: a pilot study. Am J Health Promot 2006;20(3):165-170. [Medline: 16422134]
12. Burke LE, Swigart V, Warziski Turk M, Derro N, Ewing LJ. Experiences of self-monitoring: successes and struggles during treatment for weight loss. Qual Health Res 2009 Jun;19(6):815-828 [FREE Full text] [doi: 10.1177/1049732309335395] [Medline: 19365099]
13. Siegel K, Mesagno FP, Karus DG, Christ G. Reducing the prevalence of unmet needs for concrete services of patients with cancer. Evaluation of a computerized telephone outreach system. Cancer 1992 Apr 1;69(7):1873-1883. [Medline: 1551071]
14. Christ G, Siegel K. Monitoring quality-of-life needs of cancer patients. Cancer 1990 Feb 1;65(3 Suppl):760-765. [Medline: 2302653]
15. Hettema JE, Hosseinbor S, Ingersoll KS. Feasibility and reliability of interactive voice response assessment of HIV medication adherence: research and clinical implications. HIV Clin Trials 2012;13(5):271-277 [FREE Full text] [doi: 10.1310/hct1305-271] [Medline: 23134627]
16. Baer L, Jacobs DG, Cukor P, O'Laughlen J, Coyle JT, Magruder KM. Automated telephone screening survey for depression. JAMA 1995 Jun 28;273(24):1943-1944. [Medline: 7783305]
17. Piette JD. Interactive voice response systems in the diagnosis and management of chronic disease. Am J Manag Care 2000 Jul;6(7):817-827 [FREE Full text] [Medline: 11067378]
18. Abu-Hasaballah K, James A, Aseltine RH. Lessons and pitfalls of interactive voice response in medical research. Contemp Clin Trials 2007 Sep;28(5):593-602. [doi: 10.1016/j.cct.2007.02.007] [Medline: 17400520]
19. Al-Tayyib AA, Rogers SM, Gribble JN, Villarroel M, Turner CF. Effect of low medical literacy on health survey measurements. Am J Public Health 2002 Sep;92(9):1478-1480. [Medline: 12197979]
20. Burke LE, Sereika S, Choo J, Warziski M, Music E, Styn M, et al. Ancillary study to the PREFER trial: a descriptive study of participants' patterns of self-monitoring--rationale, design and preliminary experiences. Contemp Clin Trials 2006 Feb;27(1):23-33. [doi: 10.1016/j.cct.2005.10.006] [Medline: 16311076]
21. Burke LE, Conroy MB, Sereika SM, Elci OU, Styn MA, Acharya SD, et al. The effect of electronic self-monitoring on weight loss and dietary intake: a randomized behavioral weight loss trial. Obesity (Silver Spring) 2011 Feb;19(2):338-344 [FREE Full text] [doi: 10.1038/oby.2010.208] [Medline: 20847736]
22. Estabrooks PA, Smith-Ray RL. Piloting a behavioral intervention delivered through interactive voice response telephone messages to promote weight loss in a pre-diabetic population. Patient Educ Couns 2008 Jul;72(1):34-41. [doi: 10.1016/j.pec.2008.01.007] [Medline: 18282679]
23. Bennett GG, Warner ET, Glasgow RE, Askew S, Goldman J, Ritzwoller DP, Be Fit, Be Well Study Investigators. Obesity treatment for socioeconomically disadvantaged patients in primary care practice. Arch Intern Med 2012 Apr 9;172(7):565-574. [doi: 10.1001/archinternmed.2012.1] [Medline: 22412073]
24. Foley P, Levine E, Askew S, Puleo E, Whiteley J, Batch B, et al. Weight gain prevention among black women in the rural community health center setting: the Shape Program. BMC Public Health 2012;12:305 [FREE Full text] [doi: 10.1186/1471-2458-12-305] [Medline: 22537222]
25. Bennett GG, Foley P, Levine E, Whiteley J, Askew S, Steinberg DM, et al. Behavioral treatment for weight gain prevention among black women in primary care practice: a randomized clinical trial. JAMA Intern Med 2013 Oct 28;173(19):1770-1777. [doi: 10.1001/jamainternmed.2013.9263] [Medline: 23979005]
26. Bennett GG, Herring SJ, Puleo E, Stein EK, Emmons KM, Gillman MW. Web-based weight loss in primary care: a randomized controlled trial. Obesity (Silver Spring) 2010 Feb;18(2):308-313 [FREE Full text] [doi: 10.1038/oby.2009.242] [Medline: 19696764]
27. Greaney ML, Quintiliani LM, Warner ET, King DK, Emmons KM, Colditz GA, et al. Weight management among patients at community health centers: The “Be Fit, Be Well” Study. Obesity and Weight Management 2009 Oct;5(5):222-228. [doi: 10.1089/obe.2009.0507]
28. Elder JP, Ayala GX, Harris S. Theories and intervention approaches to health-behavior change in primary care. Am J Prev Med 1999 Nov;17(4):275-284. [Medline: 10606196]
29. Cox TL, Krukowski R, Love SJ, Eddings K, DiCarlo M, Chang JY, et al. Stress management-augmented behavioral weight loss intervention for African American women: a pilot, randomized controlled trial. Health Educ Behav 2013 Feb;40(1):78-87. [doi: 10.1177/1090198112439411] [Medline: 22505570]
30. Jeffery RW, Bjornson-Benson WM, Rosenthal BS, Lindquist RA, Kurth CL, Johnson SL. Correlates of weight loss and its maintenance over two years of follow-up among middle-aged men. Prev Med 1984 Mar;13(2):155-168. [Medline: 6739444]
31. 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]
32. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
33. Sherbourne CD, Stewart AL. The MOS social support survey. Soc Sci Med 1991;32(6):705-714. [Medline: 2035047]

34. Friedman GD, Cutter GR, Donahue RP, Hughes GH, Hulley SB, Jacobs DR, et al. CARDIA: study design, recruitment, and some characteristics of the examined subjects. J Clin Epidemiol 1988;41(11):1105-1116. [Medline: 3204420]
35. 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]
36. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey Protocol. Hyattsville, MD: National Center for Health Statistics; 2007. URL: http://www.cdc.gov/nchs/nhanes/nhanes2007-2008/current_nhanes_07_08.htm [accessed 2014-04-11] [WebCite Cache ID 6OlK1ByeU]
37. Burke L, Elci O, Wang J, Ewing L, Conroy M, Acharya S, et al. Self-monitoring in behavioral weight loss treatment: SMART trial short-term results. Obesity 2009 Nov;17:S273.
38. 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]
39. 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]
40. Shapiro JR, Koro T, Doran N, Thompson S, Sallis JF, Calfas K, et al. Text4Diet: a randomized controlled study using text messaging for weight loss behaviors. Prev Med 2012 Nov;55(5):412-417. [doi: 10.1016/j.ypmed.2012.08.011] [Medline: 22944150]
41. Kumanyika SK, Gary TL, Lancaster KJ, Samuel-Hodge CD, Banks-Wallace J, Beech BM, et al. Achieving healthy weight in African-American communities: research perspectives and priorities. Obes Res 2005 Dec;13(12):2037-2047. [doi: 10.1038/oby.2005.251] [Medline: 16421334]
42. Osei-Assibey G, Kyrou I, Adi Y, Kumar S, Matyka K. Dietary and lifestyle interventions for weight management in adults from minority ethnic/non-White groups: a systematic review. Obes Rev 2010 Nov;11(11):769-776. [doi: 10.1111/j.1467-789X.2009.00695.x] [Medline: 20059708]
43. Smith A. Internet and American Life Project. Washington, DC: Pew Research Center; 2010 Jul 07. Mobile Access 2010: Pew Internet and the American Life Project URL: http://www.pewinternet.org/2010/07/07/mobile-access-2010/ [accessed 2014-04-11] [WebCite Cache ID 6OlK9WiWL]
44. 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

ANOVA: analysis of variance
BMI: body mass index
ICC: intraclass correlation coefficient
iOTA: interactive obesity treatment approach
IVR: interactive voice response
MOS-SSS: Medical Outcomes Study-Social Support Survey
PDA: personal digital assistant
PHS: Piedmont Health Services

Edited by H Riper; submitted 30.09.13; peer-reviewed by B Turner-McGrievy, S Tevelde; comments to author 23.12.13; revised version received 25.02.14; accepted 20.03.14; published 29.04.14.

Please cite as:
Steinberg DM, Levine EL, Lane I, Askew S, Foley PB, Puleo E, Bennett GG
Adherence to Self-Monitoring via Interactive Voice Response Technology in an eHealth Intervention Targeting Weight Gain Prevention Among Black Women: Randomized Controlled Trial
J Med Internet Res 2014;16(4):e114
URL: http://www.jmir.org/2014/4/e114/
doi:10.2196/jmir.2996
PMID:24780934

Original Paper

A Spanish Pillbox App for Elderly Patients Taking Multiple Medications: Randomized Controlled Trial

José Joaquín Mira[1,2,3], MSc, PhD; Isabel Navarro[2], BA; Federico Botella[4], PhD; Fernando Borrás[5], PhD; Roberto Nuño-Solinís[6], MSc; Domingo Orozco[7], MD, PhD; Fuencisla Iglesias-Alonso[8], MD; Pastora Pérez-Pérez[9], PhD; Susana Lorenzo[10], MD, MPH, PhD; Nuria Toro[6], BA

[1]Sant Joan-Alicante Health District, Consellería Sanidad, Alicante, Spain
[2]Health Psychology Department, Miguel Hernández University, Elche, Spain
[3]Research Network on Health Services in Chronic Diseases (REDISSEC), Madrid, Spain
[4]Center of Operations Research University Institute, Miguel Hernández University, Elche, Spain
[5]Statistics, Mathematics and Informatics Department, Miguel Hernández University, Elche, Spain
[6]Basque Institute for Healthcare Innovation (O+berri), Barakaldo, Spain
[7]Department of Family Medicine, Faculty of Medicine, Miguel Hernández University, Alicante, Spain
[8]Castilla La Mancha Health Service (SESCAM), Toledo, Spain
[9]Andalusian Patient Safety Observatory, Andalusian Agency for Health Care Quality, Seville, Spain
[10]Hospital Universitario Fundación Alcorcón, Madrid, Spain

Corresponding Author:
Roberto Nuño-Solinís, MSc
Basque Institute for Healthcare Innovation (O+berri)
Torre del BEC (Bilbao Exhibition Centre)
Ronda de Azkue 1
Barakaldo, 48902
Spain
Phone: 34 944536896
Fax: 34 944530465
Email: oberri@bioef.org

Abstract

Background: Nonadherence and medication errors are common among patients with complex drug regimens. Apps for smartphones and tablets are effective for improving adherence, but they have not been tested in elderly patients with complex chronic conditions and who typically have less experience with this type of technology.

Objective: The objective of this study was to design, implement, and evaluate a medication self-management app (called ALICE) for elderly patients taking multiple medications with the intention of improving adherence and safe medication use.

Methods: A single-blind randomized controlled trial was conducted with a control and an experimental group (N=99) in Spain in 2013. The characteristics of ALICE were specified based on the suggestions of 3 nominal groups with a total of 23 patients and a focus group with 7 professionals. ALICE was designed for Android and iOS to allow for the personalization of prescriptions and medical advice, showing images of each of the medications (the packaging and the medication itself) together with alerts and multiple reminders for each alert. The randomly assigned patients in the control group received oral and written information on the safe use of their medications and the patients in the experimental group used ALICE for 3 months. Pre and post measures included rate of missed doses and medication errors reported by patients, scores from the 4-item Morisky Medication Adherence Scale (MMAS-4), level of independence, self-perceived health status, and biochemical test results. In the experimental group, data were collected on their previous experience with information and communication technologies, their rating of ALICE, and their perception of the level of independence they had achieved. The intergroup intervention effects were calculated by univariate linear models and ANOVA, with the pre to post intervention differences as the dependent variables.

Results: Data were obtained from 99 patients (48 and 51 in the control and experimental groups, respectively). Patients in the experimental group obtained better MMAS-4 scores ($P<.001$) and reported fewer missed doses of medication ($P=.02$). ALICE only helped to significantly reduce medication errors in patients with an initially higher rate of errors ($P<.001$). Patients with no experience with information and communication technologies reported better adherence ($P<.001$), fewer missed doses ($P<.001$), and fewer medication errors (*P*=.02). The mean satisfaction score for ALICE was 8.5 out of 10. In all, 45 of 51 patients (88%) felt that ALICE improved their independence in managing their medications.

Conclusions: The ALICE app improves adherence, helps reduce rates of forgetting and of medication errors, and increases perceived independence in managing medication. Elderly patients with no previous experience with information and communication technologies are capable of effectively using an app designed to help them take their medicine more safely.

Trial Registration: Clinicaltrials.gov NCT02071498; http://clinicaltrials.gov/ct2/show/NCT02071498 (Archived by WebCite at http://www.webcitation.org/6OJjdHVhD).

(J Med Internet Res 2014;16(4):e99) doi:10.2196/jmir.3269

KEYWORDS

medication; patient nonadherence; mobile apps; patient safety; elderly

Introduction

Background

Chronic disorders are more common among elderly people. In Spain, persons older than 65 years have 3 chronic conditions on average [1]. As many as 94% take 5 or more drugs every day [1,2] and more than half are under the care of more than one doctor [3]. These rates are similar those found in other countries [4,5].

Safe Medication Use Among Elderly Individuals

The complexity of treatment regimens can lead to an accumulation of medications, confusion because medications look similar (especially in the case of generic medications), and a lack of coordination between the different levels of care. Patients' intrinsic factors (eg, cognitive impairment and false beliefs regarding some drugs) have been also cited among the most common causes of nonadherence [6] and of involuntary errors related to taking medication at home [3] in Spain. Similar patterns have been observed in other countries [7-13].

Nonadherence

Improving treatment adherence is a priority because nonadherence to long-term treatments has a negative impact on the health of patients and leads to unnecessary expenditures [8,14,15]. The causes of nonadherence are multiple [16], but it is known that as the complexity of medication regimens increases, there is more likely to be unintentional nonadherence, as observed in diabetes patients [17,18]. That is, nonadherence may be intentional when patients chose not to take the medicines as prescribed, but also unintentional when patients experience difficulties taking the medications [19]. Among other reasons, these difficulties are attributable to one-third of elderly patients failing to assimilate the information required to continue prescribed drug regimens correctly at home [20-23].

Patient Errors

Taking the wrong dose or the wrong medication accounts for 5% of hospital admissions [24]. In Spain, 35.8% of patients with diabetes admitted to having committed 4 or more medication errors in the previous year [12], and among those who had had diabetes for longer, a higher rate of patients reported drug interactions from taking together drugs that should not be mixed [25]. Sakar et al [10] found that as many as 59% of patients with type 2 diabetes made mistakes with the self-administration of medications and 21% with diet.

The Current Study

The participation of patients in their own care [26,27] has been shown to help improve use of drugs, for example, associating medications with specific meals, writing notes on the packaging, and using pillboxes [28,29]. Recently, with the growth in use of tablets and smartphones, various interventions [30-32] have been designed to improve adherence and it has been found that such tools are effective and help to increase patient independence [33].

Usage of apps for tablets and smartphones (eg, iOS, Android, Windows Phone, and Blackberry OS) designed to improve adherence has increased exponentially [34]. However, these apps have various limitations [35,36]: they are conceived for patients familiar with these technologies, they have not been written for elderly patients (assuming that they would not benefit from this type of tool), and in most cases, patients have not been consulted about the design [37]. Few are available in Spanish or in languages other than English, in general, reducing their use among those who are not English speakers. Furthermore, no studies have tested the effectiveness of these apps with the elderly.

The objective of this study was to design, implement, and evaluate a medication self-management app in Spanish for elderly patients taking multiple medications by using tablets (Android or iOS) to increase adherence to treatment regimens and achieve safer use of drugs.

Methods

ALICE Design

A tablet-based medication self-management app (called ALICE) was designed to help patients to remember to take all their medications at the correct doses, distinguish between drugs to avoid confusions, avoid known potential interactions and common errors in use of the medications, and know how to properly store the medications. ALICE was also designed to remember doctors' recommendations for healthy habits, such as physical exercise and diet. The app design was based on suggestions extracted from 3 nominal groups with a total of 23 patients and a focus group with 7 professionals (3 physicians and 4 pharmacists). The guide with questions for patients and professionals was drawn up considering the classification of medication errors in the study by Field et al [13] at the Meyers Primary Care Institute in Massachusetts and previous research by our group on patients' perception of safety and the frequency and characteristics of patients' medication errors [3].

Tablets Used

The tablet used was selected on the basis of the need for a device with an at least 7-inch easy-to-use touch screen, ensuring that users would only have to follow simple instructions and tap on some icons on the screen. Specifically, the BQ Verne Plus 3G with an liquid-crystal display (LCD) tactile screen was chosen for an Android and an iPad 2 with Wi-Fi and 3G for the iOS.

ALICE Characteristics

The ALICE app was designed to work with personalized prescriptions and recommendations given to patients, with a function making it possible to store details of all their prescriptions and related instructions, as well as images for each of the medications (even allowing pictures to be taken of the packaging or the appearance of the medication itself), and recommendations of the various different doctors seeing the same patient. A second function established a customized system of alerts and reminders for a given alert, to remind patients when to take their medications and to put into practice healthy habits, using the approaches suggested by participants in the nominal groups (eg, the association of the intake of medicines with particular meals or daily activities, and various techniques for splitting pills). Lastly, a third function was to enable monitoring of the level of adherence to the prescriptions and medical advice, the tablet connecting via a wireless or 3G network with the study monitoring system, with the health care provider and with a relative or caregiver when authorized by the patient (Figure 1).

When it is time to take a medication, an alarm sounds and the patient accesses the main menu of the app. Displayed is the name of the medication, dose, time, and any warnings of how the patient should take the medication (if applicable). The patient may consult his/her medication anytime. The medication usually is associated with breakfast, lunch, snack, or dinner daily. ALICE follows this normal use. The ALICE app reports medications the patient must take in a day and reports medicines that the patient has forgotten to take that day.

Once the tool was designed and before the experimental phase, we assessed its feasibility, verifying that all the app characteristics proposed in the previous phase were included in the design process. For this purpose, 8 elderly patients assessed the user-friendliness of the tool, its degree of intuitiveness, whether the font was sufficiently large, the contrast of the text and images, and the quality of the photographs. Identified problems were the low volume of alerts and a proposal to simplify the system for taking photos of drugs to include images on ALICE. Both problems were solved.

The ALICE app has the following features: a user-friendly interface to introduce text and images; various medication reminder alerts and messages for patients, including sounds or flashing and also messages sent to relatives or caregivers; and a complete list for the caregivers of all prescriptions regardless of the number of doctors involved including a summary of patient adherence behavior.

Figure 1. Example of a ALICE screen.

MEDICAMENTO: Omeprazol 20 mg

DOSIS: 1

HORA TOMA: 9:00

Debe administrarse preferentemente por las mañanas

Study Design

To evaluate ALICE we opted for a single-blind experimental design with 2 groups (control and experimental) and pre and post assessments using a randomized controlled trial (NCT02071498). Patients were randomly assigned to the control or experimental group. The control group was composed of participants who did not use ALICE, and the experimental group was composed of people who used this tool for 3 months. To maintain the blinding and be able to link the pre and post measurements, patients were assigned codes based on their date of birth and initials.

Patients Enrolled

We randomly selected 102 patients with a digital medical history from 13 health centers in the health districts of Alicante and Bilbao. We defined the following inclusion criteria for the study: multimorbid patients taking multiple medications, older than 65 years, with a Barthel [38] score of more than 60, living in their own home, and able to manage the administration of their medication at home. The sample size was calculated to detect a difference between means of at least 10 points with a statistical power of 90% at a level of significance of alpha=.05 (in a 2-tailed test). We requested the informed consent of patients from both control and experimental groups. This study was approved and financed by the Spanish Ministry of Health, Equality, and Social Policy. The Spanish Research Health Agency (FIS), Independent Clinical Research, project number EC11-527. The Experimental Research Ethics Committee of Miguel Hernández University (DPS-JJM-003-11) approved the trial.

Measures

All participating patients completed a questionnaire to assess the rates of missed doses and of medication errors, and adherence to treatment measured by the 4-item *Morisky Medication Adherence Scale (MMAS-4)*, as well as self-perceived health status, the number of doctors seen, whether they used a physical pillbox, and who organized their medication (over the previous 3 months). Additionally, we recorded data on the sociodemographic characteristics of participants, namely sex, age, and civil status. Those in the control group received oral and written information regarding the main risks related to their medications and the most common errors of patients when taking medications. Participants in the experimental group were given a BQ tablet or an iPad with the ALICE app installed and personalized according to the medications they had been prescribed as listed in their medical record. Patients in this group attended individual sessions of up to 2 hours to be shown how to use the app. During the study period, patients from the experimental group had a contact telephone number for any query regarding the use of the tablet or the app.

Three months later, the measurements made preintervention were repeated and extra information was collected regarding treatment adherence based on the data provided by ALICE (post). Additionally, patients from the experimental group answered a series of questions to ascertain whether they, prior to the study, had experience with using tablets, smartphones, mobile phones, and computers. They were also asked to evaluate ALICE (its performance, functionality, usability, reliability, acceptability, usefulness, design, simplicity, accessibility, and problem-solving power, as well as overall satisfaction with the tool), whether they would recommend the app to relatives, friends, and acquaintances, and their perception of the degree of independence they had achieved because of ALICE.

Statistical Analysis

To assess the effectiveness of ALICE, we built various univariate linear models. Where there were intergroup differences in preintervention measurements, we performed univariate linear model and ANOVA tests, using as the dependent variables the differences between the pre and post intervention measurements in MMAS-4 scores, self-perceived health status, levels of glycated hemoglobin (HbA1), low-density lipoprotein (LDL) cholesterol, blood pressure, number of medication errors, and missed doses related to medication reported by patients. Allocation to the experimental or control group was considered as the independent variable. In the experimental group, we carried out Pearson correlation analysis between MMAS-4 scores and treatment adherence assessed objectively by the number of alerts that had not been acknowledged or had been acknowledged late (data provided by ALICE). Additionally, to allow for the potential effect of the level of experience of patients with information and communication technology (ICT), we compared the pre and post intervention measurements in the experimental group adjusting for the effect of familiarity with ICTs. We compared the rating of ALICE by patients in the experimental group with and without ICT experience by using the chi-square test and their level of satisfaction with ALICE (on a scale of 0 to 10) by using the Student *t* test. In all cases, we checked that the assumptions of the statistical tests used were met. We considered $P<.05$ to be statistically significant.

Results

Qualitative Study: ALICE Characteristics

According to the 23 patients and 7 health professionals that participated in the qualitative research process, the design of the ALICE app had to satisfy the following criteria: it needed be easy to use, have sufficiently large font size, include photographs of the medications, and offer a large variety of alerts to be set to suit the preferences of patients with a menu enabling patients to easily adjust the alerts to be associated with their breakfast, lunch, dinner, and other meals and snacks, or other daily activities occurring at regular times. For each alert, it had to be possible to program up to 5 reminders and include information to avoid confusion between medicines, and advice on how to take the medication and how to store it properly. It was important to include all the medications patients were on regardless of the prescribing doctor and include automatic reminders of other medical recommendations (eg, to do physical exercise, including what type and for how long). Further, ALICE had to record the level of treatment adherence, send short message service (SMS) text messages to relatives or caregivers if patients failed to adhere to treatments, and alert a monitoring system in the event of a malfunction. Lastly, ALICE had to provide information about all the medications taken by patients, regardless of the number of prescribing doctors, and send reports on medication alerts that were not acknowledged so that when patients attended appointments, their doctors would know the extent to which they were adhering to treatments prescribed.

Experimental Study: Participant Characteristics

This study was carried out between June 2012 and May 2013. Data were obtained from 99 patients (48 and 51 in the control and experimental groups, respectively). Three patients in the control group declined to participate (Figure 2). The characteristics of those who participated are summarized in Table 1. A total of 72 of 99 patients (73%) took more than 5 types of drugs per day and 36 of 99 (36%) were under the care of more than one doctor. In the experimental group, 22 of 51 (43%) had a computer and 19 of 51 (37%) had an Internet connection at home, whereas 39 of 51 (76%) had a mobile phone and 9 of 51 (18%) a smartphone. On the other hand, 28 of 51 individuals (55%) were not familiar with ICTs, having never used a computer, tablet, or smartphone.

Figure 2. Study flow diagram.

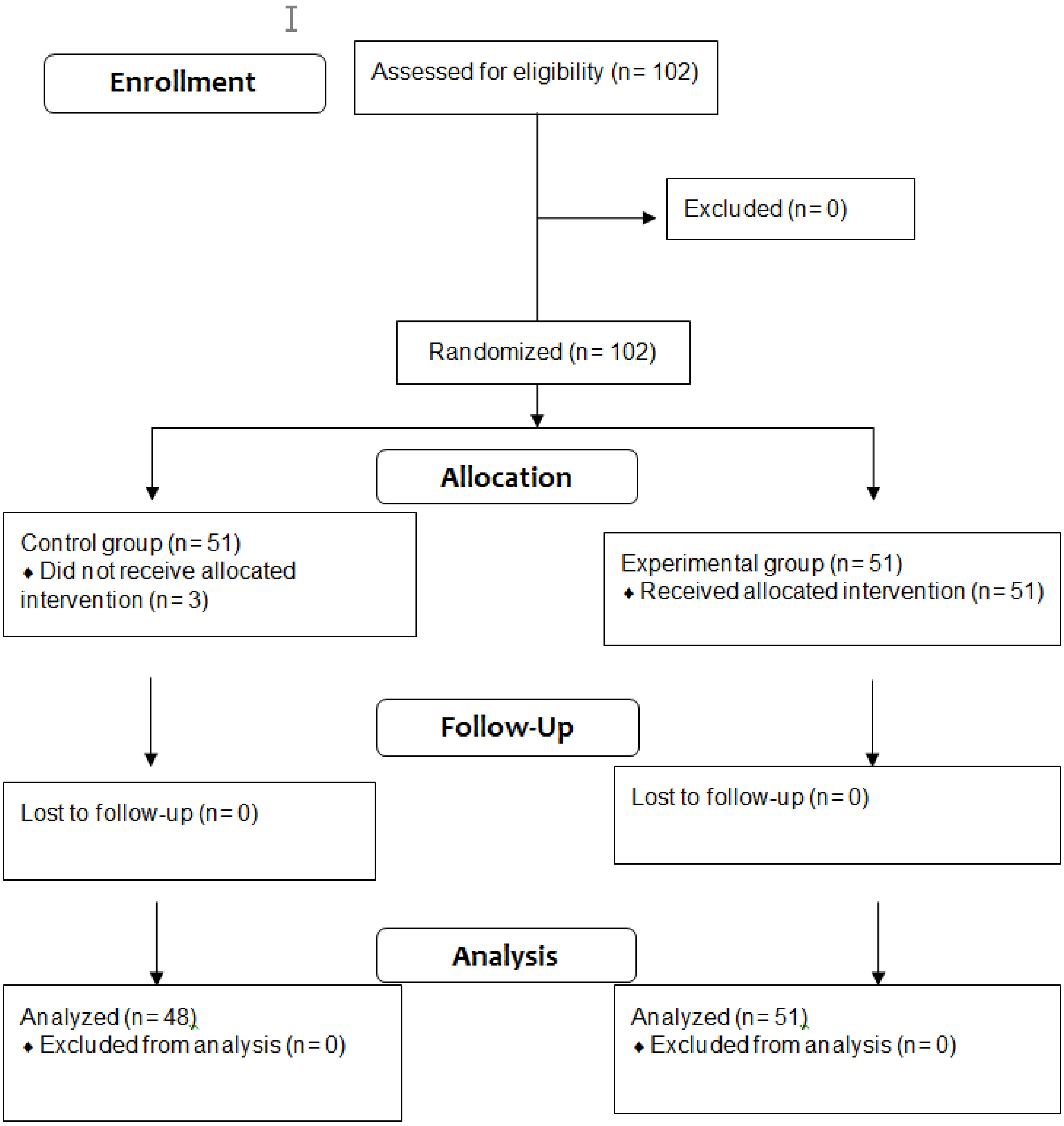


Table 1. Description of patients in the experimental and control groups (N=99).

Demographic characteristics	Control n=48	Experimental n=51	*P* [a]
Age (years), mean (SD)	72.9 (6)	70.9 (8)	.16
Sex, n (%)			
Women	23 (48)	21 (41)	.50
Men	25 (52)	30 (59)	
Civil status, n (%)			
Single	1 (2)	1 (2)	.10
Married	39 (81)	32 (63)	
Widowed	8 (17)	14 (27)	
Divorced	0 (0)	4 (8)	
Living arrangements, n (%)			
Alone	9 (19)	12 (23)	.56
With partner/relative	39 (81)	39 (76)	
Who organized their medication, n (%)			
Patient himself/herself	41 (85)	47 (92)	.29
Partner/relative/caregiver	7 (15)	4 (8)	
Self-perceived health status, n (%)			
Poor	5 (10)	2 (4)	.64
Fair	14 (29)	17 (33)	
Good	26 (54)	28 (55)	
Excellent	3 (6)	4 (8)	
Disorders [b]			
Diabetes	46 (96)	43 (84)	.06
Insulin-dependent patients[c]	9 (20)	15 (36)	.09
Depression/anxiety	5 (10)	4 (8)	.66
Hypercholesterolemia	24 (50)	28 (55)	.62
Benign prostatic hyperplasia[d]	3 (75)	7 (14)	.22
High blood pressure	40 (83)	38 (74)	.28
Other cardiovascular diseases	22 (46)	21 (41)	.64
Arthrosis	11 (23)	9 (18)	.51
Renal failure	8 (89)	6 (12)	.48
Chronic obstructive pulmonary disease	10 (21)	9 (18)	.69
Digestive disorders	11 (23)	4 (8)	.04
Number of drugs prescribed, mean (SD)	7.9 (3)	7.6 (3)	.55
Number of doctors involved, mean (SD)	2.3 (1)	2.3 (2)	.99

[a] Based on the Pearson chi-square test or the Student *t* test for independent samples.

[b] Patients could have more than 1 disorder.

[c] Percentage calculated over the total number of patients with diabetes.

[d] Percentage calculated over the total of men in the sample.

Effects of ALICE

Patients from the experimental group reported greater treatment adherence (measured using the MMAS-4) and a lower rate of missed doses at the end of the study (Table 2). Specifically, treatment adherence (measured by the scale scores) increased by 28.3% and the rate of missed doses fell by 27.3%. On the other hand, the univariate linear analysis of pre-post differences in the control and experimental groups indicated that ALICE was not effective in reducing the rate of medication errors. For this variable, we conducted ANOVA to control for the effect of the difference in the rate of medication errors between the experimental and control groups before the intervention. This analysis indicated that ALICE did help to reduce the medication errors, but only in patients who had recognized that they had made 2 or more errors before the study (Table 2). Among these patients, the rate of medication errors fell by 41.2% compared to before the study.

The patients in the control and experimental groups had similar levels of glycated hemoglobin and blood pressure and similar self-perceived health status before and after the study. However, the pre-post measurements of cholesterol were statistically different; levels increased by 5% compared to before the study (Table 2).

A total of 31,383 medication alerts were programmed. Between Monday and Sunday, 3293 (10.49%) alerts were not been dealt with and 2978 (9.49%) were dealt with after the first reminder. The Pearson correlation coefficient between the number of alerts that were not dealt with and the MMAS-4 score was .3 (*P*=.04). This result showed the positive and significant relationship between the adherence of the MMAS-4 and the information provided by ALICE on the level of patient compliance.

Table 2. The 4-item Morisky Medication Adherence Scale (MMAS-4) scores and rates of missed doses and medication errors reported by patients in the control and experimental groups.

Measures	Control group (n=48)		Experimental group (n=51)		Pre-post difference		Effect size Δ	Between-group difference, *P*	
	Pre	Post	Pre	Post	Control	Experimental		ULM[a]	ANOVA
MMAS-4, mean (SD)	7.2 (0.9)	7.3 (0.7)	6.6 (1.2)	7.4 (0.9)	0.1	0.8	0.7	<.001	—
Self-perceived health status, n (%)	68.3 (21)	69.1 (20)	71.27 (17)	74.6 (17)	0.9	3.3	1.2	.54	—
Glycated hemoglobin (mmol/mol), mean (SD)	7.1 (1.1)	6.7 (1.4)	7.1 (1.4)	7.4 (2.7)	–0.4	0.3	0.4	.36	—
Cholesterol (mg/dL), mean (SD)	105.4 (29.9)	101.9 (28.1)	107.0 (29.8)	112.7 (45.8)	–3.5	5.7	12.2	.04	—
Blood pressure (mm Hg), mean (SD)									
Diastolic	75.8 (10.5)	76.6 (11.2)	72.3 (10)	70.6 (8.8)	0.8	1.7	0.2	.89	—
Systolic	137.3 (12.4)	140.5 (14.6)	130.9 (15)	128.6 (20.9)	3.2	2.3	2.6	.28	—
Medication errors, [b] **n (%)**					0.0	–0.1	0.2	.21	—
0	42 (87)	43 (90)	38 (74)	43 (84)					
1	6 (12)	3 (6)	9 (18)	6 (12)					.95[c]
2	0 (0)	2 (4)	4 (8)	2 (4)					<.001
Number of missed doses, n (%)					0.2	–0.3	0.5	.02	—
0	28 (58)	20 (42)	18 (35)	27 (53)					
1	12 (25)	17 (35)	21 (41)	16 (31)					
2	6 (12)	9 (19)	8 (16)	5 (10)					
≥3	0 (0)	2 (4)	4 (8)	3 (6)					

[a] ULM: univariate linear model.

[b] Wrong drug taken (attributed to confusion between drugs that appear similar) or incorrect doses.

[c] For the subgroup of 0 or 1 medication errors reported at the first assessment.

Experience With Information and Communication Technology

Patients from the experimental group who were not familiar with ICTs obtained higher MMAS-4 scores and reported fewer medication errors and missed doses after the intervention (Table 3). The pattern was similar in patients from the experimental group who had experience with ICTs, except that there was not a significant reduction in the number of missed doses.

Table 3. Influence of ALICE and patients' previous experience with information and communication technology (ICT) in the experimental group (n=51).

Measures	Some ICT experience (n=23)			No ICT experience (n=28)		
	Pre	Post	P [a]	Pre	Post	P [a]
Morisky Medication Adherence Scale, mean (SD)	6.6 (1.4)	7.4 (0.8)	.01	6.6 (1.1)	7.2 (1.0)	<.001
Medication errors, [b] **n (%)**						
0	18 (78)	19 (83)		20 (71)	24 (86)	
1	3 (13)	3 (13)		6 (21)	3 (11)	
2	2 (9)	1 (4)	.01	2 (7)	1 (4)	.02
Missed doses reported by patients, n (%)						
0	9 (39)	14 (61)		9 (32)	13 (46)	
1	9 (39)	6 (26)		12 (43)	10 (36)	
2	4 (17)	2 (9)	.14	4 (14)	3 (11)	<.001
≥3	1 (4)	1 (4)		3 (11)	3 (7)	

[a] Based on the Wilcoxon test for paired samples (MMAS-4) or differences based on phi and Cramer's V (medication errors and missed doses).

[b] Wrong drug taken (attributed to confusion between drugs that appear similar) or incorrect errors.

ALICE Functioning

More than half of the patients from the experimental group (30/51, 59%) required individual support once they joined the study to solve problems related to the use of ALICE. Most of these (9/29, 31%) concerned charging the battery and restarting the system (Table 4).

Table 4. Patient assessment of the ALICE functioning.

Patient report	n (%)
Has the tablet worked?	
Yes, it has consistently worked well from the beginning	21 (41)
Yes, though I have had some problems	29 (57)
No, it has consistently failed to work	1 (2)
In the event that it has failed to work, what was the problem?	
It kept turning off	6 (21)
The alert didn't trigger	10 (34)
The sound was poor	1 (3)
The battery ran down very quickly	3 (10)
It froze and/or crashed	5 (17)
Other	4 (14)
Was the problem solved quickly and effectively?	
No, the problem was not solved	1 (3)
Yes, the problem was always solved	28 (96)

Participant Satisfaction and Medication Self-Management

The level of satisfaction with ALICE (in the experimental group) was similar in those with some and with no previous ICT experience (Table 5). The mean scores were more than 8 out of a maximum of 10. More than half of the patients in the experimental group (30/51, 59%) reported that the ALICE app improved their medication use. Another 15 (29.4%) considered that ALICE helped to a certain extent, whereas 6 patients (11.8%) indicated that ALICE did not help at all (Table 5).

Table 5. Patient satisfaction with ALICE app.

ALICE satisfaction	Some ICT experience n=23	No ICT experience n=28	Total N=51	P [a]
Patient report, n (%)				
I like the design of the messages and alerts in ALICE	23 (100)	28 (100)	51 (100)	—
The photos of the pills/capsule help me take the correct drug	23 (100)	28 (100)	51 (100)	—
The font size is sufficiently large	23 (100)	27 (96)	50 (98)	.36
The photos of the medication packaging help me take my medication correctly	23 (100)	27 (96)	50 (98)	.36
The instructions for using ALICE have been clear, correct and complete	22 (96)	27 (100)	49 (98)	.27
ALICE is easy to use and manage	21 (91)	27 (96)	48 (94)	.53
It is sufficiently large to see the screen well	18 (78)	27 (96)	45 (88)	.13
It is easy to tap on the correct icon with my finger	21 (91)	24 (86)	45 (88)	.28
In general, it is easy to operate with my finger	20 (87)	24 (86)	44 (86)	.46
The audio alerts are loud enough	20 (87)	22 (79)	42 (82)	.27
Overall satisfaction with ALICE, mean (SD)	8.3 (1.4)	8.7 (1.4)	8.5 (1.4)	.26

[a] Based on the chi-square test *t* test for unpaired samples (overall satisfaction).

Discussion

Principal Results

Treatment adherence was higher in patients in the experimental group than in the control group. It is important to highlight that the ALICE app helped to solve a common problem in elderly patients; namely, remembering whether they have taken their medication. By using ALICE, the number of errors decreased, although the patients who benefited in this respect were those who previously made the most errors. Our data suggest that ALICE does contribute to reducing systematic errors, but not all medication errors.

The self-perceived health status, levels of glycated hemoglobin, and blood pressure remained similar from the beginning to the end of the study, and although cholesterol levels increased, the changes were not clinically relevant. That is, ALICE did not improve the clinical status of patients; however, 3 months may not be long enough to observe differences. It did, however, improve patients’ perception of independence in the management of their medications.

Elderly patients with complex drug regimens, even those without previous experience in the use of tablets, smartphones, and computers, or who had never used the Internet, were found to be capable of effectively using an app designed to improve safe medication use. ALICE was designed to cope with attrition of users using these devices. ALICE was used daily and only 1 patient left the study because of a problem with the tablet and not with ALICE.

Comparison With Previous Studies

This is one of the few experimental studies assessing the effectiveness of a virtual pillbox for tablets and smartphones targeted to elderly patients with multiple health problems. The design of ALICE took into account the findings of other studies, but was based on the views and preferences of the target patients and the experience of the primary care and pharmacy personnel caring for them. The suggestion to assess the usability during the development of this intervention was also applied [39]. This approach to designing an app has not been widely implemented, but it has been proposed in other studies focused on developing such tools. Indeed, this development process gave rise to one of its strengths, the fact that it can be adapted to the habits, timetables, and lifestyles of each patient as suggested by Tatara et al [37]. In adherence studies, the personalization of tools has not been widely considered and our results underline the importance of the participation of patients and professionals in the design of applications.

Traditional pillboxes improve treatment adherence, but are only useful in the case of solid dosage medications [40]. Further, although they help with the organization of medications at home, elderly patients may find it difficult to put their pills into the compartments [41]. One study found that approximately 22% of patients were unable to effectively use physical pillboxes [42]. It might be possible to overcome these limitations by using a virtual pillbox for tablets and smartphones.

The level of adherence, assessed in 3 different ways (direct reporting, MMAS-4 scores, and statistics from ALICE on unacknowledged alerts), was good in all cases compared to the mean nonadherence figures reported for chronic patients in general [19] (approximately 50%), being slightly higher in patients with diabetes, which was one of the most common diseases in our study population, and was slightly lower than in another study of a pillbox for smartphones that obtained adherence of 36%. Similar to other studies, ALICE showed capacity to modify the behavior of adult patients increasing adherence to the therapeutic regimen and reducing medication errors in some cases [43].

Relevance of This Study

ALICE-like apps make it possible to provide information that contributes to safer use of drugs by tuning in to the needs of specific patients, for example, by including advice and preventive measures that the doctors involved consider the most appropriate (given the medication dose or timings). Further, they can be adapted to the habits and lifestyle of patients, to make it easier and simpler for them to take their medication because it is known that these factors are the main cause of nonadherence. These issues have been taken into account in the design of ALICE and it seems this app does have a beneficial effect on the level of adherence and patients perceive that it improves the way they take their medication.

Rates of medication errors are particularly high in patients on multiple medications [11,12,27]. One of the objectives when designing ALICE was to reduce medication errors among elderly patients with multimorbidity by providing a tool that they could use themselves. This objective is particularly important from a clinical point of view in the target population. Field et al [13] found that most medication errors made by patients concerned oral hypoglycemic agents (29%) and more than one-third (36%) of our patients were on this type of drug. However, this objective was only partly achieved.

Limitations

The small number of participants and the number of months using ALICE affected our ability to detect differences between the group using the ALICE app and the control group (eg, in relation to biomarkers) as well as our ability to generalize the results. This type of app does not address the issue of where to store medication at home to have it accessible when it should be taken. The data collected do not include details about the number of drugs to be taken daily. No count was made of pills nor did we use any other system to check if patients actually took any medication after acknowledging the alert. Further, we cannot be sure that they took the correct dose or even the correct drug. We do not know whether ALICE would continue to be effective in the longer term. This is important because a notable reason patients cite for stopping taking medications is the feeling that they are not working for them. There is some evidence that the *MMAS-4* overestimates adherence, yielding higher rates than those obtained from pill counts. This study did not consider medication reconciliation. In some cases, if there has been inappropriate prescribing, following treatment regimens could pose a risk. This app only addresses unintentional nonadherence. In the event of a patient being reluctant or opposed to taking a certain medicine [44] (this known to affect approximately 43% of cases), this type of tool might be useful for its educational features, but exploring this issue is beyond the scope of this study.

Recommendations for Practice and Research

Most apps have been designed for patients with less complex health problems and/or experience with ICTs. This study should change the expectations of developers and mobile phone companies, encouraging them to develop apps and devices suited to older patients with multimorbidity who are normally excluded from studies thought to be too complex because such tools could improve the capacity of these individuals to manage their illnesses.

Glucose monitors and other devices currently in use in telemedicine programs could have add-ons to help individual patients use their medications more safely. Such devices should not only include alerts for medicines, but also reminders of how to put advice on healthy habits into practice.

Further studies on virtual pillboxes for tablets and smartphones could explore whether adherence can be improved by personalization of treatment regimens as suggested by other authors [45]. Specifically, future research should assess to what extent these tools are useful for older individuals living alone, a situation that is expected to be the reality for a growing number of patients in the near future.

ALICE and other similar apps have a broad potential not only for patients, but also for professionals because they can provide useful information about how patients adapt the therapeutic regimen to their lifestyle. Future studies could explore how to improve physician recommendations to increase adherence to treatment based on personalized information.

Acknowledgments

Alicia Peralta assisted in many phases of this study. This study was financed by the Spanish Ministry of Health, Equality, and Social Policy. The Spanish Research Health Agency (FIS), project number EC11-527.

Authors' Contributions

JJM, RN and FB conceived of the study. JJM, RN, FB, and FBt participated in the design of the study. IN performed the statistical analysis. DO, VG, PP, FI, SL, and NT coordinated the qualitative research and the trial, captured and prepared data to design and improve ALICE. FB and FBt designed the app. JJM, IN, and RN prepared a first version of this original. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V1.6.2 [46].

[PDF File (Adobe PDF File), 996KB - jmir_v16i4e99_app1.pdf]

References

1. Osakidetza. ESCAV’07. Bilbao; 2007. URL: http://www.osakidetza.euskadi.net/r85-gkgnrl00/es/contenidos/informacion/encuesta_salud/es_escav/r01hRedirectCont/contenidos/informacion/encuesta_salud_comarcas/es_escav/comarcas.html [accessed 2014-03-28] [WebCite Cache ID 6OPeVtAia]
2. Ministerio de Sanidad y Política Social. Unidad de Pacientes Pluripatológicos. Estándares y recomendaciones. Madrid: Ministerio de Sanidad y Política Social; 2009. URL: http://www.msssi.gob.es/organizacion/sns/planCalidadSNS/docs/EyR_UPP.pdf [accessed 2014-03-26] [WebCite Cache ID 6OMl9rQY1]
3. Mira JJ, Orozco-Beltrán D, Pérez-Jover V, Martínez-Jimeno L, Gil-Guillén VF, Carratala-Munuera C, et al. Physician patient communication failure facilitates medication errors in older polymedicated patients with multiple comorbidities. Fam Pract 2013 Feb;30(1):56-63 [FREE Full text] [doi: 10.1093/fampra/cms046] [Medline: 22904014]
4. Bodenheimer T, Berry-Millett R. Follow the money--controlling expenditures by improving care for patients needing costly services. N Engl J Med 2009 Oct 15;361(16):1521-1523. [doi: 10.1056/NEJMp0907185] [Medline: 19797277]
5. Koper D, Kamenski G, Flamm M, Böhmdorfer B, Sönnichsen A. Frequency of medication errors in primary care patients with polypharmacy. Fam Pract 2013 Jun;30(3):313-319. [doi: 10.1093/fampra/cms070] [Medline: 23132894]
6. Crespillo-García E, Rivas-Ruiz F, Contreras E, Castellano P, Suárez G, Pérez-Tobar E. Conocimientos, percepciones y actitudes que intervienen en la adherencia al tratamiento en pacientes ancianos polimedicados desde una perspectiva cualitativa. Rev Calid Asist 2013;28:56-62. [doi: 10.1016/j.cali.2012.09.009]
7. Vermeire E, Hearnshaw H, Van Royen P, Denekens J. Patient adherence to treatment: three decades of research. A comprehensive review. J Clin Pharm Ther 2001 Oct;26(5):331-342. [Medline: 11679023]
8. World Health Organization. Adherence to Long-Term Therapies: Evidence for Action. Geneva: WHO; 2003.
9. Ose D, Mahler C, Vogel I, Ludt S, Szecsenyi J, Freund T. Let's talk about medication: concordance in rating medication adherence among multimorbid patients and their general practitioners. Patient Prefer Adherence 2012;6:839-845 [FREE Full text] [doi: 10.2147/PPA.S35498] [Medline: 23226007]
10. Sarkar U, Handley MA, Gupta R, Tang A, Murphy E, Seligman HK, et al. What happens between visits? Adverse and potential adverse events among a low-income, urban, ambulatory population with diabetes. Qual Saf Health Care 2010 Jun;19(3):223-228. [doi: 10.1136/qshc.2008.029116] [Medline: 20378619]
11. al Mahdy H, Seymour DG. How much can elderly patients tell us about their medications? Postgrad Med J 1990 Feb;66(772):116-121 [FREE Full text] [Medline: 2349180]
12. Mira JJ, Ortiz L, Lorenzo S, Royuela C, Vitaller J, Pérez-Jover V. Oversights, Confusions and Misinterpretations Related to Self-Care and Medication in Diabetic and Renal Patients. Med Princ Pract 2014 Feb 7. [doi: 10.1159/000358225] [Medline: 24513645]
13. Field TS, Mazor KM, Briesacher B, Debellis KR, Gurwitz JH. Adverse drug events resulting from patient errors in older adults. J Am Geriatr Soc 2007 Feb;55(2):271-276. [doi: 10.1111/j.1532-5415.2007.01047.x] [Medline: 17302666]
14. Cleemput I, Kesteloot K, DeGeest S. A review of the literature on the economics of noncompliance. Room for methodological improvement. Health Policy 2002 Jan;59(1):65-94. [Medline: 11786175]
15. Haynes RB, Ackloo E, Sahota N, McDonald HP, Yao X. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2008(2):CD000011. [doi: 10.1002/14651858.CD000011.pub3] [Medline: 18425859]
16. Pyatak EA, Florindez D, Weigensberg MJ. Adherence decision making in the everyday lives of emerging adults with type 1 diabetes. Patient Prefer Adherence 2013;7:709-718 [FREE Full text] [doi: 10.2147/PPA.S47577] [Medline: 23935361]
17. Grissinger M, Lease M. Misadventures in insulin therapy: are you at risk? J Natl Med Assoc 2003 Feb;95(2 Suppl 1):1S-16S. [Medline: 12653373]
18. DiMatteo MR. Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. Med Care 2004 Mar;42(3):200-209. [Medline: 15076819]
19. Clyne W, White S, McLachlan S. Developing consensus-based policy solutions for medicines adherence for Europe: a Delphi study. BMC Health Serv Res 2012;12:425 [FREE Full text] [doi: 10.1186/1472-6963-12-425] [Medline: 23176439]
20. Metlay JP, Cohen A, Polsky D, Kimmel SE, Koppel R, Hennessy S. Medication safety in older adults: home-based practice patterns. J Am Geriatr Soc 2005 Jun;53(6):976-982. [doi: 10.1111/j.1532-5415.2005.53308.x] [Medline: 15935020]
21. Leal M, Abellán J, Casa MT, Martínez J. Paciente polimedicado: ¿conoce la posología de la medicación?, ¿afirma tomarla correctamente? Aten Primaria 2004;33:451-456. [doi: 10.1157/13061580]
22. Britten N, Stevenson FA, Barry CA, Barber N, Bradley CP. Misunderstandings in prescribing decisions in general practice: qualitative study. BMJ 2000 Feb 19;320(7233):484-488 [FREE Full text] [Medline: 10678863]
23. Mira JJ, Martínez-Jimeno L, Orozco D, Iglesias F, Lorenzo S, Nuño R, et al. What older complex chronic patients need know about the medication they take everyday for a safe drug use. Expert Opinion On Drug Safety 2014 under review(forthcoming).

24. Einarson TR. Drug-related hospital admissions. Ann Pharmacother 1993;27(7-8):832-840. [Medline: 8364259]
25. Mira JJ, Lorenzo S, Vitaller J, Ignacio E, Aranaz J. ¿Qué hacen y qué deben hacer los pacientes diabéticos para evitar errores con el tratamiento? Endocrinol Nutr 2012;59:416-422. [doi: 10.1016/j.endonu.2012.05.007]
26. Tricco AC, Ivers NM, Grimshaw JM, Moher D, Turner L, Galipeau J, et al. Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. Lancet 2012 Jun 16;379(9833):2252-2261. [doi: 10.1016/S0140-6736(12)60480-2] [Medline: 22683130]
27. Schwappach DL, Frank O, Buschmann U, Babst R. Effects of an educational patient safety campaign on patients' safety behaviours and adverse events. J Eval Clin Pract 2013 Apr;19(2):285-291. [doi: 10.1111/j.1365-2753.2012.01820.x] [Medline: 22332730]
28. Branin JJ. The role of memory strategies in medication adherence among the elderly. Home Health Care Serv Q 2001;20(2):1-16. [doi: 10.1300/J027v20n02_01] [Medline: 11987652]
29. Raehl CL, Bond CA, Woods T, Patry RA, Sleeper RB. Individualized drug use assessment in the elderly. Pharmacotherapy 2002 Oct;22(10):1239-1248. [Medline: 12389875]
30. Burda C, Haack M, Duarte AC, Alemi F. Medication adherence among homeless patients: a pilot study of cell phone effectiveness. J Am Acad Nurse Pract 2012 Nov;24(11):675-681. [doi: 10.1111/j.1745-7599.2012.00756.x] [Medline: 23088699]
31. Morak J, Schwarz M, Hayn D, Schreier G. Feasibility of mHealth and Near Field Communication technology based medication adherence monitoring. Conf Proc IEEE Eng Med Biol Soc 2012;2012:272-275. [doi: 10.1109/EMBC.2012.6345922] [Medline: 23365883]
32. Hayakawa M, Uchimura Y, Omae K, Waki K, Fujita H, Ohe K. A smartphone-based medication self-management system with realtime medication monitoring. Appl Clin Inform 2013;4(1):37-52 [FREE Full text] [doi: 10.4338/ACI-2012-10-RA-0045] [Medline: 23650486]
33. Årsand E, Frøisland DH, Skrøvseth SO, Chomutare T, Tatara N, Hartvigsen G, et al. Mobile health applications to assist patients with diabetes: lessons learned and design implications. J Diabetes Sci Technol 2012 Sep;6(5):1197-1206 [FREE Full text] [Medline: 23063047]
34. Dayer L, Heldenbrand S, Anderson P, Gubbins PO, Martin BC. Smartphone medication adherence apps: potential benefits to patients and providers. J Am Pharm Assoc (2003) 2013;53(2):172-181 [FREE Full text] [doi: 10.1331/JAPhA.2013.12202] [Medline: 23571625]
35. El-Gayar O, Timsina P, Nawar N, Eid W. Mobile applications for diabetes self-management: status and potential. J Diabetes Sci Technol 2013;7(1):247-262 [FREE Full text] [Medline: 23439183]
36. McCall C, Maynes B, Zou CC, Zhang NJ. An automatic medication self-management and monitoring system for independently living patients. Med Eng Phys 2013 Apr;35(4):505-514. [doi: 10.1016/j.medengphy.2012.06.018] [Medline: 22824724]
37. Tatara N, Årsand E, Skrøvseth S, Hartvigsen G. Long-term engagement with a mobile self-management system for people with type 2 diabetes. JMIR Mhealth Uhealth 2013;1(1):e1. [doi: 10.2196/mhealth.2432]
38. Wade DT, Collin C. The Barthel ADL Index: a standard measure of physical disability? Int Disabil Stud 1988;10(2):64-67. [Medline: 3042746]
39. Voncken-Brewster V, Moser A, van der Weijden T, Nagykaldi Z, de Vries H, Tange H. Usability evaluation of an online, tailored self-management intervention for chronic obstructive pulmonary disease patients incorporating behavior change techniques. JMIR Res Protoc 2013 Jan;2(1):e3 [FREE Full text] [doi: 10.2196/resprot.2246] [Medline: 23612363]
40. Morales Suárez-Varela MT, GEMECOR. [Study on the use of a smart pillbox to improve treatment compliance]. Aten Primaria 2009 Apr;41(4):185-191. [doi: 10.1016/j.aprim.2008.07.003] [Medline: 19328598]
41. Anderson K, Jue SG, Madaras-Kelly KJ. Identifying patients at risk for medication mismanagement: using cognitive screens to predict a patient's accuracy in filling a pillbox. Consult Pharm 2008 Jun;23(6):459-472. [Medline: 18764676]
42. Carlson MC, Fried LP, Xue QL, Tekwe C, Brandt J. Validation of the Hopkins Medication Schedule to identify difficulties in taking medications. J Gerontol A Biol Sci Med Sci 2005 Feb;60(2):217-223. [Medline: 15814866]
43. Gotsis M, Wang H, Spruijt-Metz D, Jordan-Marsh M, Valente TW. Wellness partners: design and evaluation of a web-based physical activity diary with social gaming features for adults. JMIR Res Protoc 2013 Feb;2(1):e10 [FREE Full text] [doi: 10.2196/resprot.2132] [Medline: 23611986]
44. Gould ON, McDonald-Miszczak L, King B. Metacognition and medication adherence: how do older adults remember? Exp Aging Res 1997;23(4):315-342. [doi: 10.1080/03610739708254034] [Medline: 9352290]
45. Pyatak EA, Florindez D, Weigensberg MJ. Adherence decision making in the everyday lives of emerging adults with type 1 diabetes. Patient Prefer Adherence 2013;7:709-718 [FREE Full text] [doi: 10.2147/PPA.S47577] [Medline: 23935361]
46. 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

ICT: information and communication technology
LCD: liquid-crystal display
LDL: low-density lipoprotein
MMAS-4: 4-item Morisky Medication Adherence Scale
SMS: short message service

Edited by G Eysenbach; submitted 22.01.14; peer-reviewed by G Bacigalupe, L Fernandez-Luque; comments to author 12.02.14; revised version received 04.03.14; accepted 13.03.14; published 04.04.14.

Please cite as:
Mira JJ, Navarro I, Botella F, Borrás F, Nuño-Solinís R, Orozco D, Iglesias-Alonso F, Pérez-Pérez P, Lorenzo S, Toro N
A Spanish Pillbox App for Elderly Patients Taking Multiple Medications: Randomized Controlled Trial
J Med Internet Res 2014;16(4):e99
URL: http://www.jmir.org/2014/4/e99/
doi:10.2196/jmir.3269
PMID:24705022

Viewpoint

Collaborative Biomedicine in the Age of Big Data: The Case of Cancer

Abdul R Shaikh[1], MHSc, PhD; Atul J Butte[2], MD, PhD; Sheri D Schully[3], PhD; William S Dalton[4], MD, PhD; Muin J Khoury[3], MD, PhD; Bradford W Hesse[3], PhD

[1]PricewaterhouseCoopers LLP, McLean, VA, United States
[2]Division of Systems Medicine, Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, United States
[3]Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD, United States
[4]DeBartolo Family Personalized Medicine Institute at the Moffitt Cancer Center, Moffitt Cancer Center, Tampa, FL, United States

Corresponding Author:
Abdul R Shaikh, MHSc, PhD
PricewaterhouseCoopers LLP
1800 Tysons Boulevard
McLean, VA, 22102
United States
Phone: 1 301 448 0057
Fax: 1 416 340 3200
Email: shaikh@us.pwc.com

Abstract

Biomedicine is undergoing a revolution driven by high throughput and connective computing that is transforming medical research and practice. Using oncology as an example, the speed and capacity of genomic sequencing technologies is advancing the utility of individual genetic profiles for anticipating risk and targeting therapeutics. The goal is to enable an era of “P4” medicine that will become increasingly more predictive, personalized, preemptive, and participative over time. This vision hinges on leveraging potentially innovative and disruptive technologies in medicine to accelerate discovery and to reorient clinical practice for patient-centered care. Based on a panel discussion at the Medicine 2.0 conference in Boston with representatives from the National Cancer Institute, Moffitt Cancer Center, and Stanford University School of Medicine, this paper explores how emerging sociotechnical frameworks, informatics platforms, and health-related policy can be used to encourage data liquidity and innovation. This builds on the Institute of Medicine’s vision for a “rapid learning health care system” to enable an open source, population-based approach to cancer prevention and control.

(J Med Internet Res 2014;16(4):e101) doi:10.2196/jmir.2496

KEYWORDS

biomedical research; crowdsourcing; health information technology; innovation; precision medicine

Introduction

Biomedicine is undergoing a revolution driven by innovation in high throughput and connective computing [1,2], big data [3,4], and evolving models of individual and population care [5,6]. Emerging informatics technologies and platforms are being used to combine molecular, clinical, and population data to better anticipate risk, target therapeutics, and manage care for cancer and other diseases [6,7]. Based on a panel discussion at the 2012 Medicine 2.0 conference (5th World Congress on Social Media, Mobile Apps, and Internet/Web 2.0 in Health and Medicine) at the Harvard Medical School conference center in Boston with representatives from the National Cancer Institute (NCI), Moffitt Cancer Center (MCC), and Stanford University School of Medicine, this paper explores how emerging technologies and innovative care models that build on the concepts of “P4” medicine (ie, predictive, personalized, preemptive, and participatory) [8] and the learning health care system [9] can help enable an open source, population-based approach to cancer prevention and control.

P4 Medicine and the Learning Health System

The growing speed and capacity of genomic sequencing technologies are advancing the utility of individual genetic profiles for anticipating risk and targeting therapeutics for cancer [10]. Combining the digital revolution with genomics and other "omics" fields, the term P4 medicine implies a systems approach to biology and medicine that brings together molecular immunology, advanced computation, biotechnology, and genomics, among other fields [8]. Standing for *predictive* profiles of risk, *preventive* clinical and wellness systems, *personalized* medicine, and *participative* research and practice, proponents of P4 medicine extend its purview beyond genomics to include multiple data vectors such as longitudinal molecular, cellular, and phenotypic data for predicting disease progression and targeting intervention [11]. Thus, P4 medicine is predicated on the notion that individual disease and broader notions of health and wellness can be quantified with advanced computation and informatics through systems approaches to decipher the inherent complexity of billions of data points surrounding patients in the future.

P4 medicine can be viewed in conjunction with the Institute of Medicine's (IOM) concept of a learning health system, which simultaneously links effective, efficient clinical health care to the biomedical research enterprise [12]. As presented in a draft proposal at the IOM National Cancer Policy Forum, the notion of a rapid learning health system for cancer utilizes basic translational, comparative effectiveness, and health services research synchronized with optimal delivery of precision care. This model of research and practice is based on two elements: (1) a sufficiently advanced digital health infrastructure that can fully utilize the phenomenon of (2) data liquidity, defined as "the rapid, seamless, secure exchange of useful, standards-based information among authorized individual and institutional senders and recipients" [13].

Both of these innovative approaches to health research and practice—P4 medicine and the learning health system—require robust technology infrastructure and data liquidity to realize the ambitious aim of transforming biomedicine for cancer and other diseases. Moreover, in presenting the rapid learning health system framework for cancer research and practice, members of the National Cancer Policy Forum's planning committee identified five challenges to developing a learning health system for cancer that are also directly relevant to the realization of P4 medicine: (1) data collection (eg, data accuracy, timeliness, and completeness), (2) incentivizing data-sharing, (3) data standards, harmonization, and computation, (4) meaningful use of health IT, and (5) the central role of government entities such as the National Institutes of Health (NIH), the Food and Drug Administration, and the Centers for Medicare and Medicaid Services [13].

One example of a rapid learning health care system that is currently being implemented for oncology is the American Society of Clinical Oncology's Cancer Learning Intelligence Network for Quality (CancerLinQ) system [14]. CancerLinQ is designed to address the growing challenge of managing the deluge of data emerging from precision medicine for cancer care. The system incorporates data from researchers, providers, and patients in order to continually improve comprehensive clinical algorithms reflecting preferred care at a series of decision nodes for clinical decision support.

"P5 Medicine": A Population Approach to Transforming Biomedicine

Adding both promise and complexity to the previously described frameworks of modern care, proponents of the public health sciences assert that a 5th P standing for a *population* perspective is needed to realize the full potential of P4 medicine [15]. Limited by a primary focus on individual health, the P4 approach to biomedicine can be augmented as follows:

- Predictive: Predicting health using systems biologic and phenotypic information augmented with the ecological model of health to account for multilevel determinants of health and life-course approaches.
- Preventive: Early disease detection and prevention also incorporate population screening principles to assess benefits, harm, and costs of primary prevention.
- Personalized: Targeted therapeutics and diagnostics enhanced by principles of evidence-based medicine using formal analytic frameworks for comparative effectiveness.
- Participatory: Engaging patients, providers, and systems including the public health enterprise (eg, policy development, regulatory science, implementation, and health services research).

In addition to the rapid learning health system approach, which incorporates notions of translational, comparative effectiveness and health services research, adding a population focus to P4 medicine explicitly addresses broader, structural issues such as costs and potential for harm that result in greater social, economic, and health disparities. Population science also helps focus on the need for enhanced population level interventions (such as education, employment, and roads) in addition to individual level interventions to improve health and prevent disease [15].

One innovative example of an effort incorporating a P5 approach to cancer biomedicine can be found in the Moffitt Cancer Center's Total Cancer Care (TCC) proposal for a new federated model for research and health care [16]. Based on a robust informatics platform allowing for real-time integration and analysis of disparate multilevel data, the TCC builds on the rapid learning health system model by incorporating development of "secondary use" of data including comparative effectiveness research. Perhaps equally important, TCC proposes a shared governance approach with a federated data model designed to promote team science, data liquidity, and access to the disparate data sources that are essential for effective transformation of the biomedical enterprise [17].

Crowdsourcing Science and the Future of Biomedicine

Scaling the biomedical research enterprise to tackle cancer and other diseases with unknown therapeutics and unclear diagnostics will require recruiting new communities of investigators such as those in engineering and computational disciplines, often earlier in their careers. In addition to exploring new models of cancer research and practice, the Medicine 2.0

conference panel also delved into how big data, emerging technologies, and commoditized access to sophisticated wet lab tools and computational methods can spark scientific innovation in basic and applied research. Individuals have greater access to potentially disruptive technologies in medicine to accelerate basic discovery science and reorient clinical practice for patient-centered care. Publically available molecular measurements can be used to discover novel biomarkers of disease [18] and can be used to find novel uses for existing therapeutics [19,20]. One example of individuals addressing big data challenges is in the field of computational immunogenics, where a challenge sponsored by Harvard Medical School was used to crowdsource solutions that significantly outperformed leading academic efforts [21]. Such immunology datasets, including clinical trials, are available at the National Institute of Allergy and Infectious Diseases ImmPort website for professionals and students.

On a policy level, examples such as the federal Open Government Initiative, the America COMPETES Act, and NIH requirements for data sharing in grant proposals, combined with public and private sector initiatives by donors, journals, and foundations, have led to unprecedented amounts of data being available for secondary research. Two examples include the Data.gov platform, which enables public access to "nearly 450,000 datasets…across 172 federal agencies" [22], and the availability of one million gene expression microarray measurements for research [23].

In addition to greater availability of data, public and private entities are leveraging prize and challenge mechanisms to accelerate innovation with health-related data. Biomedically related open innovation challenges such as these often involve the release of data first, with the expectation that the "winner" of the challenge is awarded a prize. This is the reverse of the typical grant funding mechanism, with money given first, potentially followed by results [24]. In the cancer arena, NCI and the Office of the National Coordinator for Health Information Technology (ONC) have had success in combining open innovation challenges with the federal Small Business Innovation Research (SBIR) grant program to support the evaluation and dissemination of evidence-based applications for cancer prevention and control [25,26]. For these agencies, the federal prize and challenge mechanism has provided a high-value approach to addressing their core agency missions through building a new ecosystem of developers, entrepreneurs, and scientists who can innovate for cancer control and public health. The most recent ONC and NCI challenge competition, focusing on technology innovation for cancer survivors, expands the Department of Health and Human Services innovation portfolio by incorporating crowdfunding to potentially enhance engagement and market validation of submitted innovations with consumer audiences [27].

Conclusions

In many respects, cancer is the prototypical workspace for applying new models of scientific discovery and medical practice. The story of cancer is a story of how the body's complex coding systems go awry through the creation of self-perpetuating errors in cellular replication and growth. Fortunately, advances in genetic sequencing technologies, high throughput data architectures, massively networked public and scientific communities, and the wide availability of sophisticated wet lab tools may be sparking the innovation in "open source" science needed to accelerate progress against the disease. As one panel member put it, "individuals in garages and dorm rooms have greater access to potentially disruptive technologies in medicine than the most well-resourced scientists of the last decade." This exciting era of distributed and open source science holds great potential for accelerating basic discovery and reorienting clinical practice for patient-centered care and population health.

Conflicts of Interest

WS Dalton is CEO of M2Gen, affiliated with Moffitt Cancer Center. No conflicts declared for all other authors.

References

1. President's Council of Advisors on Science and Technology. Realizing the Full Potential of Health Information Technology to Improve Healthcare for Americans: The Path Forward. Washington, DC: The White House; 2010.
2. President's Council of Advisors on Science and Technology. Designing a Digital Future: Federally Funded Research and Development in Networking and Information Technology. Washington, DC: The White House; 2010.
3. Big Data: Community Cleverness Required. [editorial] Nature. 2008. URL: http://www.nature.com/nature/journal/v455/n7209/full/455001a.html [accessed 2014-03-30] [WebCite Cache ID 6OSsZ5F61]
4. Shaikh AR, Prabhu Das I, Vinson CA, Spring B. Cyberinfrastructure for consumer health. Am J Prev Med 2011 May;40(5 Suppl 2):S91-S96. [doi: 10.1016/j.amepre.2011.02.012] [Medline: 21521603]
5. Institute of Medicine. In: Murphy S, editor. A Foundation for Evidence-Driven Practice: A Rapid Learning System for Cancer Care: Workshop Summary. Washington, DC: National Academies Press; 2010.
6. Butte AJ, Shah NH. Computationally translating molecular discoveries into tools for medicine: translational bioinformatics articles now featured in JAMIA. J Am Med Inform Assoc 2011;18(4):352-353 [FREE Full text] [doi: 10.1136/amiajnl-2011-000343] [Medline: 21672904]
7. Dalton WS, Sullivan DM, Yeatman TJ, Fenstermacher DA. The 2010 Health Care Reform Act: a potential opportunity to advance cancer research by taking cancer personally. Clin Cancer Res 2010 Dec 15;16(24):5987-5996 [FREE Full text] [doi: 10.1158/1078-0432.CCR-10-1216] [Medline: 21169252]

8. Zewail AH. A Systems Approach to Medicine Will Transform Healthcare. In: Physical biology: from atoms to medicine. London: Imperial College Press; 2008.
9. Olsen L, Aisner D. The learning healthcare system: workshop summary. Washington, DC: National Academies Press; 2007.
10. American Association of Cancer Research. Cancer Progress Report 2012: Making Research Count for Patients: A New Day. 2012. URL: http://cancerprogressreport.org/2012/Documents/2012_Report.pdf [accessed 2012-12-15] [WebCite Cache ID 6CvqufmCg]
11. Hood L, Friend SH. Predictive, personalized, preventive, participatory (P4) cancer medicine. Nat Rev Clin Oncol 2011 Mar;8(3):184-187. [doi: 10.1038/nrclinonc.2010.227] [Medline: 21364692]
12. Series THHR, Institute of Medicine, Grossman C. Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care: Workshop Series Summary (The Learning Health System Series). Washington, DC: National Academies Press; 2011.
13. Kean MA, Abernethy AP, Clark AM, Dalton WS, Pollock BH, Shulman LN, et al. Achieving Data Liquidity in the Cancer Community: Proposal for Coalition of All Stakeholders. 2012. Achieving Data Liquidity in the Cancer Community: Proposal for Coalition of All Stakeholders URL: http://www.iom.edu/Global/Perspectives/2012/DataLiquidityCancerCommunity.aspx [accessed 2012-12-15] [WebCite Cache ID 6CvqYTYo7]
14. CancerLinq. 2013. URL: http://www.asco.org/institute-quality/cancerlinq [accessed 2013-05-08] [WebCite Cache ID 6GSy5HHAS]
15. Khoury MJ, Gwinn ML, Glasgow RE, Kramer BS. A population approach to precision medicine. Am J Prev Med 2012 Jun;42(6):639-645 [FREE Full text] [doi: 10.1016/j.amepre.2012.02.012] [Medline: 22608383]
16. Fenstermacher DA, Wenham RM, Rollison DE, Dalton WS. Implementing personalized medicine in a cancer center. Cancer J 2011;17(6):528-536 [FREE Full text] [doi: 10.1097/PPO.0b013e318238216e] [Medline: 22157297]
17. Dalton WS. A Partnership to Develop a National Health & Research Information Exchange (NHRIE). 2012. URL: http://ncip.nci.nih.gov/ncip-launch-meeting-presentations/1230B-Dalton-Slides-for-5-31-Informatics-SessionVer1.pdf [accessed 2012-12-15] [WebCite Cache ID 6Cvqrh1Ju]
18. Chen R, Sigdel TK, Li L, Kambham N, Dudley JT, Hsieh SC, et al. Differentially expressed RNA from public microarray data identifies serum protein biomarkers for cross-organ transplant rejection and other conditions. PLoS Comput Biol 2010;6(9):e1000940 [FREE Full text] [doi: 10.1371/journal.pcbi.1000940] [Medline: 20885780]
19. Sirota M, Dudley JT, Kim J, Chiang AP, Morgan AA, Sweet-Cordero A, et al. Discovery and preclinical validation of drug indications using compendia of public gene expression data. Sci Transl Med 2011 Aug 17;3(96):96ra77 [FREE Full text] [doi: 10.1126/scitranslmed.3001318] [Medline: 21849665]
20. Dudley JT, Sirota M, Shenoy M, Pai RK, Roedder S, Chiang AP, et al. Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. Sci Transl Med 2011 Aug 17;3(96):96ra76 [FREE Full text] [doi: 10.1126/scitranslmed.3002648] [Medline: 21849664]
21. Lakhani KR, Boudreau KJ, Loh PR, Backstrom L, Baldwin C, Lonstein E, et al. Prize-based contests can provide solutions to computational biology problems. Nat Biotechnol 2013 Feb;31(2):108-111. [doi: 10.1038/nbt.2495] [Medline: 23392504]
22. Data.gov celebrates third anniversary. URL: http://www.data.gov/whats-new [accessed 2012-12-15] [WebCite Cache ID 6CvqxyO1z]
23. Baker M. Gene data to hit milestone. Nature 2012 Jul 19;487(7407):282-283. [doi: 10.1038/487282a] [Medline: 22810669]
24. Lakhani KR, Panetta JA. The Principles of Distributed Innovation: Technology, Governance, Globalization. Innovations: Technology, Governance, Globalization 2007;2(3):1034 [FREE Full text]
25. National Cancer Institute, Office of the National Coordinator for Health Information Technology. Using Public Data for Cancer Prevention and Control: From Innovation to Impact. URL: http://www.health2con.com/devchallenge/using-public-data-for-cancer-prevention-and-control-from-innovation-to-impact-2/ [accessed 2012-12-15] [WebCite Cache ID 6Cvr2QgPU]
26. National Cancer Institute. SBIR Program Announcement: Innovative Health Information Technology for Broad Adoption by Healthcare Systems and Consumers. URL: http://sbir.cancer.gov/resource/hit/ [accessed 2012-12-15] [WebCite Cache ID 6Cvr4dCr5]
27. Office of the National Coordinator for Health Information Technology, National Cancer Institute. Crowds Care for Cancer: Supporting Survivors. URL: http://challenge.gov/ONC/529-crowds-care-for-cancer-supporting-survivors [accessed 2013-05-08] [WebCite Cache ID 6GSiEL1Rj]

Abbreviations

CancerLinQ: Cancer Learning Intelligence Network for Quality
IOM: Institute of Medicine
MCC: Moffitt Cancer Center
NCI: National Cancer Institute
NIH: National Institutes of Health

ONC: Office of the National Coordinator for Health Information Technology
P4: predictive, personalized, preemptive, participatory
SBIR: Small Business Innovation Research
TCC: Total Cancer Care

Edited by G Eysenbach; submitted 15.12.12; peer-reviewed by L Toldo, Y Zhu; comments to author 27.03.13; revised version received 10.05.13; accepted 03.03.14; published 07.04.14.

Please cite as:
Shaikh AR, Butte AJ, Schully SD, Dalton WS, Khoury MJ, Hesse BW
Collaborative Biomedicine in the Age of Big Data: The Case of Cancer
J Med Internet Res 2014;16(4):e101
URL: http://www.jmir.org/2014/4/e101/
doi:10.2196/jmir.2496
PMID:24711045

Original Paper

Would You Tell Everyone This? Facebook Conversations as Health Promotion Interventions

Jonathan Syred[1], BSc (Hons), MSc; Carla Naidoo[2], BSc (Hons), DipHE(PN); Sarah C Woodhall[3], BSc, MSc; Paula Baraitser[1,4], MBBS, MD, FFPH

[1]HIV & Sexual Health Research Group, King's College London, London, United Kingdom
[2]National Chlamydia Screening Programme, Public Health England, London, United Kingdom
[3]HIV & STI Department, Public Health England, London, United Kingdom
[4]Department of Sexual Health and HIV, Kings College Hospital NHS Foundation Trust, London, United Kingdom

Corresponding Author:
Paula Baraitser, MBBS, MD, FFPH
HIV & Sexual Health Research Group
King's College London
Weston Education Centre
10 Cutcombe Road
London, SE5 9RJ
United Kingdom
Phone: 44 020 7848 5776
Fax: 44 020 7848 5769
Email: paula_baraitser@mac.com

Abstract

Background: Health promotion interventions on social networking sites can communicate individually tailored content to a large audience. User-generated content helps to maximize engagement, but health promotion websites have had variable success in supporting user engagement.

Objective: The aim of our study was to examine which elements of moderator and participant behavior stimulated and maintained interaction with a sexual health promotion site on Facebook.

Methods: We examined the pattern and content of posts on a Facebook page. Google analytics was used to describe the number of people using the page and viewing patterns. A qualitative, thematic approach was used to analyze content.

Results: During the study period (January 18, 2010, to June 27, 2010), 576 users interacted 888 times with the site through 508 posts and 380 comments with 93% of content generated by users. The user-generated conversation continued while new participants were driven to the site by advertising, but interaction with the site ceased rapidly after the advertising stopped. Conversations covered key issues on chlamydia and chlamydia testing. Users endorsed testing, celebrated their negative results, and modified and questioned key messages. There was variation in user approach to the site from sharing of personal experience and requesting help to joking about sexually transmitted infection. The moderator voice was reactive, unengaged, tolerant, simplistic, and was professional in tone. There was no change in the moderator approach throughout the period studied.

Conclusions: Our findings suggest this health promotion site provided a space for single user posts but not a self-sustaining conversation. Possible explanations for this include little new content from the moderator, a definition of content too narrow to hold the interest of participants, and limited responsiveness to user needs. Implications for health promotion practice include the need to consider a life cycle approach to online community development for health promotion and the need for a developing moderator strategy to reflect this. This strategy should reflect two facets of moderation for online health promotion interventions: (1) unengaged and professional oversight to provide a safe space for discussion and to maintain information quality, and (2) a more engaged and interactive presence designed to maintain interest that generates new material for discussion and is responsive to user requests.

(J Med Internet Res 2014;16(4):e108) doi:10.2196/jmir.3231

KEYWORDS

social media; chlamydia; health promotion

Introduction

Background

Health promotion interventions on social networking sites (SNS) harness user-generated content and the power of networks to improve health [1]. They reach large numbers of people and have the capacity to communicate tailored messages quickly [2]. Stimulating user-generated content is important to the success of this approach because interactivity influences reach with new participants joining the conversation as they observe the activity of their online friends [3]. Observing or participating in such conversations shares information on the beliefs and experience of peers, an important influence on behavior [4], and interaction with health promotion information, (commenting, challenging, modifying) facilitates engagement with this material and therefore learning from it [5]. User-generated content personalizes generic health promotion messages and adapts them for specific populations, increasing their relevance to new audiences [6].

Although user-generated content has important advantages, it may also lead to widespread sharing of poor quality information and material that is offensive to individuals or groups [6,7]. As sharing of personal experience is an integral part of online community activity, SNS health promotion interventions carry risks for participants. These risks are potentially high where they involve sharing of sexual health information between young people [8], and interventions of this sort require careful moderation to prevent harmful activity.

Although the interactivity of SNS health promotion interventions is an important element of their capacity to deliver health promotion messages, they have been variably successful in stimulating and/or supporting user-generated content and many remain inactive [9]. Levels of interactivity are likely to reflect individual or group motivations for engaging with the site, the technology required to access it, the relationships formed within the group, and the structures for interaction including the presence of a moderator and moderator behavior [10-12]. The role of the moderator in health promotion campaigns on SNS are not fully understood [13], with a perceived trade-off being between encouraging activity and retaining control of content.

A recent systematic review of the use of SNS for health promotion identified a lack of evidence on effective approaches to stimulating interaction [9]. Although there is significant literature on interactions within online communities, particularly those that are business oriented, there is less research in relation to health promotion interventions.

The key metrics to understand interactions on social networking sites are the number of posts and their content [12]. To understand the factors that facilitate interaction on SNS for health promotion, we looked at the volume, pattern, and content of interaction within a national health promotion campaign using the SNS site, Facebook. The health campaign, “Say Yes to the Test”, aimed to encourage testing for genital chlamydia infection among young people in England between January and June 2010 by generating discussion on chlamydia and chlamydia testing. The SNS page was one element of a larger multimedia campaign by the English Department of Health and the Health Protection Agency (now Public Health England) entitled “Sex Worth Talking About”.

We collected quantitative data on the volume of interaction and qualitative data on its content. We sought to identify and study what stimulated and maintained interaction on this Facebook-based sexual health promotion campaign with particular attention to the role of the moderator in stimulating interaction and the role of users in modifying messages. Our research question was “What elements of moderator and participant behavior stimulated and maintained interaction with a sexual health promotion site on Facebook?”

Setting and Approach

Chlamydia trachomatis is the most commonly diagnosed bacterial sexually transmitted infection in the United Kingdom. The prevalence is highest among young people aged 15-24 and infection is usually asymptomatic [14]. The English National Chlamydia Screening Programme aims to offer all sexually active young people under the age of 25 years testing for this infection annually or on change of sexual partner [15].

The “Chlamydia Worth Talking About” strand of the “Sex Worth Talking About” Campaign was a multimedia campaign launched in January 2010 that aimed to encourage (1) open discussion about chlamydia and (2) acceptance of a chlamydia test when offered by a health professional.

The campaign included television, radio, and billboard advertisements and a Facebook page that encouraged discussion about chlamydia testing, entitled “Say Yes to the Test”. The Facebook site was promoted via a separate digital media strategy. The target audience for the campaign was young people (male and female) in England aged 15-24 years. Advertisements (traffic drivers) placed on sites used by young people suggested either that chlamydia “knew” someone they knew or that it had “poked” them. It was anticipated that curiosity arising from these ads would cause the user to click on them. The welcome page for the health promotion intervention offered options to become a fan of the page and post comments on the “comment wall”. This gave users the opportunity to affirm their positive attitude among their own friends, spreading the word and encouraging conversations.

The moderation style employed on the “Say Yes to the Test” site was non-interventionist. Only comments that were completely unrelated to sexual health or that were directly offensive were removed. Factually incorrect comments were left on the site giving the peer group opportunity to respond before the moderation team intervened.

Facebook Environment

A Facebook community page is organized around a publically displayed wall where new content (messages, media, or links) can be added by the owner (or moderator) and other Facebook users [16]. Any form of interaction, whether liking or posting content, can be seen on the user’s newsfeed. This may be shared automatically with people within their network depending on the level of interaction that they have specified with the other user [17].

Methods

The literature on online communities suggests that the key metrics for evaluation of this type of intervention are the volume of member contributions and the quality of the online relationships formed [12]. We therefore used a combination of quantitative data to describe the volume of contributions and completed a qualitative analysis of the content of the interaction. We described the users and their patterns of use and analyzed the text posted. We looked at data from the first 5 months of the campaign from January 18 to June 27, 2010, as this was the period of highest activity on the “Say Yes to the Test” site. The total number of fans, wall posts, and comments over time, fan demographics (gender, age, and country where page was accessed) were obtained from the Facebook page administrators. Google analytics were used to document numbers using the site and viewing patterns. We captured the page content from within the study period using the NCapture function of NVivo 10 software and used simple counting to describe user and moderator content, discussion thread patterns, and moderator intervention.

The qualitative analysis was completed by 2 investigators (JS and PB) using the framework approach and initially coding together to ensure consistency. Disagreements were discussed until a consensus was reached. All coding was done using NVivo 10.

We completed an initial process of familiarizing ourselves with the data by reading and re-reading the posts, and from this, we identified an initial set of themes to describe interaction with the site: (1) patterns of posting, and (2) content of posts. From this, we developed three coding strategies each building on the previous one (Table 1). We kept the two main elements of coding—patterns and content—constant.

We applied for ethical approval for this study from a local National Health Service Research Ethics Committee. The committee reported that ethical approval was not required since our analysis was of publically available data that we did not link to personal profiles. Our use of the data is consistent with guidelines on the ethical conduct of qualitative research on online communities [18]. Furthermore, we have followed recent recommendations from research in similar contexts and used non-verbatim quotes to prevent identification of the users through a search engine [19]. These were constructed through interchanging the words from several posts from different authors that were thematically similar.

Table 1. Coding categories.

	Patterns of posting	Content of posts
Round 1 coding: simple description of patterns and content of interaction (content codes)	Conversation length: conversations coded by number of posts	Attitudes to chlamydia/testing Experience of testing Offensive/stigmatizing/inaccurate material
	Conversation source: content coded by source of posts (users/moderator)	Requests for information/advice
Round 2 coding: factors that triggered interactivity (response codes)	Conversation length: factors associated with longer conversation length	Responses to inaccurate information, challenges to key messages or stigmatizing or offensive material
	Conversation source: factors that triggered user or moderator intervention	Responses to key messages (eg, questioning/endorsing)
Round 3 coding: responses to initial interactions	Interactivity: patterns of questions and answers between users and other users, and users and the moderator.	Content of unresolved and complex issues and moderator responses to these

Results

Patterns of Interaction

There were 191,072 page views during the 5-month study period. The largest cumulative total number of fans was 68,174 fans at Week 7. Two-thirds (64%) of the fans were female, 96% of fans were aged 13-24 years (Table 2), and 95% of fans were from the United Kingdom.

The number of fans increased from the launch of the site until Week 7, when the advertising ceased and when the total number of fans accumulated during this 7-week period reached 68,174 (6/3/2010). It subsequently decreased by about 100 fans per week (Figure 1). The majority of the activity was seen between Weeks 2 and 6 (Figure 2).

We collected 888 interactions (508 posts, 380 comments) from 576 unique users; 93% of content was from users. In contrast to the 68,174 “likes” for the site, content interaction received a much smaller number with 483 “likes” for 156 posts or comments.

Most users interacted once (77.6%) and fewer than 2% posted more than five times, with 17 posts the highest number of interactions from a single user.

There were 164 threads where one or more users commented on an original post. These were usually short (one comment followed by one post), although a minority were longer with the longest being made up of 56 comments.

The number of fans and posts related closely to the advertising campaign and activity decreased very rapidly once the advertising stopped, suggesting that the interactions on the site themselves were insufficient to maintain user input or to generate new participants.

Table 2. Age of fans of “Say yes to the test” by sex.[a]

Age range	Male, %	Female, %	All, %
13-17	43	54	50
18-24	54	43	47
25-34	2	2	2
35-44	<1	<1	<1
55+	<1	<1	<1

[a] The proportion in each age group calculated on a daily basis and averaged across the data collection period.

Figure 1. Total fans, Weeks 1 (18/01/2010) to 10.

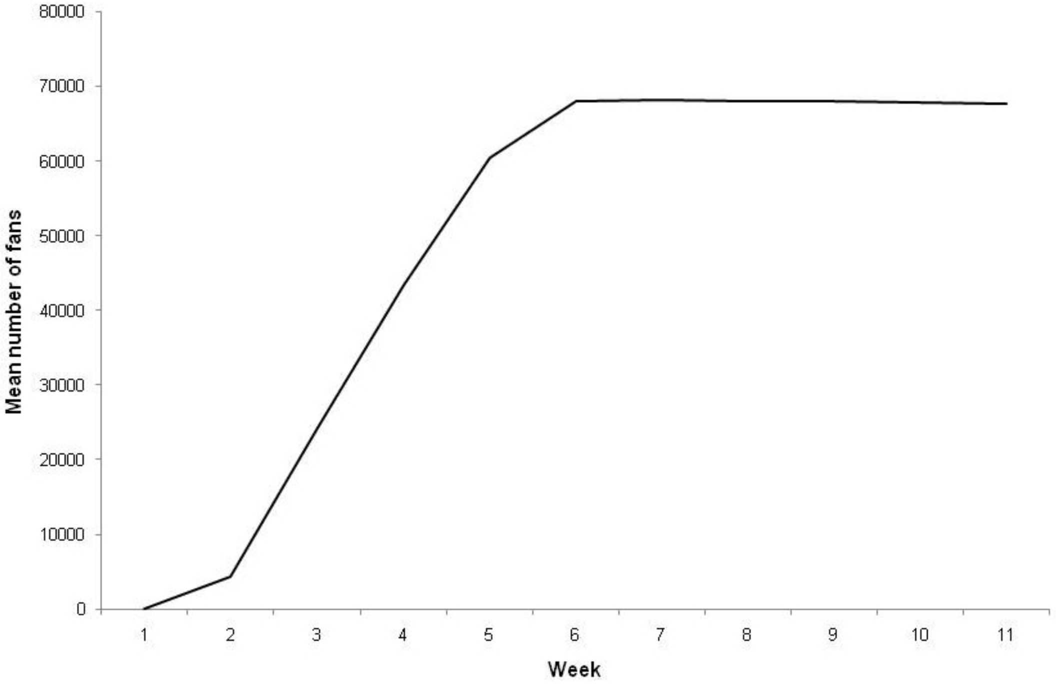
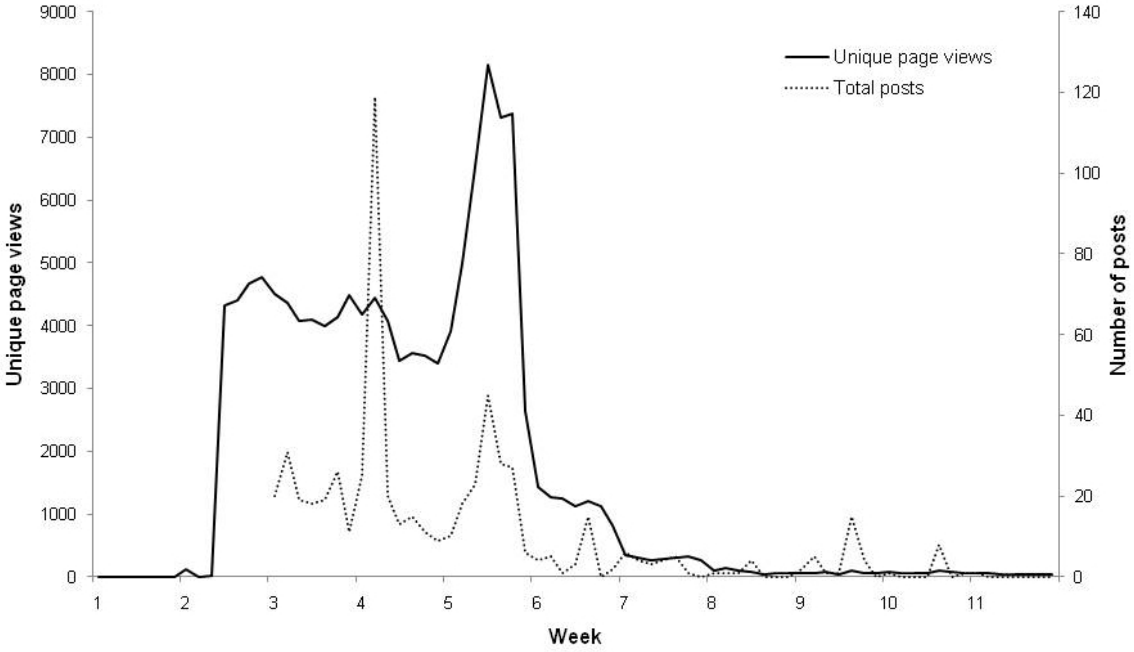


Figure 2. Total wall posts/comments and unique page views, Weeks 1(18/01/2010) to 10.



Content of Interaction

Overview

Conversations covered key issues on chlamydia and chlamydia testing including the lack of symptoms, consequences of infection, and experience of getting tested and treated.

Attitudes to Chlamydia Testing

Posts endorsed chlamydia testing as the right thing to do, often reproducing the campaign message “Say Yes to the Test”. Sixty-two posts encouraged others to get tested for peace of mind because it is easy to do and to prevent subfertility:

> *Bout time peeps startd tikin about this…y not go gt check 4eurything its 20mins out ur life…tht ultimately cud save ur life and tht of ur future children and new partners...sense...people..gud on ya if ya dne it...an if u ant y…x* [Poster 1]

Although testing is generally presented as the “right thing to do”, there was disagreement about its implications. Some participants felt that testing suggests risk of infection through multiple partners with either negative implications (promiscuity) or positive ones (sexual experience). Others suggested that it signals irresponsible behavior (no condoms used), carelessness (poor condom use), or bad luck. Some felt that it is a routine part of looking after yourself:

> *People who say like “oh im not ashamed, I have this test all the time”...stop having sex with so many diff men, then you wouldn’t need to have it done all the time =]*
>
> *It’s making it “shameful” that prevents people from going and getting tested. The skanky ones are the ones that don’t get tested you fool. If anything everyone that is a fan of this group are the absolute opposite.*
>
> *I went for it its all good. I got the all clear too. just go on a random moment, just means your looking after yourself*

Uncertainty Around Testing Messages

User-generated content identified and did not resolve uncertainty around some aspects of chlamydia testing. This maintained conversations by generating disagreement and encouraged sharing personal experience. It identified questions that were important to participants but failed to resolve them because the approach to moderation was simple, factual responses without engaging in conversation.

The effectiveness of condoms in preventing chlamydia was an example of this and a subject of concern. If condoms protect against chlamydia, why should those who use condoms consistently still be tested? Similarly, users were confused about the risk of subfertility following chlamydia infection. Some posters reported that after a single episode of infection becoming pregnant was impossible and that chlamydia infection could therefore be used as a method of contraception.

None of these questions have simple answers since the effectiveness of condoms depends on how consistently and carefully they are used. The impact of chlamydia on subfertility is important but contested. It is very difficult to measure accurately and not completely resolved by treatment.

> *the result was.....positive. no kids for me shammmme init* [Poster 1]
>
> *oviously twat* [Poster 2]
>
> *u cn still have kids if u get it treated* [Poster 3]

Results of Getting Tested

Most of those who discuss their experience of testing report their results as negative and celebrate this (71/128) with “I’m clean” (63/128) and “never thought I had it and now I know I don’t” as common posts associated with a tangible sense of relief after the anxiety of waiting for results.

Many fewer comments (28/128) report a positive chlamydia test. Most of these fall into the “just joking” category, often emphasizing “I don’t care” or “I’m pleased” or exaggerating the consequences, for example, “my penis will drop off”. The minority of posts that asked for help were important opportunities for participants to offer support and challenge stigma:

> *im really worried because I think I have chlamydia, … WHAT SHOULD I DO?!* [Poster 1]
>
> *have u got any symptoms like a rash or anything like urination being different ive added u ive had it b4 ill chat to u if u like im xxxxxx unlike some idiots on here* [Poster 2]

There were marked differences in approach to the topic between those who were seriously concerned about chlamydia, from a personal or public health perspective, and those who used the topic as an opportunity to post in a joking manner.

Modification and Repetition of Health Promotion Messages

Many posts and comments (n=158) contained messages that adapted but were consistent with key campaign messages about testing and condom use, for example, “dont be fool wrap your tool, strap up before you whack up boys”.

Whether to share personal information in a public forum was discussed referencing both the need for openness to provide information and support and the personal consequences of sharing information online where it is accessible by both those from offline social networks and strangers:

> *had it, got rid of it, get tested few times a year just coz I can* [Poster 1]
>
> *why wud u tell everyone this?? Nice lass u r eh!! well I see u deleted ur comment nd I mean why wud u want everyone to know u had chlamydia twice?? Just makes u look like a skank tbh them things are usually kept private* [Poster 2]

Harmful or Offensive Messages

There were 39 posts classified as potentially harmful or embarrassing, as a named individual or poster of a previous message was the subject of offensive language or inappropriate remarks. We were unable to establish the tone of many of the challenging messages without further context including knowledge of the relationship between posters, for example, "I got chlamydia off Stevey ". Some harmful posts were ignored by the moderator to discourage continued posting. We were unable to view those deleted with a private warning message. Of all 164 comment threads, harmful or offensive comments appeared in 31 (18.9%) of them at least once. Of these message chains, 40 had four or more comments following an original post. Of these 40 longer threads, offensive content appeared in 12 (30%) at least one or more times. We had hypothesized that potentially harmful content or challenges of key messages might stimulate message chains, but this content did not appear to drive longer threads.

The Role of the Moderator

Of all 888 posts and comments, 62 (6.9%) were by the moderator. The approach adopted was to choose from a list of standardized responses and repeat these as the same issue occurred. There was no attempt to engage in the conversation, stimulate discussion, or reproduce the language used by posters. This means that the tone of posts by the moderator was very different from that of the users. The moderator did not respond to misinformation posted on the site immediately but allowed time for other posters to correct it and then endorsed the corrections:

> *Hi xxx! Just to clarify, neither chlamydia infection or the health consequences of untreated chlamydia infection would lead to penis amputation! However, untreated chlamydia can cause painful inflammation in one or both testicles… To find out how it is important to "Say yes to the test" log on to….*

Of the moderator's interactions, 53/62 (85%) were a response to an existing conversation. Of these comments, 43/53 (81%) were the last or second to last comment in the thread.

Users appreciated moderator comments with 192 likes (40%) for moderator interactions. The highest number of likes for a user interaction was 10 and for a moderator interaction 63. There were no cases where the posters expressed dissatisfaction with comments from the moderator.

A total of 80% (51/64) posts or comments that challenged campaign messages or contained inaccurate information or questions were corrected, with 47 (73%) responses within four posts, that is, appearing in the same window as the post on Facebook newsfeeds. Over half of these responses came from users:

> *where you get free test from??* [Poster 1]
>
> *from the docs just go and ask for one or go sex health clinc x* [Poster 2]
>
> *boots do them for free aswell x* [Poster 3]
>
> *any boots?* [Poster 1]
>
> *You can order one online. Its great. You get a pippette :D* [Poster 4]
>
> *i got mine from school, they were giving them out* [Poster 5]

Discussion

Principal Findings

We aimed to study what stimulated and maintained interaction on a sexual health promotion site on Facebook with particular emphasis on the role of the moderator. The stated aim of the site studied was to promote chlamydia testing, to reinforce key messages about chlamydia infection, and to promote the attitudes that make testing more likely [20]. The potential outcomes of participation included information exchange, social support and social interaction [12], as well as better sexual health through increased likelihood of chlamydia testing. The potential risks of participation include the consequences of sharing personal information online and exposure to harmful online behavior such as the posting of inflammatory, hostile, or insulting behavior.

The "Say Yes" site attracted a very large number of potential users, but there was no evidence of sustained interaction. Single visits to the site could have provided information on testing but more active engagement with the site (repeat visits or posting) would be more likely to change attitudes [4]. More sustained interaction from a significant number of people would be associated with the development of an online community, either self-sustaining or supported by new stimuli for discussion from the moderator. One definition of an online community is "a collective group of entities, individuals or organizations that come together either temporarily or permanently through an electronic medium to interact in a common problem or interest space" [21]. The idea of a shared area of interest is important for our discussion. If online communities must make a transition from early engagement to self-sustaining interaction [22] and if key metrics to monitor the success of online communities include the volume of member contributions and the quality of the online relationships formed [12], our analysis suggests that this intervention was not successful in making the transition to a mature online community. There was a low volume of posts once the advertising ceased and the development of superficial relationships as evidenced by the short discussion threads and lack of continued engagement.

Major influences on the amount and type of interaction with SNS-based health promotion interventions include the structure of relationships between users [10], the role of the moderator [11], and the content of the online discussion [23]. Drawing on this work, we propose that barriers to sustained interaction could include an audience that did not have a sufficiently shared approach to this topic, a lack of new content from the moderator, a definition of content that was too narrow to hold the interest of participants, and a lack of moderator responsiveness to participant needs. The evidence for these proposals from our study are explored below.

On the "Say Yes" site, there was some indication of shared interest in chlamydia as users had "liked" the site and there was reference to a common experience of chlamydia testing among some of those who posted. People join online discussions where they identify others with similar interests and viewpoints [21,24]. An effective online community will attract users with sufficient common interest to provide a safe space for discussion and sufficient difference to provide material for conversation. Our findings suggest marked differences in approach to the topic, for example wide differences in how seriously they approached the content with some users sharing personal experience of and concerns about chlamydia infection while others ridicule the idea of sexually transmitted infection. Since support from others is an important incentive to stay engaged with online communities [24,25], the lack of a supportive environment from some users might have limited the willingness of those who did post topics for discussion to pursue their interest in this forum.

We think that the nature of the subject matter, a mainly short-term condition that is easily treated, meant that the incentives to stay engaged with the site are very different from those that support users with long-term conditions. Strategies to maintain engagement could have included broadening the discussion to other areas of sexual health over time so that new content was generated for discussion, providing a greater incentive to stay engaged and a promise of more information on new, but related topics.

The approach to moderation was reactive (it did not generate new material), unengaged (it provided simple factual responses), tolerant (it gave participants time to respond to inaccurate or challenging material), simplistic (it ignored controversial or complex material and repeated simple health promotion messages), and professional in tone. This fulfilled the important function of maintaining the quality of information on the site [6] and provided space for user-generated content, but it was also repetitive and ultimately uninteresting and did not provide sufficient encouragement for interaction at this early stage of online community development. In particular, it did not respond directly to the needs of users, other than with repetitive health promotion messages and it was not part of a strategy that acknowledged the changing role of the moderator as an online community develops [12]. The idea that online communities progress from small groups with a common interest and informal rules to larger and more established ones with more formal rules and high levels of participation is well documented [12,22]. Groups of the sort that we studied that are created by advertising for a purpose identified by an external agency may follow a different pattern of development closer to online communities developed for consumers by businesses (eg, [26]). Here the early stages of online community development are described slightly differently with an "on board" stage where the community is highly dependent on founder participation and the nature of participation is unclear, to an "established" stage where relationships within the membership base have been established and are less reliant on founder participation.

Delaying response was important to provide space for user-generated content but also carries risks. Users answered almost all of the questions posed by other users but corrected only a third of the inaccurate information posted. With only the last four comments of a discussion thread appearing on Facebook newsfeeds, inaccurate content should be corrected within three comments but this risks closing down discussions. Leaving material that is potentially harmful to a named individual is dangerous [8,27-29] and, from our analysis, is not associated with stimulating conversation. Although removal of this material could encourage repeated posting, we recommend that it be removed. While personally offensive responses were removed, the moderator could not respond to private offensive messages on individual Facebook pages. More information is needed to quantify the risk to participants of sharing information in this way and the responsibilities of the public health agencies who run this type of intervention.

We did not set out to measure the impact of the intervention. The impact of the whole media campaign (of which this intervention was one element) on chlamydia testing rates has been evaluated and shown to be associated with an increased number of positive tests linked to increased testing of high risk individuals although not an increased number of tests overall [25]. We do not know what impact, if any, was related specifically to the SNS intervention; however, the site did have some characteristics of an effective health promotion intervention. Those who visited the site and reviewed the conversation would have received correct information on chlamydia, information on attitudes to chlamydia and chlamydia testing among their peers including normalization of chlamydia testing, reassurance that chlamydia can be easily treated, signposting to additional sources of information, and promotion of condoms to prevent chlamydia infection. They would also have had an opportunity to actively engage with this material, modifying and personalizing it—both are activities associated with active learning and effective health promotion interventions [5].

Limitations

Our findings are limited by our lack of data on those who participated on this site. Without this, we do not know how they are similar or different from the rest of the population targeted by this campaign. Similarly, we have no data on interactions with the site that did not involve posting, for example, we do not know how long each participant spent on the site, what material they read or whether they returned to the site, except where they returned to the site to post. In addition, we were not able to obtain data on posts that were removed by the moderator or warnings sent to individual participants. Finally, we have no data on the impact of this SNS campaign on the knowledge, attitudes, and behaviors of those who viewed or participated in this intervention.

Conclusions

Our findings suggest this health promotion site provided a space for single user posts but not a self-sustaining conversation. Possible explanations for this include little new content from the moderator, a definition of content that was too narrow to hold the interest of participants, and limited responsiveness to user needs.

Implications for health promotion practice include the need to consider a life cycle approach to online community development for health promotion and the need for a developing moderator strategy to reflect this. This strategy should reflect two facets of moderation for online health promotion interventions: (1) unengaged and professional oversight to provide a safe space for discussion and to maintain information quality and (2) a more engaged and interactive presence designed to maintain interest that generates new material for discussion and is responsive to user requests.

Acknowledgments

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The authors would like to thank Pippa Pollard as well as Rebecca Humphreys for comments on the initial analysis of the qualitative data.

Conflicts of Interest

PB, SW, and CM have worked for the English National Chlamydia Screening Programme. CM worked on the development of the “Say Yes to the Test” Campaign.

References

1. 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]
2. Chou WY, Prestin A, Lyons C, Wen KY. Web 2.0 for health promotion: reviewing the current evidence. Am J Public Health 2013 Jan;103(1):e9-18. [doi: 10.2105/AJPH.2012.301071] [Medline: 23153164]
3. Cormode G, Krishnamurthy B. Key differences between Web 10 and Web 2.0. First Monday 2008 2008 Jun 02;13(6):00. [doi: 10.5210/fm.v13i6.2125]
4. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
5. Cairncross S, Mannion M. Interactive Multimedia and Learning: Realizing the Benefits. Innovations in Education and Teaching International 2001 Jan;38(2):156-164. [doi: 10.1080/14703290110035428]
6. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
7. Kamel Boulos MN, Wheeler S. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. Health Info Libr J 2007 Mar;24(1):2-23. [doi: 10.1111/j.1471-1842.2007.00701.x] [Medline: 17331140]
8. Moreno MA, Parks MR, Zimmerman FJ, Brito TE, Christakis DA. Display of health risk behaviors on MySpace by adolescents: prevalence and associations. Arch Pediatr Adolesc Med 2009 Jan;163(1):27-34. [doi: 10.1001/archpediatrics.2008.528] [Medline: 19124700]
9. Gold J, Pedrana AE, Sacks-Davis R, Hellard ME, Chang S, Howard S, et al. A systematic examination of the use of online social networking sites for sexual health promotion. BMC Public Health 2011;11:583 [FREE Full text] [doi: 10.1186/1471-2458-11-583] [Medline: 21777470]
10. Kietzmann JH, Hermkens K, McCarthy IP, Silvestre BS. Social media? Get serious! Understanding the functional building blocks of social media. Business Horizons 2011 May;54(3):241-251. [doi: 10.1016/j.bushor.2011.01.005]
11. Lindsay S, Smith S, Bellaby P, Baker R. The health impact of an online heart disease support group: a comparison of moderated versus unmoderated support. Health Educ Res 2009 Aug;24(4):646-654 [FREE Full text] [doi: 10.1093/her/cyp001] [Medline: 19251770]
12. Iriberri A, Leroy G. A life-cycle perspective on online community success. ACM Comput. Surv 2009 Feb 01;41(2):1-29. [doi: 10.1145/1459352.1459356]
13. 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]
14. Kalwij S, Macintosh M, Baraitser P. Screening and treatment of Chlamydia trachomatis infections. BMJ 2010;340:c1915. [Medline: 20410164]
15. NHS: National Chlamydia Screening Programme. 2013. URL: http://www.chlamydiascreening.nhs.uk/ [accessed 2014-01-03] [WebCite Cache ID 6MMNH2q8N]
16. Facebook: using Facebook. 2013. URL: https://www.facebook.com/unsupportedbrowser [accessed 2014-01-03] [WebCite Cache ID 6MMQlrhkt]
17. Facebook: How News Feed Works. URL: https://www.facebook.com/unsupportedbrowser [accessed 2014-01-03] [WebCite Cache ID 6MMPrh8rV]
18. Eysenbach G, Till JE. Ethical issues in qualitative research on internet communities. BMJ 2001 Nov 10;323(7321):1103-1105 [FREE Full text] [Medline: 11701577]
19. Bond CS, Ahmed OH, Hind M, Thomas B, Hewitt-Taylor J. The conceptual and practical ethical dilemmas of using health discussion board posts as research data. J Med Internet Res 2013;15(6):e112 [FREE Full text] [doi: 10.2196/jmir.2435] [Medline: 23748147]

20. Department of Health. 2009. Teenage Pregnancy and Sexual Health Marketing Strategy URL: http://www.nhs.uk/sexualhealthprofessional/Documents/Sexual_Health_Strategy_2009.pdf?wt.mc_id=21103 [accessed 2014-04-08] [WebCite Cache ID 6OgRuyrXg]
21. Plant R. Online communities. Technology in Society 2004 Jan;26(1):51-65. [doi: 10.1016/j.techsoc.2003.10.005]
22. Andrews DC. Audience-specific online community design. Commun. ACM 2002 Apr 01;45(4):64-68. [doi: 10.1145/505248.505275]
23. Wise K, Hamman B, Thorson K. Moderation, Response Rate, and Message Interactivity: Features of Online Communities and Their Effects on Intent to Participate. J Comp Mediated Comm 2006 Oct;12(1):24-41. [doi: 10.1111/j.1083-6101.2006.00313.x]
24. Porter CE. A typology of virtual communities: A multi-disciplinary foundation for future research. Journal of Computer-Mediated Communication 2004;10(1):00. [doi: 10.1111/j.1083-6101.2004.tb00228.x]
25. Gobin M, Verlander N, Maurici C, Bone A, Nardone A. Do sexual health campaigns work? An outcome evaluation of a media campaign to increase chlamydia testing among young people aged 15--24 in England. BMC Public Health 2013 May 17;13(1):484 [FREE Full text] [doi: 10.1186/1471-2458-13-484] [Medline: 23683345]
26. Mashable. Mashable. How to: Manage a sustainable online community URL: http://mashable.com/2010/07/30/sustainable-online-community/ [accessed 2014-04-08] [WebCite Cache ID 6OgDyMDbZ]
27. Juvonen J, Gross EF. Extending the school grounds?--Bullying experiences in cyberspace. J Sch Health 2008 Sep;78(9):496-505. [doi: 10.1111/j.1746-1561.2008.00335.x] [Medline: 18786042]
28. Ybarra ML, Mitchell KJ. How risky are social networking sites? A comparison of places online where youth sexual solicitation and harassment occurs. Pediatrics 2008 Feb;121(2):e350-e357 [FREE Full text] [doi: 10.1542/peds.2007-0693] [Medline: 18227194]
29. Ybarra ML, Mitchell KJ, Wolak J, Finkelhor D. Examining characteristics and associated distress related to Internet harassment: findings from the Second Youth Internet Safety Survey. Pediatrics 2006 Oct;118(4):e1169-e1177 [FREE Full text] [doi: 10.1542/peds.2006-0815] [Medline: 17015505]

Abbreviations

SNS: social networking site

Edited by G Eysenbach; submitted 07.01.14; peer-reviewed by A Moorhead, R Webster; comments to author 31.01.14; revised version received 05.03.14; accepted 24.03.14; published 11.04.14.

Please cite as:
Syred J, Naidoo C, Woodhall SC, Baraitser P
Would You Tell Everyone This? Facebook Conversations as Health Promotion Interventions
J Med Internet Res 2014;16(4):e108
URL: http://www.jmir.org/2014/4/e108/
doi:10.2196/jmir.3231
PMID:24727742

Original Paper

Mapping Physician Twitter Networks: Describing How They Work as a First Step in Understanding Connectivity, Information Flow, and Message Diffusion

Ranit Mishori[1], MD, MHS; Lisa Oberoi Singh[2*], PhD; Brendan Levy[3*], MD; Calvin Newport[2*], PhD

[1]Department of Family Medicine's Center for Health Communication, Media and Primary Care, Georgetown University School of Medicine, Washington, DC, United States

[2]Department of Computer Science, Georgetown University, Washington, DC, United States

[3]Georgetown University/Providence Hospital Residency in Family Medicine, Colmar Manor, MD, United States

[*]these authors contributed equally

Corresponding Author:
Ranit Mishori, MD, MHS
Department of Family Medicine's Center for Health Communication, Media and Primary Care
Georgetown University School of Medicine
Pre-clinical GB-01D
3900 Reservoir Rd NW
Washington, DC, 20007
United States
Phone: 1 202 687 3011
Fax: 1 202 687 1651
Email: mishorir@georgetown.edu

Abstract

Background: Twitter is becoming an important tool in medicine, but there is little information on Twitter metrics. In order to recommend best practices for information dissemination and diffusion, it is important to first study and analyze the networks.

Objective: This study describes the characteristics of four medical networks, analyzes their theoretical dissemination potential, their actual dissemination, and the propagation and distribution of tweets.

Methods: Open Twitter data was used to characterize four networks: the American Medical Association (AMA), the American Academy of Family Physicians (AAFP), the American Academy of Pediatrics (AAP), and the American College of Physicians (ACP). Data were collected between July 2012 and September 2012. Visualization was used to understand the follower overlap between the groups. Actual flow of the tweets for each group was assessed. Tweets were examined using Topsy, a Twitter data aggregator.

Results: The theoretical information dissemination potential for the groups is large. A collective community is emerging, where large percentages of individuals are following more than one of the groups. The overlap across groups is small, indicating a limited amount of community cohesion and cross-fertilization. The AMA followers' network is not as active as the other networks. The AMA posted the largest number of tweets while the AAP posted the fewest. The number of retweets for each organization was low indicating dissemination that is far below its potential.

Conclusions: To increase the dissemination potential, medical groups should develop a more cohesive community of shared followers. Tweet content must be engaging to provide a hook for retweeting and reaching potential audience. Next steps call for content analysis, assessment of the behavior and actions of the messengers and the recipients, and a larger-scale study that considers other medical groups using Twitter.

(J Med Internet Res 2014;16(4):e107) doi:10.2196/jmir.3006

KEYWORDS

social networking; network analysis; information science; dissemination science; infodemiology; physician communication; physician networks; Twitter

Introduction

Background

Social media, including Facebook and Twitter, is fast becoming an important tool in health care. In editorials, essays, and blogs, physicians have been urged to become active participants in social media as a form of engagement with the larger health community, patients, and peers [1], and as a way to "start an online dialogue" with policy makers and stakeholders [2]. Twitter, the microblogging medium, has been hailed as "an essential tool for every physician leader" [3], one that is "crucial to the development of medicine today" [4], and "just what the doctor ordered" [5].

This rapidly growing social network has approximately 500 million users worldwide, 140 million of them in the United States [6]. Some say it has "democratized" health information exchanges because is it a highly participatory medium where patients, physicians, health care organizations, and other stakeholders can interact on equal footing. Social media, including Twitter, has been described as one of the main tenets of what some scientists have called Medicine 2.0, or "next generation medicine" [7]. Still, despite the increasing interest and the huge potential for information diffusion and its analysis, it is not entirely clear who in the medical world is using Twitter, how much they are using it, and for what purpose [8].

In order to make inferences about group behavior and predictions about "best practices" for dissemination and diffusion of information through these networks, it is important to analyze the networks first. Social network analysis (SNA) is a well-established technique in sociology that can be adapted and used to systematically explore virtual communities, such as those that exist within the world of medicine. Applied graph theory is an overlapping area focused on using graphs to represent structures and networks, and theory developed about graphs to explain applications in a variety of fields, from computer science, to biology and chemistry, to mathematics and linguistics, to name a few. SNA and applied graph theory have been used to analyze structural patterns of social relationships, to explore influential information brokers, and to visualize the formal or informal personal networks within and between organizations.

Online networks have been studied before in relation to their topological structure, patterns of propagation of information, homophily (the tendency of individuals to associate and bond with similar others), and the types of tie formations and decays [9-11]. These frameworks can be applied to characterize medical communities. In a recent study, SNA was applied in an investigation of the network characteristics of the group Health Care Social Media Canada (HCSMCA), particularly as they relate to the formation of an online community [12]. Beyond community formation, an assessment of the network structure can help characterize the actual and potential flow of information between different professional physician groups. The characterization of social activity and information flow is a first step toward understanding the visibility of each network within the online medical community and the potential to transmit health information rapidly and effectively.

Twitter can be thought of as an information sharing network because of its highly skewed distribution of followers, or listeners, and its low rate of reciprocated connections (most information sharers are not followers of their followers) [11]. A number of computer science studies have attempted to characterize how far and how quickly information flows on Twitter [11-13]. This general idea is referred to as information diffusion. It is important to first describe a few models of information diffusion to better understand information diffusion in the context of online medical communities.

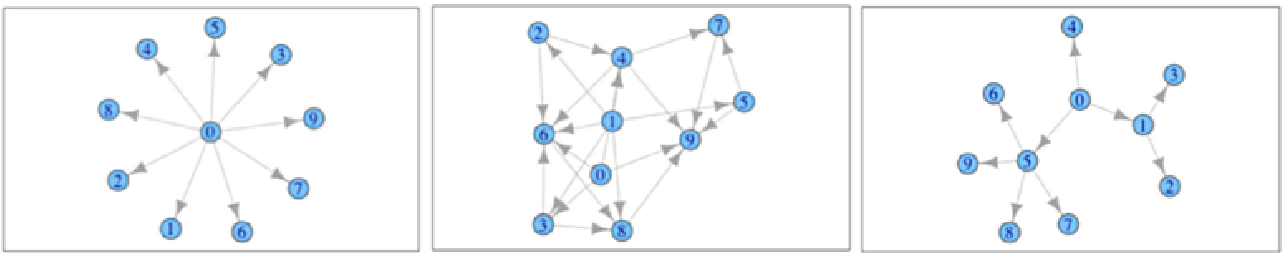
Connectivity Variation Across Network Models

Overview

There are several network topologies (structures or models of the network) that highlight common information diffusion patterns. Figure 1 shows examples of three of these topologies for small directed networks. In each example, the circles are referred to as "nodes" and the lines as "edges".

The number of edges connected to a node is referred to as the degree of the node. The direction of the edge in each network indicates the direction of information flow. The difference between these networks is the degree distribution, ie, the number of incoming and outgoing edges of each node in the network. The pattern of these edge connections defines the structure of the network and dictates how quickly a message travels.

Figure 1. Network configurations - Star, Random, Small World (left to right).



Star Networks

In star networks, the degree distribution of the nodes is heavily skewed. Every time node 0 sends a message, every node in the network receives it immediately. However, when node 8 sends a message, no one ever receives it. This is equivalent to a Twitter subnetwork where a Twitter account or Twitter user has followers who do not connect to each other (they do not have a direct communication channel to each other). When users in a subnetwork are well connected, we say that they form a "cohesive community". The amount of cohesion is defined as the number of common neighbors a group of individuals have, divided by the total number of neighbors. This measure is a variant of a more traditional local “clustering coefficient” measure. A clustering coefficient is user specific and measures the number of triangles a user is involved in. Because connectivity is limited in a star network (no neighbors have edges between them), it does not form a cohesive community. However, for messages sent from the center node, this network is optimal for basic information dissemination since everyone receives the message right away and no extra messages are sent. Messages and ideas sent from the periphery though, do not spread. At the same time, while star structures within a real world network are ideal for quick information diffusion from the center node, they are problematic in terms of network resilience and community development.

Random Networks

In random networks, the degree of the nodes follows a normal distribution. It is unusual to see a node that has a very high or low degree and the dissemination power of each node tends toward the mean. When node 0 sends a message, it takes three steps, or three hops, before it reaches everyone in the network. When node 1 sends a message, it reaches everyone in two hops. If everyone who receives the message sends it forward, some nodes will receive the message multiple times. Since extra messages are being sent, the dissemination is not considered efficient. In other words, because of the random connectivity pattern, spreading a message requires more individuals to participate and is thus not efficient. Further, this network is not a cohesive community since only a small number of neighbors are connected to each other. In general, information networks, social networks, epidemics, and other such networks exhibit non-random connectivity patterns [14].

Small World Networks

In small world networks, the diameter of the graph (the furthest distance between any two nodes in the graph) is low and the amount of cohesion is higher than in a random network. A network in which the degree of the nodes follows a power law distribution indicates a small world network. In this network, a few nodes are very well connected, but most are not. In our example, if node 0 sends a message, it takes two hops to reach everyone without extra messages (in a larger example, we would expect a small number of redundant messages). This network is more efficient than a random network. Well-connected users in this network have comparatively high dissemination power and act as hubs, but messages from the periphery can also be efficiently disseminated because the diameter of small world networks is low. This means that even though most nodes are not neighbors of each other, the number of hops needed to reach every node is small. Small world networks tend to have pockets of cohesive communities throughout the network. Another interesting property of small world networks is that they are more resilient to removal of random nodes from the network than are random networks. Because most random nodes will have a small degree, deleting them will not increase the diameter or decrease the cohesion/clustering coefficient significantly [15].

Many networks have been shown to follow small world properties, including social networks, protein networks, and voter networks. When celebrities are excluded, the degree distribution of nodes on Twitter approximates a power law distribution [14].

Potential Information Dissemination Based on Network Structure

A network with a power law structure has the theoretical capacity to spread information, even arising from the periphery, efficiently to many users. In generated networks of this type, a message can be disseminated to everyone in the network using a simple dissemination strategy and a small number of resends (logarithmic in the network size) [16]. Turning to the specific dynamics of Twitter in which followers observe messages from those they follow and then decide whether or not to retweet them, a common method to capture this behavior is the independent cascade model [17]. In this model, each person resends a message with some fixed independent probability, which captures the likelihood they will find a message interesting. In this setting, for generated power law graphs (which have structures similar to Twitter), once these probabilities pass a certain reasonable threshold, there is a high probability that a message dissemination will become a “long-lived” epidemic [18]. Researchers [19] provided a technique for identifying influential individuals in this model, ie, users who, due to their position in the network, are likely to instigate large information cascades. In their simulations on a real social network topology of a collaboration network among physics researchers, they showed that if a message was resent 10% of the time, termed a “uniform resend probability” of 10%, they could identify a message source that would cause a message to spread to thousands of other users. Kwak et al [13] analyzed message spread on a 2009 snapshot of Twitter containing 41.7 millions user profiles and 106 million tweets and found that over 96% of tweets were not retweeted and that the tweets with the highest dissemination during this period were generally retweeted by 12% to 30% of the sender’s followers. However, even with fewer than 1000 followers, if a message began disseminating quickly, information cascades were much larger than the size of the original follower’s network. This was the case even when the number of initial retweeters was small.

To summarize, computer science theory on information dissemination elicits two things about networks with the properties observed of the Twitter follower graph: (1) Twitter resembles a small world graph with a degree distribution that follows a power law distribution, (2) dissemination to a large number of nodes in a small amount of time is possible, and (3) these large scale disseminations can be achieved with simple resend rules (ie, they do not require sophisticated centralized planning).

It should be mentioned that there is a natural trade-off between information dissemination and community cohesion. If there is high community cohesion, members of the community will have quick access to information. However, members outside of the community will not. In contrast, if cohesion is low, information can disseminate to a broader audience. However, the actual amount of dissemination in a subnetwork without active community participants can be low. For information transmission networks, developing a community with moderate cohesion will increase the dissemination of information to a broader audience.

In this preliminary study, we sought to employ applied graph theory and a basic SNA framework in order to characterize and understand information diffusion on social media within a subset of the medical community by examining the Twitter networks of a few medical professional societies.

Methods

Overview

Social network analysis and network configuration models were used to characterize community structure and information dissemination of four professional physician groups that have a presence on Twitter. The core groups in this analysis are: the American Medical Association (AMA), the American Academy of Family Physicians (AAFP), the American Academy of Pediatrics (AAP), and the American College of Physicians (ACP). Explanations of the metrics used in this study are presented in Table 1.

Table 1. Description of metrics from Twitter.

Metric	What it measures	Description and purpose of metric
Number of followers	Actual information dissemination	How many people/groups received your message? AND How many people/groups may resend (retweet) your tweet? A larger number indicates that a higher probability exists for retweeting the message. Therefore, the potential for large-scale information flow increases.
Number of Level 2 followers	Level 2 information dissemination potential	How many followers (active listeners) do your followers have? This number represents the number of Level 2 followers who will see your tweet if all of your followers retweet it and none of your followers have the same Level 2 followers.
Dissemination network size	Information dissemination potential	How many people can see your message if all of your followers retweet it? This number represents the followers and the Level 2 followers who see your tweet if all of your followers retweet the message and none of your followers have the same Level 2 followers.
Number of information sharers	Active sources of information	Who are the other people or groups on Twitter that you are getting information from? A larger number indicates more information sources to retweet messages from. These are active sources because they send you information.
Number of tweets	Frequency of information disseminated	How often do you share information with your followers? A large number indicates that you regularly post tweets/re-tweets for your followers to view.
Number of retweeters	Actual number of information disseminators	How many people retweeted a particular tweet you sent? A large number indicates that many people shared your tweet with their followers.
Retweeter network size	Number of Level 2 followers	How many people receive the tweet when some of the followers retweet it?

Network Characterization

The Twitter API (api.twitter.com) was used to characterize the network: determine the number of followers and the number of tweets for each of these groups. These data were collected between July 2012 and September 2012. Accounts that were disabled, private, or not recognizable by our automated programs were ignored. This amounted to less than 1% (1257/238,853) of the accounts we had access to. Similar to other studies in computer science, the number of followers was used as an indicator of actual information dissemination, the Level 2 followers as an indicator of the information dissemination potential, the number a user is following as a way to identify potential sources of information, and the number of tweets as an indicator of the frequency of information dissemination for a particular group or individual.

We approximated community cohesion for this dataset by measuring the amount of overlap in followers between the professional groups as a percentage [(A intersect B)/min(|A|,|B|)*100]. Overlap is necessary to develop a cohesive community that has common sets of followers and connections between subsets of the followers. As previously mentioned, too much overlap may reduce the amount of information that disseminates outside of the community. We also used visualization to better understand this overlap.

Information Flow

The actual information flow of the tweets for each of these groups for a one-month period from August 1, 2012 to September 1, 2012 was analyzed. Tweets sent by the four professional physician groups were examined using Topsy, a Twitter data aggregator [20], to determine how many times each tweet was retweeted. Topsy is a Twitter partner that has indexed all public tweets since Twitter was founded in 2006. Of note, since this data is dynamic, more retweets can occur after the data has been collected.

Tweet Propagation Analysis

For each group, tweets sent between July 1, 2012 and September 12, 2012 were identified and assessed, including the number of retweets and details of their dissemination. This was followed by the identification of individuals who retweeted the message, the determination of the number of followers for each retweeter and the computation of the retweeter network size to measure actual information flow. Looking at how many times a message was retweeted, compared to the number of times it could have been retweeted if all of an organization's followers retweeted the message, provides a measurement of how close the actual tweet dissemination is compared to the theoretical best.

Results

Network Characterization

Table 2 shows the number of followers, the number a user is following, the number of tweets, and the information dissemination potential for each professional medical group in this analysis. These statistics show that all of these professional groups have thousands of followers and a theoretical information dissemination potential ranging from 6.9 to 122 million people. This value is the sum of each professional group's followers and Level 2 followers. For example, if all the followers of the AAFP retweet a message sent by the AAFP, 6,959,092 people will see the message. Given the large potential for dissemination, we can view each of these professional groups as information brokers within the Twittersphere.

The information dissemination potential for the followers of each group was plotted using a cumulative frequency graph in Figure 2. Each line represents one organization and shows cumulatively what percentage of their followers has a given size follower network. Of note, the graph is a log-scale. For example, over 25% of AMA followers have fewer than 10 individuals following them. We see that the distribution of followers for these subnetworks is consistent with a power law distribution, with approximately half the nodes having fewer than 100 followers and very few nodes having hundreds of thousands of followers.

In AAFP and ACP, it emerged that over half of the followers have a strong listener network (Level 2 follower network) with at least 100 listeners. The median number of listeners for each of the followers of AAFP, ACP, and AAP are 120, 165, and 81 respectively. In contrast, the majority of followers of AMA have smaller (quieter) listener networks, with over 50% (119,560/213,122) having fewer than 50 listeners. The potential to disseminate widely exists for each of these organizations since all of the organizations have a large percentage of listeners who are themselves information brokers.

A detailed social network visualization of the networks' overlap is shown in Figure 3. This visualization only shows the three smaller physicians group networks since the largest one, the AMA, has approximately ten times the number of followers as the other three networks combined.

In the networks under study, there is evidence of the beginnings of a collective community, where large percentages of individuals (13%-55%) are following more than one of the professional groups. As the illustration shows, the majority of followers are specific to one of the groups. The pink nodes are following all three of the professional groups, while the green, orange, and purple nodes are following two of the professional groups. The overall common overlap across all four groups is only 471 individuals, a very small percentage of the overall networks for these groups, indicating a limited amount of community cohesion and cross-fertilization, but still allowing for efficient channels (fewer redundant messages) for information dissemination.

Table 2. Dissemination potential and professional group statistics during the study period.

Professional group	Number of followers	Number following	Number of tweets	Information dissemination potential
American Academy of Family Practice (AAFP) (@AAFP)	7546	298	2788	6,959,092
American College Physicians (ACP) (@ACPinternists)	5955	2023	2979	11,228,160
American Academy of Pediatrics (AAP) (@AmerAcadPed)	11,768	132	1184	14,496,559
American Medical Association (AMA) (@AmerMedicalAssn)	213,122	5729	7065	122,066,397

Figure 2. Information dissemination potential for each professional physicians group - American Medical Association (AMA), American Academy of Family Physicians (AAFP), American Academy of Pediatrics (AAP), and American College of Physicians (ACP).

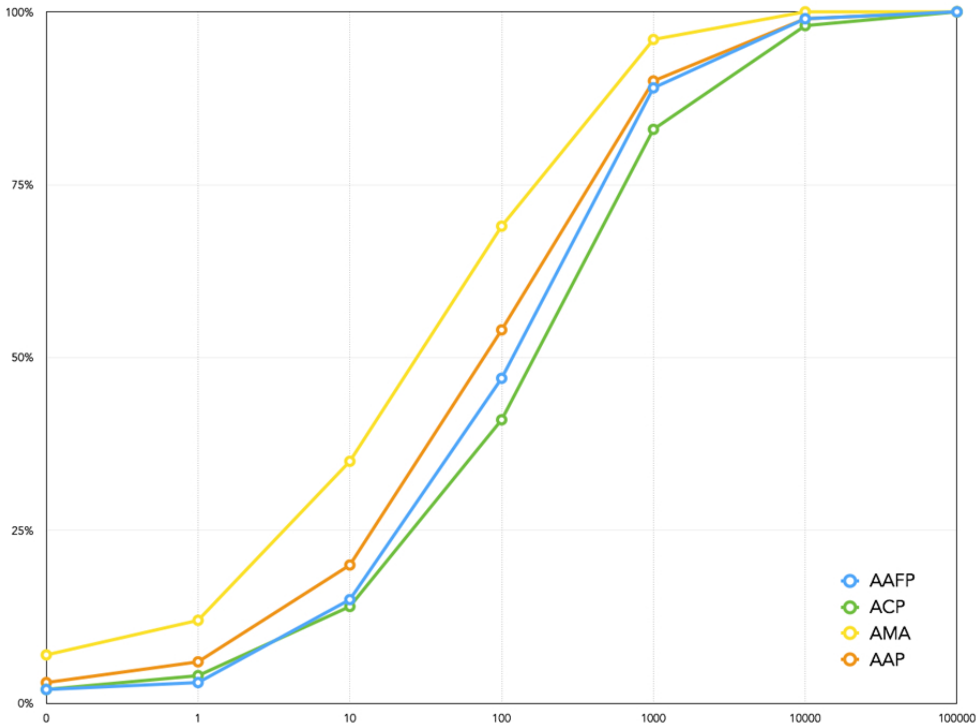
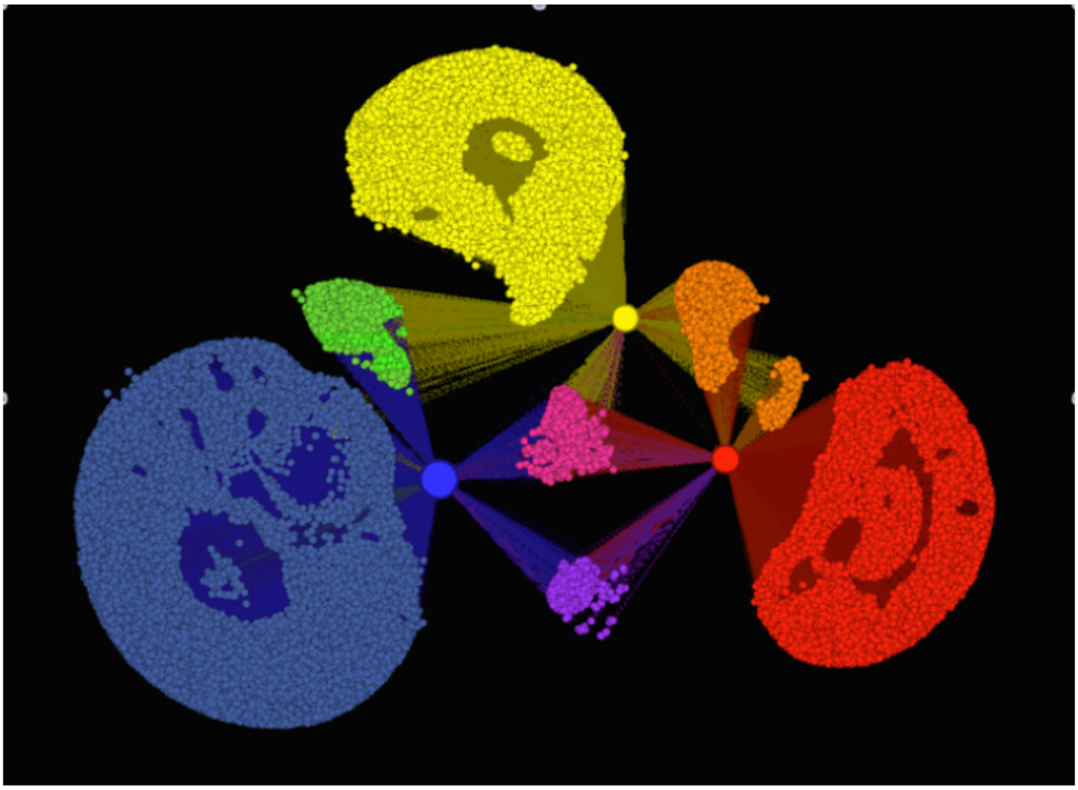


Figure 3. Follower network of American Academy of Family Physicians (yellow), American College of Physicians (red), and American Academy of Pediatrics (blue). Size of group nodes based on number of followers.

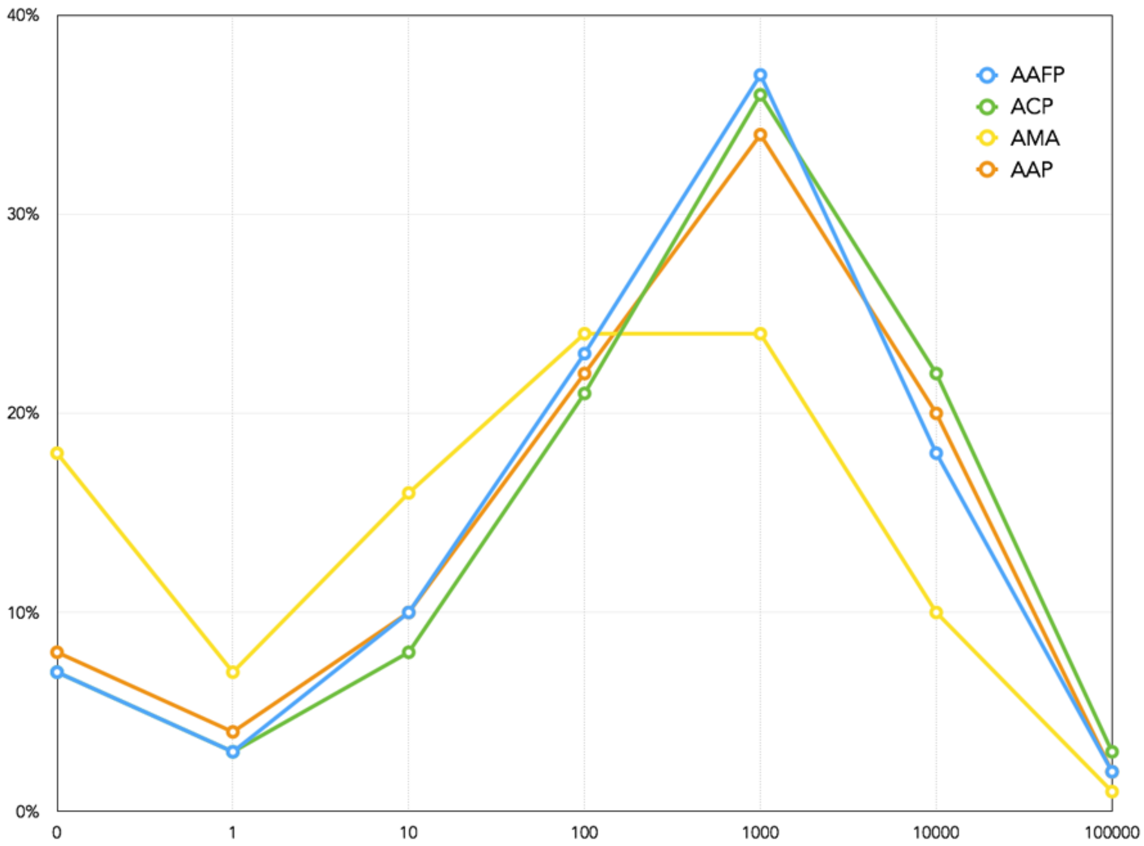


Information Flow

When considering information diffusion potential, it is reasonable to exclude followers who have never tweeted or retweeted, since it is likely that those individuals will not retweet a message from one of the professional groups. The more tweets a group's followers send, the higher the likelihood for larger dissemination. Figure 4 shows a continuous histogram of the volume of tweets of followers for each of our professional groups. The x-axis shows logarithmically the volume of tweets sent by followers of each professional group. The y-axis shows the percentage of followers who have sent that volume of tweets. For example, the percentage of followers that have not sent any tweets is 6.92% for AAFP (522/7546), 8.17% (962/11,768) for AAP, 7.22% (430/5955) for ACP, and 18.43% (39,275/213,122) for AMA.

If these followers and their Level 2 followers are removed from the dissemination network, the overall information dissemination potential decreases by less than 1% for all four professional groups. This indicates that the followers who do not send any tweets/retweets have a small number of followers themselves and are not essential information brokers. When all the followers who have sent only 10 or fewer tweets are removed, then the AMA professional group information dissemination potential is reduced by over 35%. The other professional groups are still impacted by less than 1%. This is an indication that the AMA followers' network is not as active as the other three professional networks. For the other three groups, there is a stronger correlation between the number of tweets disseminated and the number of followers.

Figure 4. Number of tweets/retweets sent by followers of the four professional groups - American Medical Association (AMA), American Academy of Family Physicians (AAFP), American Academy of Pediatrics (AAP), and American College of Physicians (ACP).



Tweet Propagation Analysis

In addition to the information dissemination potential, the actual retweet propagation of a sample of tweets was assessed. Figure 5 shows the propagation of actual tweets during the month of August 2012 (ie, the number of retweets for each message sent by the different professional medical associations.)

The x-axis represents each tweet where the tweets are sorted by number of retweets. The y-axis represents the number of retweets. The AMA posted the largest number of tweets (164), while the AAP posted the fewest during this time period. Each organization had a number of tweets that were not retweeted by anyone. The largest number of retweets for any of these organizations during this month was 24. Given that each of these groups has thousands of followers, this level of retweeting leads to information dissemination that is far below the information dissemination potential shown in Table 2. While unlikely, even with a small number of followers retweeting a message, the diffusion can still be large at the third and fourth hop. We have not computed the third and fourth hop networks here, but previous literature supports this pattern of diffusion; Bakshy et al [21] analyzed a data set from 2009 containing 1.6 million Twitter users in 2009 and identified common information cascade patterns of their tweets. Many of these patterns involved transmission with third and fourth hop users. Therefore, we cannot discount that dissemination occurs beyond the Level 2 followers.

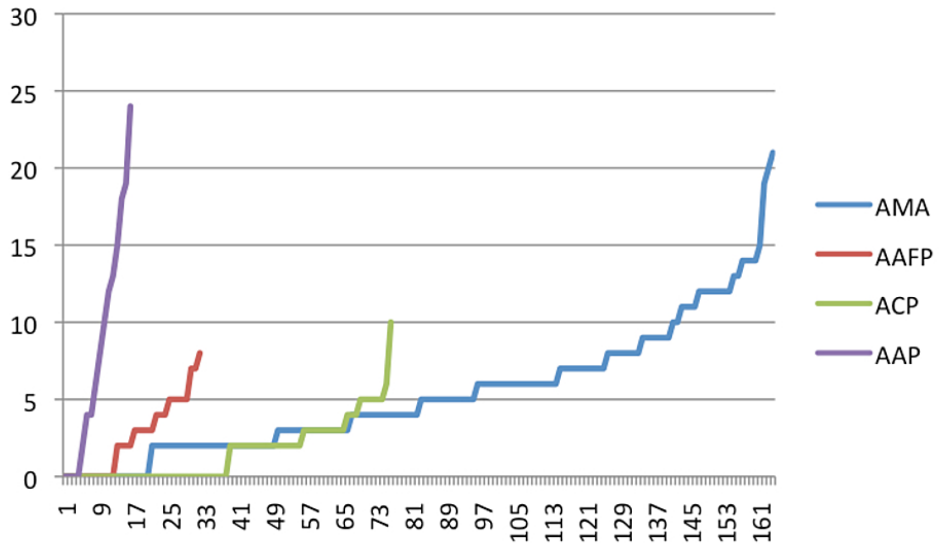
Finally, the dissemination of a particular tweet was considered: how does the dissemination of an actual tweet compare to the theoretic best? Here, we focus on the propagation of the tweet as opposed to the content of the tweet. Are there any tweets that are disseminating to a large fraction of this medical community? Table 3 compares the tweet with the highest retweet dissemination (actual information dissemination) for tweets sent between July 1, 2012 and September 12, 2012 to the dissemination potential for each professional group.

Overall, the number of retweets and the number of individuals who received the tweet is less than 0.2% of the total population dissemination potential, with the tweet from the ACP disseminating the least.

Table 3. Top tweets for each professional group.

Professional group	Tweet	Number of retweets	Actual information dissemination	Fraction of information dissemination potential
AAFP	“Ask your Doctor if medical advice from a TV commercial is right for you…”	10	9558	0.00137
ACP	“Interaction between proton-pump inhibitors clopidogrel clinically unimportant…”	7	489	0.000044
AAP	“Tragedy in CO – in the wake of news about another act of gun violence, how to talk with children and teens…”	25	25,482	0.00176
AMA	“September is Women in Medicine Month, a time to celebrate growing number, influence of women physicians	45	200,778	0.00164

Figure 5. Number of retweets for messages sent in August 2012. [American Medical Association (AMA), American Academy of Family Physicians (AAFP), American Academy of Pediatrics (AAP), and American College of Physicians (ACP)].



Discussion

Principal Findings

At the time of our study, the AMA had the largest number of followers—and thus, information diffusion potential—and was trailed by the AAP, AAFP, and ACP, respectively. However, each of the smaller organizations had a strong network of followers among which were individuals who themselves are potentially strong information brokers. We also began to see interconnectedness among these groups as evidenced by a group of users who follow all three smaller organizations. This preliminary analysis shows possibly large information diffusion potential, yet when we analyzed actual tweets sent, the actual dissemination was well below the calculated potential.

With the growing popularity of social media and Twitter, medical organizations are urged to engage in social media and actively share information. Therefore, it is important to determine what metrics can be used to measure the effectiveness of this as a medium. This study attempted to describe the characteristics of four medical networks and analyze their theoretical information dissemination potential, their actual information dissemination, their information sharers, and their propagation and distribution of tweets.

Limitations

This study has several weaknesses. First, we captured our data at one point in time. As such, it is only a snapshot of the Twitter networks described. Social media networks tend to be dynamic with followers added and dropped from moment to moment. So, in all likelihood, these networks may look different today than they did during the study period. Additionally, the overall trend on Twitter is expansion, with increased number of users. In fact, at the time of manuscript revision submission, each of the organizations described in this manuscript has shown to have a much larger following. This also applies to the data captured for the individual tweets, which may have diffused beyond the study period. Second, our analysis does not allow for an investigation of the inter-activity, interactions, and engagement within the networks. As a result, it is impossible to draw conclusions about motivation for dissemination or how content drives diffusion. Third, though the total percentage of accounts we ignored in our analysis was less than 1%, we could not determine the percentage of private accounts that we did not have access to. According to Beevolve [22], in 2012, approximately 12% of Twitter users had private or protected accounts. Finally and most importantly, the dissemination potential described should be interpreted with caution. The number of followers may not necessarily mean each follower is a valid or relevant one, or one that can further propagate the message. Many followers may be family members, friends, commercial entities, organizations, or individuals with no interest in the topic. Further studies will need to look at the specific nature and identity of the followers. Additionally, because information is disseminated on Twitter in a continuous manner, there is no guarantee that all the followers will receive or will have seen all the messages, as patterns of use vary and many followers check their accounts in a more sporadic manner, thus “missing” many tweets in real time (and, as a result, the opportunity to retweet or act on the message in other ways).

This work is merely the first step toward understanding the information power and potential of several medical professional groups on Twitter.

Conclusions

This analysis indicates that these medical groups participate in subnetworks with small world type tendencies. This structure allows for large-scale information dissemination; however, actual dissemination is well below potential for all four professional groups. This is consistent with many other groups on Twitter. Large-scale information diffusion in Twitter is driven by information brokers who have at least a moderate number of followers, some of whom are active followers. In other words, it is more valuable to a network to be well connected to a few influential information brokers than to have a large number of first degree followers. Although having a large number of followers is beneficial, small networks can still achieve a high potential distribution if they have a few information brokers who themselves are active and well connected. Developing a community that is active (in terms of retweeting) and engaged (in terms of content, mutuality, and reciprocity) is important for strong dissemination. As demonstrated in a previous study [23], social media engagement is a complex system to describe and measure. Engagement may be thought of as a continuum (low-medium-high) and may follow specific patterns and hierarchical structures. Indeed, one metric of engagement, as described in Neiger et al, is retweeting. While these professional groups all have some individuals who retweet heavily, those individuals were not very engaged during our study period. Encouraging these information brokers to retweet the shared content will help increase dissemination. Other ways to increase dissemination include tweeting more regularly (increasing the volume of tweets to show active engagement in the Twitter platform) and retweeting more often. Others will be less interested in retweeting content if the organizations themselves do not retweet content. These strategies will help develop a more cohesive community of shared followers, for cross-fertilization of information.

The content of the messages is of course of utmost importance. Even with strong channels for dissemination, tweets must be timely and engaging in order to provide the hook for followers to retweet and begin to reach the vast potential audience.

In the past few years, reports have been published about the use of Twitter for various purposes within the field of medicine: as a support tool for patients with chronic conditions [24], by ministries of health for health promotion [25], as an educational tool during conferences [26], to track public health message diffusion [27], and as a teaching tool in health profession education [28-34]. A few researchers are developing metrics that may enable this medium to assist in measuring journal article impact [35] and to predict citations of journal articles [36]. However, as a field of research, Twitter usage—the flow of information, content analysis, profiles of message generators and of message recipients, overall effects on public behavior—is still in its infancy.

This study is one example of the development of theoretical models of knowledge dissemination that could have practical implications in how we use this medium to empower patients, disseminate important public health messages, or promote our ideas and specialties. As researchers attempt to characterize best practices in the use of social networks for knowledge transfer and dissemination, they will need to look at the networks themselves, conduct content analysis of the messages, and assess the behavior and actions of the messengers as well as the recipients. This calls for more multi-disciplinary research, involving experts in computer science, communications, linguistics, and cultural studies, to develop and advance this field of inquiry.

Conflicts of Interest

None declared.

References

1. Goldenberg J, Libai B, Muller E. STFM. 2012 Feb 15. Engaging the family medicine community through social media URL: http://blog.stfm.org/2012/02/15/engaging-the-family-medicine-community-through-social-media [accessed 2013-10-07] [WebCite Cache ID 6OdtKMwsO]
2. Sevilla M. Healthcare Communication News. 2011 May 12. Why family medicine needs social media URL: http://www.healthcarecommunication.com/Physicians/Articles/_Why_family_medicine_needs_social_media_6774.aspx [accessed 2013-10-07] [WebCite Cache ID 6KBVFuWt9]

3. Bottles K. Twitter: an essential tool for every physician leader. Physician Exec 2011;37(3):80-82. [Medline: 21675319]
4. Chasin L, Pho K. KevinMD. 2012 Jun. Social media is crucial to the development of medicine today URL: http://www.kevinmd.com/blog/2012/06/social-media-crucial-development-medicine-today.html [accessed 2013-10-07] [WebCite Cache ID 6KBVOwu9T]
5. Blau A. VentureBeat. 2012 Apr 26. Improving modern medicine: why social media is just what the doctor ordered URL: http://venturebeat.com/2012/04/26/social-media-medicine/ [accessed 2013-10-07] [WebCite Cache ID 6KBVedWJ0]
6. Chou WY, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
7. 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]
8. Chou WY, Prestin A, Lyons C, Wen KY. Web 2.0 for health promotion: reviewing the current evidence. Am J Public Health 2013 Jan;103(1):e9-18. [doi: 10.2105/AJPH.2012.301071] [Medline: 23153164]
9. Mislove A, Lehmann S, Ahn YY, Onnela JP, Rosenquist JN. Understanding the demographics of Twitter users. : Association for the Advancement of Artificial Intelligence; 2011 Presented at: Fifth International AAAI Conference on Weblogs and Social Media; July 17-21, 2011; Barcelona, Spain URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/download/2816/3234
10. Grabowicz PA, Ramasco JJ, Moro E, Pujol JM, Eguiluz VM. Social features of online networks: the strength of intermediary ties in online social media. PLoS One 2012;7(1):e29358 [FREE Full text] [doi: 10.1371/journal.pone.0029358] [Medline: 22247773]
11. Bakshy E, Rosen I, Marlow C, Adamic L. The role of social networks in information diffusion. New York, NY: Assn for Computing Machinery; 2012 Apr Presented at: 21st International Conference on World Wide Web; April 16-20, 2012; Lyon, France p. 519-528 URL: http://dl.acm.org/citation.cfm?id=2187836 [doi: 10.1145/2187836.2187907]
12. Gruzd A, Haythornthwaite C. Enabling community through social media. J Med Internet Res 2013 Oct;15(10):e248 [FREE Full text] [doi: 10.2196/jmir.2796] [Medline: 24176835]
13. Kwak H, Lee C, Park H, Moon S. What is Twitter, a social network or a news media? New York, NY: Assn for Computing Machinery; 2003 Presented at: 19th International Conference on World Wide Web; April 26-30, 2010; Raleigh, NC, USA URL: http://dl.acm.org/citation.cfm?doid=1772690.1772751 [doi: 10.1145/1772690.1772751]
14. Kempe D, Kleinberg J. Who says what to whom on Twitter? New York: Assn for Computing Machinery; 2003 Presented at: 20th International Conference on World Wide Web; 2011-03-28; Lyon, France p. 137-146. [doi: 10.1145/1963405.1963504]
15. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature 1998 Jun 4;393(6684):440-442. [doi: 10.1038/30918] [Medline: 9623998]
16. Elsässer R, Ogierman A. Efficient broadcasting in random power law networks. In: Thilikos DM, editor. Graph-Theoretic Concepts in Computer Science: 36th International Workshop. New York: Springer; 2010:279-291.
17. Goldenberg J, Libai B, Muller E. Talk of the network: a complex systems look at the underlying process of word-of-mouth. Marketing Letters 2001;12(3):211-223. [doi: 10.1023/A:1011122126881]
18. Ganesh A, Massoulié L, Towsley D. The effect of network topology on the spread of epidemics. Piscataway, N.J: IEEE; 2005 Presented at: 24th Annual Joint Conference of the IEEE Computer and Communications Societies; March 13 -17, 2005; Miami, Florida. [doi: 10.1109/INFCOM.2005.1498374]
19. Kempe D, Kleinberg J, Tardos E. Maximizing the spread of influence through a social network. New York: Assn for Computing Machinery; 2003 Presented at: 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 3, 2003; Washington, DC.
20. Topsy Website. URL: http://www.topsy.com [accessed 2014-04-09] [WebCite Cache ID 6Oi2CawlE]
21. Bakshy E, Hofman JM, Mason WA, Watts DJ. Everyone's an influencer: quantifying influence on Twitter. 2011 Presented at: 4th ACM International Conference on Web Search and Data Mining (WSDM '11); Feb. 9, 2011; Hong Kong p. 65-74. [doi: 10.1145/1935826.1935845]
22. Beevolve Website. URL: http://www.beevolve.com/twitter-statistics/ [accessed 2014-04-09] [WebCite Cache ID 6Oi2PTc0y]
23. Neiger BL, Thackeray R, Burton SH, Giraud-Carrier CG, Fagen MC. Evaluating social media's capacity to develop engaged audiences in health promotion settings: use of Twitter metrics as a case study. Health Promot Pract 2013 Mar;14(2):157-162. [doi: 10.1177/1524839912469378] [Medline: 23271716]
24. De la Torre-Díez I, Díaz-Pernas FJ, Antón-Rodríguez M. A content analysis of chronic diseases social groups on Facebook and Twitter. Telemed J E Health 2012;18(6):404-408. [doi: 10.1089/tmj.2011.0227] [Medline: 22650380]
25. Carrillo-Larco RM. Social networks and public health: use of Twitter by ministries of health. Int J Public Health 2012 Aug;57(4):755-756. [doi: 10.1007/s00038-012-0387-4] [Medline: 22735993]
26. McKendrick DR. Smartphones, Twitter and new learning opportunities at anaesthetic conferences. Anaesthesia 2012 Apr;67(4):438-439. [doi: 10.1111/j.1365-2044.2012.07068.x] [Medline: 22409809]
27. Ayers JW, Althouse BM, Allem JP, Ford DE, Ribisl KM, Cohen JE. A novel evaluation of World No Tobacco day in Latin America. J Med Internet Res 2012;14(3):e77 [FREE Full text] [doi: 10.2196/jmir.2148] [Medline: 22634568]

28. Bahner DP, Adkins E, Patel N, Donley C, Nagel R, Kman NE. How we use social media to supplement a novel curriculum in medical education. Med Teach 2012;34(6):439-444. [doi: 10.3109/0142159X.2012.668245] [Medline: 22449268]
29. Mistry V. Critical care training: using Twitter as a teaching tool. Br J Nurs 2011;20(20):1292-1296. [Medline: 22068003]
30. George DR, Dellasega C. Use of social media in graduate-level medical humanities education: two pilot studies from Penn State College of Medicine. Med Teach 2011;33(8):e429-e434. [doi: 10.3109/0142159X.2011.586749] [Medline: 21774639]
31. Fox BI, Varadarajan R. Use of Twitter to encourage interaction in a multi-campus pharmacy management course. Am J Pharm Educ 2011 Jun 10;75(5):88 [FREE Full text] [doi: 10.5688/ajpe75588] [Medline: 21829262]
32. Trueman MS, Miles DG. Twitter in the classroom: twenty-first century flash cards. Nurse Educ 2011;36(5):183-186. [doi: 10.1097/NNE.0b013e3182297a07] [Medline: 21857333]
33. Skiba DJ. Nursing education 2.0: Twitter & tweets. Can you post a nugget of knowledge in 140 characters or less? Nurs Educ Perspect 2008;29(2):110-112. [Medline: 18459627]
34. Forgie SE, Duff JP, Ross S. Twelve tips for using Twitter as a learning tool in medical education. Med Teach 2013;35(1):8-14. [doi: 10.3109/0142159X.2012.746448] [Medline: 23259608]
35. Evans P, Krauthammer M. Exploring the use of social media to measure journal article impact. In: AMIA Annu Symp Proc. 2011 Presented at: AMIA Annual Symposium; October 22, 2011; Washington, DC p. 374-381 URL: http://europepmc.org/abstract/MED/22195090
36. Eysenbach G. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. J Med Internet Res 2011;13(4):e123 [FREE Full text] [doi: 10.2196/jmir.2012] [Medline: 22173204]

Abbreviations

AAFP: American Academy of Family Physicians
AAP: American Academy of Pediatrics
ACP: American College of Physicians
AMA: American Medical Association
SNA: social network analysis

Edited by G Eysenbach; submitted 07.10.13; peer-reviewed by B McGowan, A Gaysynsky, C Giraud-Carrier; comments to author 11.11.13; revised version received 03.01.14; accepted 20.02.14; published 14.04.14.

Please cite as:
Mishori R, Singh LO, Levy B, Newport C
Mapping Physician Twitter Networks: Describing How They Work as a First Step in Understanding Connectivity, Information Flow, and Message Diffusion
J Med Internet Res 2014;16(4):e107
URL: http://www.jmir.org/2014/4/e107/
doi:10.2196/jmir.3006
PMID:24733146

Original Paper

Still a Hard-to-Reach Population? Using Social Media to Recruit Latino Gay Couples for an HIV Intervention Adaptation Study

Omar Martinez[1], JD, MPH, MSc; Elwin Wu[2], PhD; Andrew Z Shultz[1], MA; Jonathan Capote[3], BA; Javier López Rios[1], BA; Theo Sandfort[1], PhD; Justin Manusov[2], BA; Hugo Ovejero[4], JD; Alex Carballo-Dieguez[1], PhD; Silvia Chavez Baray[5], PhD; Eva Moya[5], PhD; Jonathan López Matos[6], BA; Juan J DelaCruz[7], PhD; Robert H Remien[1], PhD; Scott D Rhodes[8], PhD

[1]Columbia University, HIV Center for Clinical and Behavioral Studies at the New York State Psychiatric Institute and Columbia University, New York, NY, United States

[2]Columbia University, School of Social Work, New York, NY, United States

[3]AID for AIDS International, New York, NY, United States

[4]Latino Commission on AIDS, New York, NY, United States

[5]University of Texas at El Paso, Department of Social Work, El Paso, TX, United States

[6]The City University of New York, Center for HIV Educational Studies & Training, New York, NY, United States

[7]The City University of New York, Department of Business and Economics - Lehman College, New York, NY, United States

[8]Wake Forest University School of Medicine, Winston-Salem, NC, United States

Corresponding Author:
Omar Martinez, JD, MPH, MSc
Columbia University
HIV Center for Clinical and Behavioral Studies at the New York State Psychiatric Institute and Columbia University
1051 Riverside Drive, Unit 15
New York, NY, 10032
United States
Phone: 1 646 774 6924
Fax: 1 212 543 6003
Email: om2222@columbia.edu

Abstract

Background: Online social networking use has increased rapidly among African American and Latino men who have sex with men (MSM), making it important to understand how these technologies can be used to reach, retain, and maintain individuals in care and promote health wellness. In particular, the Internet is increasingly recognized as a platform for health communication and education. However, little is known about how primarily Spanish-speaking populations use and engage with each other through social media platforms.

Objective: We aimed to recruit eligible couples for a study to adapt “Connect ‘n Unite” (an HIV prevention intervention initially created for black gay couples) for Spanish-speaking Latino gay couples living in New York City.

Methods: In order to successfully design and implement an effective social media recruitment campaign to reach Spanish-speaking Latino gay couples for our ongoing “Latinos en Pareja” study, our community stakeholders and research team used McGuire’s communication/persuasion matrix. The matrix guided our research, specifically each marketing “channel”, targeted “message”, and target population or “receiver”. We developed a social media recruitment protocol and trained our research staff and stakeholders to conduct social media recruitment.

Results: As a result, in just 1 month, we recruited all of our subjects (N=14 couples, that is, N=28 participants) and reached more than 35,658 participants through different channels. One of the major successes of our social media recruitment campaign was to build a strong stakeholder base that became involved early on in all aspects of the research process—from pilot study writing and development to recruitment and retention. In addition, the variety of “messages” used across different social media platforms (including Facebook, the “Latinos en Pareja” study website, Craigslist, and various smartphone applications such as Grindr, SCRUFF, and Jack’d) helped recruit Latino gay couples. We also relied on a wide range of community-based organizations across New York City to promote the study and build in the social media components.

Conclusions: Our findings highlight the importance of incorporating communication technologies into the recruitment and engagement of participants in HIV interventions. Particularly, the success of our social media recruitment strategy with Spanish-speaking Latino MSM shows that this population is not particularly “hard to reach”, as it is often characterized within public health literature.

(J Med Internet Res 2014;16(4):e113) doi:10.2196/jmir.3311

KEYWORDS

social media; online recruitment strategies; Spanish-speaking Latino men who have sex with men (MSM); Latino gay couples; Latino MSM; HIV prevention

Introduction

In the United States, online social networking features are widely used, as reflected in the Health Online 2013 survey conducted by the Pew Research Center’s Internet and American Life Project. Their report noted, for example, that more than one-third of US individuals turn to the Internet to learn about health problems [1]. Moreover, although it is difficult to gather exact figures, it is estimated that somewhere between 20 million and 40 million people in the United States visit online dating websites each month [2]. An online social network is a website or online application that allows individuals and communities to connect and communicate by sharing pictures, messages, and other forms of multimedia communication [3]. However, new media technologies are constantly emerging along with new definitions of what constitutes a social network, making it challenging to maintain a rigid definition capable of encompassing all of these rapid changes.

The rise of new communication technologies and social media in recent years has brought both new opportunities and challenges for public health professionals. Peer-to-peer exchanges through websites, online communities, and smartphone applications represent a major shift in the way people meet potential sexual partners, especially for men who have sex with men (MSM) [4-7]. Mobile accessibility through popular applications like Grindr, SCRUFF, Jack’d, and OkCupid offer new ways for MSM to engage in peer-to-peer exchanges, whether for gaining support and sharing information with fellow members of their community or to find sexual and/or romantic partners. At the same time, there has been a dramatic increase in the availability of information promoting healthy behaviors and advertising human immunodeficiency virus (HIV) and acquired immunodeficiency syndrome (AIDS) resources on the Internet, although evaluation of such strategies is often lacking [8-10].

Most of the research on social media has been focused on “risk behaviors” [4,11-15] and developing new HIV prevention and health promotion “media interventions” [16-21]. Our research responds to a key component of the research and program engagement that includes a discussion of recruitment strategies and methods for engaging subjects and participants in research and prevention programs. This particular area of research merits attention as Institutional Review Boards across the country try to comply with ethical standards and protection of participants in both online and offline contexts [22].

In particular, online social networking usage has increased rapidly among African American and Latino MSM [16,17,23], making it important to understand how these technologies could be used to reach, retain, and maintain individuals in research and medical care [24,25]. While most research on the use of social media among minorities has focused on African American and acculturated Latino MSM [26,27], little is known about those who are less acculturated, monolingual, and/or primarily Spanish-speaking Latino MSM. Interestingly, this particular group has been previously labeled as “hard to reach” by scholars and researchers [28-30]. Thus, in order to better recruit and engage this community, we assessed and implemented a recruitment campaign to enroll Latino gay couples in an HIV intervention adaptation study using both social media platforms as well as traditional approaches with community-based organizations. The goal of the study is to adapt “Connect ‘n Unite” (CNU), a couple-based intervention for stimulant-using black MSM, to reduce the disproportionate HIV burden borne by Latino MSM. CNU has been developed, implanted, and evaluated as a 4-module couple-based HIV preventive intervention. The focus of our paper will center on the need to better understand the Latino gay community, its engagement with social media, and the potential for effective delivery of wellness promotion interventions and HIV prevention programs through these media.

Methods

Summary

The Institutional Review Board at the New York State Psychiatric Institute approved this study. We undertook a mixed method approach to conduct our research. For the purpose of this paper, data analysis is restricted to the first month of the study implementation and recruitment, during which time we recruited our total sample size (N=14 couples). We relied on community stakeholders to help guide recruitment of participants through social media, from the development of materials to the wording and message of postings. A total of 12 stakeholders from a wide range of community-based organizations and Latino gay couples met weekly to develop social media recruitment materials. The principal investigator (PI) and research assistant took notes at all of the meetings (Figure 1). Stakeholders were introduced to methodological approaches in order to guide the development of social media recruitment materials. We also used McGuire’s communication/persuasion matrix as a model for understanding and creating an effective recruitment campaign [31]. Other scholars have used this model to guide evaluation of recruitment strategies for HIV prevention and wellness programs [32]. The model identifies the variables relating to effective persuasive communication, including receiver, channel, message, and source. Each of these elements is further defined and discussed below in relation to our “Latinos en Pareja” study.

Figure 1. Collection of photos from a stakeholder meeting.

Receivers

The receiver is the intended recipient of a message. For this study, the goal was to recruit 14 Spanish-speaking Latino gay couples to participate in an intervention adaptation workshop, composed of three sessions. We aimed to adapt CNU, a couple-based intervention created for stimulant-using black MSM, to reduce the disproportionate HIV burden borne by Latino MSM. CNU has been developed, implemented, and evaluated as a 4-module couple-based HIV preventive intervention [33,34]. The modules cover several topics including self-care (eg, information about HIV/AIDS, stimulant use), communication (eg, use of effective communication styles), relationship strengthening (eg, identification of unwritten rules and sexual decision making), and couple problem solving (eg, identification support mechanism for each partner). Couples were eligible for the “Latinos en Pareja” study if they met the following criteria: (1) both partners were 18 years or older, (2) both partners considered the other male as their “main partner”, which is operationalized as (a) a male with whom he has had an ongoing sexual relationship over the prior 3 months, (b) a male considered a “boyfriend, domestic partner, spouse, ongoing lover, or regular partner”, and (c) a stated intention to remain together for at least 12 months, (3) at least one partner self-identified as Latino or Hispanic (ie, a native or inhabitant of Latin America; a person of Latin American origin living in the United States), (4) at least one partner had limited English proficiency and both partners were proficient in Spanish, (5) at least one partner reported one or more unprotected acts of anal intercourse in the past year, within or outside of the relationship, and (6) at least one partner reported using illicit substances or other drugs/substances not prescribed by a doctor that change mood or thinking in the past 3 months, or drinking more than 4 drinks in a single period or 14 drinks per week in the past 3 months. Couples were excluded if either partner reported the occurrence of one or more incidents of severe intimate partner violence within the relationship in the past year. Participants called our research telephone number to find out whether they were eligible through the phone screening.

Channels

Most of the participants were recruited using social media. We diversified our social media channels by reaching a wide range of social network sites, including Craigslist and Facebook as well as iPhone and other smartphone applications including Jack’d, Grindr, Twitter, Instagram, and SCRUFF (Figure 2). We created study profiles on the popular dating websites OkCupid and Adam4Adam, but site administrators quickly deleted the profiles since they promoted research. Therefore, we decided not to create new profiles on these sites even though Adam4Adam has been known as a site frequently visited by gay, bi, MSM, trans, and other sexual minority men of color. We also developed a website using Wix, a free website builder tool [35].

Along with recruiting participants through social media platforms, research staff who were involved in the recruitment and the adaptation of the study focused on other channels to distribute recruitment materials including local barbershops visited by gay Latinos, community-based organizations and social venues, press releases and public service announcements published through local media, and the deployment of outreach recruiters in gay venues (eg, bars, bathhouses, community events, and local support groups aimed at gay men). We also shared on our social media sites information about these venues as well as the services they offered (Figure 3). However, although we reached more than 50 community-based organizations in New York City, most of the participants came from the Hispanic AIDS Forum, Betances Health Center, Latino Commission on AIDS, and Make the Road New York.

Figure 2. Collection of photos from our online social media sites.

Latinos En Pareja shared **Latinos En Pareja**'s photo.
Yesterday at 6:26 PM

Promoviendo relaciones saludables entre hombres latinos y sus parejas del mismo sexo.
Llámenos ya: (646) 322-7918
Email: latinosenpareja@gmail.com

Proyecto sobre parejas gay latinas. Promoviendo relaciones saludables entre hombres latinos y sus parejas del mismo sexo. Tel: 646 322 7918 #ColumbiaUniversity
New York, NY • facebook.com/latinos.enpare...

Foro No Heterosexual @enehache_com 1d
Sexo con amigos ¿buena decisión? tinyurl.com/k3363tc

Bitacora Sexodiversa @BitacoraDSx 1d
#Uruguay se lanza al agua bit.ly/11g1Yci #LGBT #matrimonioigualitario pic.twitter.com/hmKtF50gqb @sentimosdiverso

Juan Sebastián @JuanSebasCabral 3d
Hablar de #VIH es hablar de 1.4 millones de latinoamericanos que viven con el virus. "Si no hablamos de lo que pasa, dejamos que pase"

Works at Columbia University in the City of New York
Lives in New York, New York
In a relationship
Born on December 20, 1985

Latinos En Pareja
Proyecto sobre Parejas gay Latinas!!! Quieres saber más? Tel: 646 322 7918. Cada participante puede recibir hasta $120.
www.facebook.com/latinos.enpareja

Figure 3. Some of the community-based venues promoted through our social media sites.

Latinos En Pareja
Wednesday at 7:23 PM in New York, NY

Promoviendo y creciendo! Compartiendo con el Lic. Fidel Bu Contreras en #betanceshealthcenter — with **Jonathan Cb** and **2 others** at **Betances Health Unit**.

Promoviendo el proyecto en la comunidad

Latinos En Pareja
7 hours ago in New York, NY

Promoviendo, creciendo y triunfando. Te sumas a la causa? Tel: 646 322 7918 **#amor #gay #gaylatino** — with **Omar Martínez-Foyedo** at **AIDS Community Research Initiative of America (ACRIA)**.

Latinos En Pareja shared Latinos En Pareja's photo.
Jan 23 ·

Latinos En Pareja at HAF NYC.

Latinos En Pareja shared Latinos En Pareja's photo.

Messages

We used a cell phone and the research center's computer to deliver recruitment and promotional messages for the study. The message refers to the information intended to reach recipients. Stakeholder meetings, which included representatives from community-based organizations, assured the research staff that the message was culturally relevant and directed at Spanish-speaking Latino gay couples. The stakeholders discussed and offered advice on the study name and logo (Figure 4), wording of the publicity messages, visual layout of the palm cards, study's website, and advertisements. The stakeholders, along with the research staff, suggested a very specific name, "Latinos en Pareja", that would attract our specific target population. Furthermore, all the materials and postings with the name of the study featured a photograph of a same-sex Latino couple. A group of Latino gay couples associated with the study consented to share their pictures for advertising and promotional purposes related to the study.

Developing the study's recruitment message for social media channels such as Facebook and the study website also included describing the program as "Promoviendo relaciones saludables entre hombres Latinos y sus parejas del mismo sexo [Promoting strong healthy relationships between Latino men and their same-sex partners]". Stakeholders also recommended including language that would conjure up images of building something for the good of the collective, "Los participantes completarán tres sesiones con otras parejas para crear programas innovadores de prevención del VIH en nuestra comunidad [Participants will complete three sessions with other couples to help build stronger HIV prevention programs in our community]". Additionally, stakeholders recommended that we include postings related to social causes such as same-sex marriage and job and volunteer opportunities (Figure 5). Stakeholders also stressed the importance of including photographs showing the diversity of the Latino gay community, in terms of age, country of origin, race, socioeconomic status, gender, body type, and sexual preferences. Potential participants who had expressed interest in the study and “friended” our pages had the opportunity to respond to these postings and share their couples stories, anecdotes, and photos on our sites.

We used a more simplistic approach in publicizing our message through the smartphone applications such as Grindr, SCRUFF, and Jack’d. Our profile heading included the words “En pareja [In a relationship?]?” and a picture of a couple holding each other. The couple was part of our stakeholder team who consented in writing to use of the picture. Once potential participants clicked on the profile, they were directed to the following heading, “Proyecto sobre Latinos en Pareja [Latino gay couples study]”, and a brief description that included the following message, “Proyecto sobre latinos en pareja. Promoviendo relaciones saludables. Cada participante puede recibir hasta $120. Nos puede llamar ahora para determinar si es elegible [Latino gay couples study. Promoting healthy relationships. Each participant can receive up to $120. Call us to determine if you are eligible]”, plus a phone number and email address. Participants also had the option to access our Latinos en Pareja Facebook page, using the social network link built into the apps.

Figure 4. Latinos en Pareja logo.

Figure 5. Collection of photos with job opportunities and information on social issues.

1/4/14 Sex Trafficking of LGBT Individuals: A Call for Service Provision, Research, and Action

INTERNATIONAL LAW NEWS

WWW.AMERICANBAR.ORG/INTLAW

Home > Publications > International Law News > 2013 > Fall 2013: Human Trafficking

Sex Trafficking of LGBT Individuals: A Call for Service Provision, Research, and Action

Vol. 42 No. 4

By Omar Martinez, Guadalupe Kelle

Omar Martinez (om2222@columbia.edu), *JD, MPH, is a postdoctoral research fellow at the HIV Center for Clinical and Behavioral Studies at the New York State Psychiatric Institute and Columbia University. Mr. Martinez’s research is supported by a National Institute of Mental Health training grant (principal investigator: Theo Sandfort, PhD).* Guadalupe Kelle (gkelle@casamateo.org), *RN, is executive director at Casa Mateo/Latino Action League, Inc., in Indianapolis, Indiana. Latino Action League Inc. provides bilingual HIV/AIDS and health education, prevention support, and other social services to the Latino community through collaboration with local and state governmental agencies and public and private sector organizations, including research institutions such as the Center for Sexual Health Promotion at Indiana University.*

Human trafficking, also referred to as trafficking in persons, is a crime typically hidden within communities and societies. According to the U.S. Department of State, human trafficking involves recruiting, supplying, transporting, and/or obtaining persons for involuntary labor or repayment of debt by the use of coercion, force, or fraud. Human trafficking can include a wide

Call for ILN Articles

Spring Issue
Deadline: December 2, 2013
Theme: International Legal Education

Summer Issue
Deadline: April 1, 2014
Theme: Southeast Asia

International Law News (*ILN*) publishes high-quality submissions on topics in our field. *ILN* editors are especially interested in articles addressing each issue’s theme, but articles on other topics are also appreciated. To learn more about submitting articles to *ILN*, please refer to the *ILN* Author Guidelines.

1/7/14 Job (Bronx): Latino Health Program Coordinator / Community Health Program - idealist.org

Latino Health Program Coordinator / Community Health Program Coordinator

Job posted by: Planned Parenthood of New York City
Posted on: *January 2, 2014*

Job description

We are currently recruiting for a full-time **Community Health Program Coordinator** based in the Education department's Bronx office located at 349 E. 149th Street.

This is an exempt position reporting to the Director of Adult & Professional Programs.

POSITION SUMMARY:

As part of a growing and dynamic Adult & Professional Programs team, the Community Health Program Coordinator will assist the Director of Adult & Professional Programs in the management of education and training programs primarily for immigrant adults. Responsibilities will include: the daily management of the *Promotores de Salud* program, (*Promotores* are concerned community members with strong ties to Latino immigrant communities throughout NYC, trained by PPNYC to promote sexual and reproductive health (SRH) education within their own communities and link community members with SRH services, including those offered at PPNYC health centers), coordination of outreach to Latino immigrants and immigrant-serving groups, including partnerships with various Consulates, and training design and facilitation for PPNYC's Training Institute. The Community Health Program Coordinator will recruit and train *Promotores*, in addition to providing ongoing supervision and support including professional development trainings, monthly team meetings, and one-on-one supervision. The Community Health Program Coordinator will collect service logs on a monthly basis, documenting health center referrals and provision of community education. The Community Health Program Coordinator will create and maintain a master database of all community outreach and education administered by *Promotores de Salud*. The Community Health

Location

349 E 149th Street, Bronx, NY, US

Details

Education requirements
4-year degree

Languages needed
Spanish required, Bilingual

Employment type
Full time

Professional level
None specified

Salary details
$40,000 annually

Benefits
Excellent benefits package.

Job function
training, Education, Education & Training, Community Health, spanish, Health and medical, Latino, Partnership Development, Professional Development

Latinos En Pareja
January 6 at 1:50 PM

!Oportuidades de Trabajo! Asesor/a en Comunicación y Gestión del Conocimiento

La Red Latinoamericana y del Caribe de Mujeres Trabajadoras Sexuales (RedTraSex) está implementando un proyecto regional del Fondo Mundial para el SIDA, la TB y la Malaria (FM) aprobado en la ronda décima.
A continuación se detalla información sobre la organi… Continue Reading

RedTraSex
La Red Latinoamericana y del Caribe de Mujeres Trabajadoras Sexuales (RedTraSex) está implementando un proyecto regional del Fondo Mundial para el SIDA, la TB y la Malaria (FM) aprobado en la ronda décima. La ...
REDTRASEX.ORG

Latinos En Pareja
January 6 at 12:44 PM

¡Nuestro equipo les desea un Feliz Día de los Reyes!

Promoviendo relaciones saludables entre hombres latinos y sus parejas del mismo sexo.

Los participantes completarán tres sesiones con otras parejas para crear

Latinos En Pareja
January 21 at 8:42 AM

© Springer Science+Business Media New York 2013

Abstract Over the past two decades, new anti-immigration policies and laws have emerged to address the migration of undocumented immigrants. A systematic review of the literature was conducted to assess and understand how these immigration policies and laws may affect both access to health services and health outcomes among undocumented immigrants. Eight databases were used to conduct this review, which returned 325 papers that were assessed for validity based on specified inclusion criteria. Forty critically appraised articles were selected for analysis; thirty articles related to access to health services, and ten related to health outcomes. The articles showed a direct relationship between anti-immigration policies and their effects on access to health services. In addition, as a result of these policies, undocumented immigrants were impacted by mental health outcomes, including depression, anxiety, and post-traumatic stress disorder. Action items were presented, including the promotion of cultural diversity training and the development of innovative strategies to support safety-net health care facilities serving vulnerable populations.

Keywords Access to health services · Health outcomes · Health status · Anti-immigration policies and laws · Undocumented immigrants

Introduction

Due to vastly different living standards caused by large income disparities between developed and developing countries, people have been moving to more promising and developed regions throughout history [1–3]. We have seen signs of this phenomenon in the 1990s when Africans crossed the Sahara desert and climbed barbed wire fences in the Spanish enclaves of Ceuta and Melilla in order to enter the European Union. During this time, the number of sub-Saharan undocumented immigrants to Europe started to rise, prompted by the rapidly changing political map of

O. Martinez (✉) · T. Sandfort · A. Carballo-Dieguez

B. Dodge

Evaluating the Impact of Immigration Policies on Health Status Among Undocumented Immigrants: A...
link.springer.com

Source

Our research study was run from two prominent research centers at Columbia University: the HIV Center for Clinical and Behavioral Studies and the Social Intervention Group (SIG). In particular, we received recruitment support, guidance, and training by the HIV Center Media Core through various opportunities, including Cross-Core Meetings, which are built to provide guidance for new studies at the HIV Center. In addition, the study received the support from research staff at the Wake Forest University School of Medicine, North Carolina, and the Department of Social Work at the University of Texas at El Paso.

Results

Ethical Considerations in Conducting Social Media Recruitment

The use of social media for recruitment of participants involves a number of new ethical considerations that must be integrated into the training of stakeholders and research staff and into study protocols to ensure proper protection of potential participants. Prior to recruitment of participants, stakeholders and research staff met to outline strategies to protect potential participants' confidentiality and to discuss proper ways to approach potential participants. Stakeholders recommended the development of a protocol, which was later developed by a collaborative team of stakeholders and research staff (Multimedia Appendix 1). In addition, research staff and stakeholders received training on how to recruit and engage participants through social media for the purposes of cultivating interest in the study.

Regarding Facebook recruitment, we agreed that the Facebook group and study page should have the same name as the study, “Latinos en Pareja”, and must be open to the public so it could be found by open searches and members could invite others to join. Importantly, all group members could post public messages, notes and pictures, and engage in private messages and chats. Facebook users were made aware that their postings, including photos, may be used for the purposes of the study via a disclaimer on our Facebook page. In addition, all of the photos published in this manuscript were shared directly with the research team by Facebook users to use for study-related deliverables. The PI monitored the Facebook group and page and responded to private messages and inquiries about the study with the approved New York State Psychiatric Institute Institutional Review Board (NYSPI IRB) language. Research staff had weekly discussions with stakeholders to ensure privacy was upheld in all of our social media channels. We did not collect identifying data for any of the potential participants who contacted us for information about the study. However, as approved by the NYPSI IRB, those who screened eligible were given the option to provide their personal information in order to be contacted for the study. All identifying data for the study comply with the Health Insurance Portability and Accountability Act (HIPAA) regulations, and storage of data is encrypted, firewalled, and password-protected. Although more ethical issues will emerge as online engagement continues, we have sought to address major ethical concerns both in the training and study methods.

Social Media

We relied on a wide range of social media and community-based venues for recruitment (Table 1). Potential participants were instructed to call our direct research line to be screened to determine their eligibility. Facebook proved to be the most successful tool to recruit participants into the study. By the end of January 2014, a month after the study was initiated, and after having successfully recruited all the eligible couples, we had 1872 members associated with the Facebook individual profile and 1370 members in the Facebook group. We reached approximately 6612 weekly through the Facebook interactive page with a total of 589 friends. Additionally, statistical analytical tools built into the Facebook group page enabled us to keep track of the “likes” and the number of individuals “reached” on a weekly basis. A total of 44 participants were screened through Facebook, and 5 of the 14 eligible couples were recruited using this site. “Liking” other causes or groups enabled “Latinos en Pareja” to share targeted recruitment materials on the pages of other groups and causes. For instance, the Hispanic AIDS Awareness Program, a nationally well-known advocacy organization, shared several of our postings to help promote the Latino gay couples study. In return, we also shared and endorsed their site on our page. Our Facebook group particularly targeted other groups catering to Latinos on topics such as Latino parties in the Jackson Heights neighborhood of Queens, New York, wellness promotion events, and social venues for Latino gay men. By the end of the first month of the study, we had hundreds of members associated with our Facebook page. Some of these partner groups on Facebook include “Gay men interracial lovers”, “Gay Latinos in NYC and Friends”, “New York Latinos Underground”, “Gay Latinos Worldwide”, and “Core-Group Latinos D”.

Table 1. Summary of recruitment reach through social media channels and community-based venues.

Social media sites (Jan. 1-31, 2014)	Number of potential participants reached or total number of messages (N=35,658)	Eligible couples (N=14)
Facebook		5
Screeners	44	
Messages	1000	
Individual profiles	1872	
Groups	1370	
Pages	589	
Total number of posts	120	
Weekly page reach	6612	
Grindr		0
Screeners	21	
Messages	300	
SCRUFF		0
Screeners	1	
Messages	200	
Jack’d		0
Screeners	5	
Messages	250	
Craigslist		0
Screeners	4	
Postings	10	
Messages	50	
Twitter		
Screeners	0	0
Following on Twitter	2003	
Followers	315	
Tweets	36	
Messages	150	
Instagram		0
Screeners	0	
Following on Instagram	667	
Followers	167	
Posts	5	
Community-based organizations		5
Hispanic AIDS forum Screeners	4	
Betances	3	
Latino Commission on AIDS Screeners	2	
BOOM! Health Screeners	2	
Make the Road New York Screeners	2	
Others		4
Screeners through couples or friends’ referral	18	

Many potential participants saw our Facebook profile as a reflection of the value of collectivism in Latino communities as well as a networking tool enabling them to share their personal stories about being a Latino gay couple. Interestingly, many potential participants also visited the profile as single Latino gay men hoping to find other partners. In addition, members shared events and parties taking place in the community. In particular, members found merit in our strategy to post jobs and volunteer opportunities in the New York City area on the group page. Stakeholders used the group page to refer jobs to member, and we also used Indeed, the popular job search engine, to advertise job and volunteer opportunities to group members. Many couples also shared their picture, stories, anecdotes, and provided advice and suggestions on how to maintain healthy relationships (Figure 6).

Our Instagram application enabled us to share pictures related to couples and HIV prevention and to send messages that Latino gay couples would identify with using the popular social media tool of hashtags, which allows social media users to create trends, share threads of information, and reach a wider general audience. Some of our hashtag messages included #latinosenpareja, #*Prevención*delVIH, #amordeparejas, #amorenpareja, #justicia, #*acción, among many others*. The Instagram application was shared and connected with our Facebook page, helping us streamline and integrate all of our communications and messages. Twitter was also used to shared news and important events affecting the Latino MSM community. The same strategies were used on Twitter by joining relevant profiles sharing similar interests and causes and targeting the same group of Latino MSM.

Our website featured recruitment information about the study, pictures of couples, positive messages including our objective in "promoviendo relaciones saludables [promoting healthy relationships]" and "crear programas innovadores de prevención del VIH en nuestra comunidad [creating innovative HIV prevention programs in our community]", as well as information about stakeholders, volunteers, and the research team. In addition, through the website, we constructed a "Contact Us" inbox where potential participants could email us with their inquiries about the study as well as other related questions [35].

These built-in tools in our channels, including Internet pages and smartphone applications, enabled us to increase our engagement through social media. In total, we received 3670 messages from social media sites, screened 75 potential participants of the total 106 screened, and recruited 7 eligible couples through social media. Most of our screeners on smartphone applications were from Grindr, with a total of 21 potential participants screened.

Engaging with various social media platforms led to an interactive dialogue between potential participants, community stakeholders, and the research team. In particular, our social media presence and messages received a great deal of positive feedback. Individuals engaged with our online social media tools shared their thoughts, pictures, and opinions on the study through a public forum that connected them to a community of like-minded individuals. For example, one individual stated on our Facebook wall:

> *Estoy promoviendo este espectacular proyecto de investigación que la Universidad de Columbia y un grupo de amigos están desarrollando en la ciudad de Nueva York. Si te consideras un hombre comprometido con tu pareja y responsable con tu entorno, no deberías quedar ajeno a esta iniciativa que pretende cambiar el modo en que enfrentamos el mundo. Además, si somos amigos es porque te considero alguien valioso y quiero lo mejor para ti. Por eso te invito a que junto a quien amas vivan esta experiencia. [I'm promoting this wonderful research study developed by Columbia University and a group of my friends in New York City. If you are a man in a committed relationship with your partner and responsible in your community, you should not miss out on this new initiative that is trying to change the way we deal with the problems we face in our world. Furthermore, if we are friends, it is because I consider you as someone who is valuable and I want the best for you. For this reason I'm inviting you and your loved one to join us in this experience.]* [Facebook user]

Another Facebook user shared "Gracias por su presencia. Existe falta de difusión de forma que la ayuda llegue más proporcionalmente. Ustedes son un buen ejemplo de ayuda en información que se necesita casi urgente. Saludos [Thank you for your presence. There is a real need to reach out and spread the word more efficiently. You are a good example of a source of useful information that is urgently needed]"

Furthermore, social media proved to be an effective tool for sharing HIV information and knowledge, testing resources, and agency referrals. Potential participants discussed and responded to our information with thoughtful insights about factors leading to HIV and barriers to medical care. One Facebook user explained:

> *Cabe destacar que el incremento del HIV tiene mucho que ver con tres factores muy importantes: 1) economía 2) educación y 3) ambiente. Aunque todos estamos expuestos a una variedad de enfermedades de transmisión sexual, pobre y ricos, las personas de bajos recursos económicos son más propensas al contagio. También existen factores como la promiscuidad y comportamiento relacionado al consumo de alcohol y drogas ilícitas. Por otra parte, la educación sexual escolar es mínima o no existe en los planteles escolares públicos, lo que contribuye a la intimidad sexual sin conocer los riesgos que implica falta de higiene y protección con anti-conceptivos u otros métodos conocidos como condón o abstinencia. Y por último, personas que viven en áreas pobres o de bajos recursos económicos, zonas urbanas o suburbios, tienden a vivir sin el cuidado de personas adultas. Por otro lado, vivir en confinamiento, apartamentos u hogares de menor espacio en relación a la cantidad de personas residiendo bajo un mismo techo, también se puede considerar como un grave problema que aumenta la posibilidad de adquirir tanto HIV como*

> *otras enfermedades relacionadas al contacto sexual. [It is worth mentioning that the rise in HIV has to do with three important factors: (1) the economy, (2) education, and (3) the environment. Although all of us are exposed to a variety of sexually transmitted diseases, both rich and poor, people with fewer economic resources are more prone to infection. There are also other factors to consider such as promiscuity and behaviors related to the consumption of alochol and illegal drugs. On the other hand, sexual education in schools is very minimal or does not exist in public schools, a fact that contributes to sexual intimacy without knowing the risks that implies a lack of hygiene or protection using contraceptives or other known methods like condoms and abstinence. And finally, people who live in poor areas or those with fewer economic resources, in urban or suburban zones, tend to live without the supervision of adults. On the other hand, living in confined spaces, in cramped apartments or homes with many people under the same roof, also could be considered a serious problem that increases the possibility of acquiring HIV and sexually transmitted infections.]*

Our focused and targeted message about the study and the participants' positive perception of our study enabled us to increase recruitment and engage others through all of the social media venues we used. In addition, through the sharing of sexual health information, we enabled users to engage in thoughtful discussions about HIV, sexually transmitted infections (STIs), and sexual health promotion.

Figure 6. Collection of photos shared directly with the research team by Facebook users with their consent to use for study-related deliverables.

Y siguen llegando historias lindas: Casi dos años juntos, tres nenes, Jose de Mexico y Pual de Vietnam...como la hacen con tantos nenes!?! Un abrazote! Gracias chicos por compartir! — at **Starbucks**.

71 Likes 5 Comments

Share

Latinos En Pareja
Tuesday at 10:04 PM in New York, NY

AMOR a pasito norteño! :-) Love it! — with **Adrian Velez** in **New York, New York**.

71 Likes 6 Comments

Share

Gracias chicos por compartir sus historias, experiencias y anécdotas con nuestro equipo de **#latinosenpareja**! TODO es POSIBLE! El AMOR llega - de diferentes formas, maneras y colores! — in **New York, New York**.

93 Likes 6 Comments

Se nos casan en verano-Mexicano y Puertorriqueño! Nos invitan a la boda?! Pa ya vamos con la guaracha de **#olgatanon** y las rancheras de **#vicentefernandez** Felicidades!!!

135 Likes 14 Comments

Community-Based Organizations and Other Recruitment Venues

We reached more than 50 community-based organizations in the five boroughs of New York City (Bronx, Brooklyn, Manhattan, Queens, and Staten Island). However, a large majority of the potential participants screened through community-based organizations came to us from Hispanic AIDS Forum (n=4), Betances Health Center (n=3), Latino Commission on AIDS (n=2), BOOM! Health Center (n=2), and Make the Road New York (n=2) (Table 1). In total, we recruited 5 eligible couples for the study through community-based organizations. In addition, 4 eligible couples were enrolled through referrals from other couples or friends. We also recruited in barbershops where most of the clientele were Latino and lesbian, gay, bisexual, and transgender (LGBT) individuals, including Pride Cuts in Manhattan.

Discussion

Principal Findings

Our findings show that monolingual Spanish-speaking Latino MSM who are less acculturated create strong and diverse networks through social media. This engagement goes beyond "joining a group" and expands into online social ties that in turn translate into social integration, informational support, communication, and engagement. This could be explained by the idea that Latino immigrants, vis-à-vis their US-born counterparts, form strong social ties and social support groups [36-38]. Our initial findings on participants' engagement with social media platforms associated with the study suggest that the value of collectivism that is part of Latino communities carries over into online communities targeting Latinos as well. In addition, our findings underscore the importance of social media ties and interactions in the development of programs and interventions to improve health outcomes. Individuals were comfortable discussing topics related to couples, including sexual health, non-monogamous relationships, disclosure of HIV status, and communication skills, among others. Individuals also shared their thoughts and the need to build couple-based programs and counseling services.

We also encourage and suggest that social media networking sites fully endorse and support research institutions in their endeavors to address complex health challenges affecting vulnerable populations. In particular, social media should help promote low-budget studies that are not able to spend on major advertisement campaigns through their sites. We do understand that this commitment comes with mutual respect and agreements. Social media sites and mobile applications, including Grindr, SCRUFF, and others, might benefit from the established networks as a result of the program or intervention. For instance, social media partners could engage those established networks by the research team through other channels including community-based organizations and Facebook. As the influence of social media expands, social media sites owe and share in the social responsibility to promote community health, and thus they should actively engage with researchers, social activists, and community stakeholders.

Another important topic to consider is the sustainability of the study sites and built networks. The PI will maintain the website as a resource channel for Latino gay couples well after the completion of the research funding cycle. This site could serve as potential space to house future recruitment initiatives for other studies targeting Latino gay couples. In case funding alternatives do not work for couple-based studies, the PI will train community stakeholders to manage the site and will encourage monthly meetings to engage in discussion and continue supporting those involved in the site. Other social media channels, such as Grindr, SCRUFF, and Jack'd, will be deactivated at the conclusion of the recruitment phase of the study.

Limitations

Given the limited resources available to evaluate the success of our social media recruitment tools to deliver information, a detailed qualitative and quantitative content analysis of the social medial channels was not feasible. In order to maintain high ethical standards using new social medial tools, we refrained from disclosing personal messages sent by participants that could have provided further insights in regards to the enhancement of participant engagement and recruitment. In addition, we could not access the impact of delivered information and shared knowledge on participants, including the effect of the shared HIV and STI information, as well as referral references and how these might have impacted participants' behaviors. Future studies should further analyze how Spanish-speaking Latino MSM might benefit from a built-in online HIV prevention intervention using existing channels such as Facebook, Grindr, and SCRUFF, to mention a few.

Strengths

Our success relied on the ongoing support from our stakeholders and community partners who were involved throughout the recruitment stage by providing critical feedback, refining messages and contents, and guiding the process of recruitment. As other studies have shown, the involvement of stakeholders in the research process is key to the success of programs, interventions, and research [39-42]. Our research staff benefited from this type of partnership and collaboration, and it will hopefully extend into the implementation of the HIV prevention intervention for Latino gay couples at the community level at a later stage of the process. Stakeholders also benefited from this collaboration. During the meetings, the research staff provided access to and discussed up-to-date research on the use of social media and technology to disseminate health information and promote wellness. In addition, three community stakeholders were introduced, connected, and enrolled in free English summer courses with the English as a Second Language (Community Impact Program) Program at Columbia University. Community involvement extends to all aspects of our adaptation research process and will continue expanding by building trust and exchanging information, sustaining relationships and commitments, and sustaining and maintaining knowledge, capacity, and values generated from the partnership.

Conclusions

Our findings provide initial support for the feasibility of recruiting a diverse group of monolingual Spanish-speaking Latino gay couples and notes the strategies, including targeted messaging, needed to recruit this population—a population that we have found to not be so “hard to reach” after all if social media platforms and community outreach efforts are combined in a strategic and culturally appropriate way. The study’s social media presence reached a substantial number of fans and continued to engage community members and serve as a platform for interacting and sharing information for Latino MSM. Our interactive approach to developing recruitment materials, implementing recruitment strategies, and filtering content feedback with community stakeholders allowed ongoing improvements to the recruitment materials. In the end, this interactive approach expanded our reach within communities of gay and other Latino MSM using these important social media network spaces. The use of multiple social media proved highly effective to provide a window into a population that is often characterized as “hard to reach”.

Acknowledgments

We want to thank all of our current and potential future users who have entrusted our site by sharing information, pictures, and comments that have benefited our study. This work was graciously supported by National Institutes of Mental Health (NIMH): Dr Robert Ramien, P30 MH43520. Mr Martinez is supported by a NIMH training grant (T32-MH19139 Behavioral Sciences Research in HIV Infection; Principal Investigator: Theo Sandfort, PhD).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Social media protocol.

[PDF File (Adobe PDF File), 111KB - jmir_v16i4e113_app1.pdf]

References

1. Kuehn BM. More than one-third of US individuals use the Internet to self-diagnose. JAMA 2013 Feb 27;309(8):756-757. [doi: 10.1001/jama.2013.629] [Medline: 23443421]
2. Altenberg M. Playing the Mysterious Game of Online Love: Examining an Emerging Trend of Limiting & 230 Immunity of the Communications Decency Act and the Effects on E-Dating Websites. Pace Law Review 2012;32(3):922-954 [FREE Full text]
3. Boyd D. Social Network Sites: Definition, History, and Scholarship. Journal of Computer-Mediated Communication 2007;13(1):210-230. [doi: 10.1111/j.1083-6101.2007.00393.x]
4. Young S, Szekeres G, Coates T. The relationship between online social networking and sexual risk behaviors among men who have sex with men (MSM). PLoS One 2013;8(5):e62271 [FREE Full text] [doi: 10.1371/journal.pone.0062271] [Medline: 23658716]
5. Hirshfield S, Chiasson MA, Joseph H, Scheinmann R, Johnson WD, Remien RH, et al. An online randomized controlled trial evaluating HIV prevention digital media interventions for men who have sex with men. PLoS One 2012;7(10):e46252 [FREE Full text] [doi: 10.1371/journal.pone.0046252] [Medline: 23071551]
6. Kubicek K, Carpineto J, McDavitt B, Weiss G, Kipke MD. Use and perceptions of the internet for sexual information and partners: a study of young men who have sex with men. Arch Sex Behav 2011 Aug;40(4):803-816 [FREE Full text] [doi: 10.1007/s10508-010-9666-4] [Medline: 20809373]
7. Wall KM, Stephenson R, Sullivan PS. Frequency of sexual activity with most recent male partner among young, Internet-using men who have sex with men in the United States. J Homosex 2013;60(10):1520-1538. [doi: 10.1080/00918369.2013.819256] [Medline: 24059971]
8. Minichiello V, Rahman S, Dune T, Scott J, Dowsett G. E-health: potential benefits and challenges in providing and accessing sexual health services. BMC Public Health 2013;13:790 [FREE Full text] [doi: 10.1186/1471-2458-13-790] [Medline: 23987137]
9. Magee JC, Bigelow L, Dehaan S, Mustanski BS. Sexual health information seeking online: a mixed-methods study among lesbian, gay, bisexual, and transgender young people. Health Educ Behav 2012 Jun;39(3):276-289. [doi: 10.1177/1090198111401384] [Medline: 21490310]
10. Barman-Adhikari A, Rice E. Sexual Health Information Seeking Online Among Runaway and Homeless Youth. J Soc Social Work Res 2011 Jun;2(2):88-103 [FREE Full text] [doi: 10.5243/jsswr.2011.5] [Medline: 22247795]
11. Kakietek J, Sullivan PS, Heffelfinger JD. You've got male: internet use, rural residence, and risky sex in men who have sex with men recruited in 12 U.S. cities. AIDS Educ Prev 2011 Apr;23(2):118-127 [FREE Full text] [doi: 10.1521/aeap.2011.23.2.118] [Medline: 21517661]

12. Al-Tayyib A. Finding sex partners on the internet: what is the risk for sexually transmitted infections. Sexually Transmitted Infections 2009;85(158):216-220. [doi: 10.1136/sti.2008.032631]
13. Grov C, Crow T. Attitudes about and HIV risk related to the "most common place" MSM meet their sex partners: comparing men from bathhouses, bars/clubs, and Craigslist.org. AIDS Educ Prev 2012 Apr;24(2):102-116. [doi: 10.1521/aeap.2012.24.2.102] [Medline: 22468972]
14. Tashima KT, Alt EN, Harwell JI, Fiebich-Perez DK, Flanigan TP. Internet sex-seeking leads to acute HIV infection: a report of two cases. Int J STD AIDS 2003 Apr;14(4):285-286. [doi: 10.1258/095646203321264926] [Medline: 12716501]
15. Ybarra ML, Bull SS, Prescott TL, Birungi R. Acceptability and feasibility of CyberSenga: an Internet-based HIV-prevention program for adolescents in Mbarara, Uganda. AIDS Care 2014 Apr;26(4):441-447. [doi: 10.1080/09540121.2013.841837] [Medline: 24093828]
16. Young SD, Zhao M, Teiu K, Kwok J, Gill H, Gill N. A social-media based HIV prevention intervention using peer leaders. J Consum Health Internet 2013 Oct 1;17(4):353-361 [FREE Full text] [doi: 10.1080/15398285.2013.833445] [Medline: 24526928]
17. Jaganath D, Gill HK, Cohen AC, Young SD. Harnessing Online Peer Education (HOPE): integrating C-POL and social media to train peer leaders in HIV prevention. AIDS Care 2012;24(5):593-600 [FREE Full text] [doi: 10.1080/09540121.2011.630355] [Medline: 22149081]
18. Noar SM, Willoughby JF. eHealth interventions for HIV prevention. AIDS Care 2012;24(8):945-952. [doi: 10.1080/09540121.2012.668167] [Medline: 22519523]
19. Wei C, Herrick A, Raymond HF, Anglemyer A, Gerbase A, Noar SM. Social marketing interventions to increase HIV/STI testing uptake among men who have sex with men and male-to-female transgender women. Cochrane Database Syst Rev 2011(9):CD009337. [doi: 10.1002/14651858.CD009337] [Medline: 21901734]
20. Reback CJ, Ling D, Shoptaw S, Rohde J. Developing a Text Messaging Risk Reduction Intervention for Methamphetamine-Using MSM: Research Note. Open AIDS J 2010;4:116-122 [FREE Full text] [doi: 10.2174/1874613601004030116] [Medline: 20657827]
21. Ko NY, Hsieh CH, Wang MC, Lee C, Chen CL, Chung AC, et al. Effects of Internet popular opinion leaders (iPOL) among Internet-using men who have sex with men. J Med Internet Res 2013;15(2):e40 [FREE Full text] [doi: 10.2196/jmir.2264] [Medline: 23439583]
22. Young SD. Recommended guidelines on using social networking technologies for HIV prevention research. AIDS Behav 2012 Oct;16(7):1743-1745 [FREE Full text] [doi: 10.1007/s10461-012-0251-9] [Medline: 22821067]
23. Muessig KE, Pike EC, Fowler B, LeGrand S, Parsons JT, Bull SS, et al. Putting prevention in their pockets: developing mobile phone-based HIV interventions for black men who have sex with men. AIDS Patient Care STDS 2013 Apr;27(4):211-222. [doi: 10.1089/apc.2012.0404] [Medline: 23565925]
24. Chou W, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
25. Sullivan, Khosropour CM, Luisi N, Amsden M, Coggia T, Wingood GM, et al. Bias in online recruitment and retention of racial and ethnic minority men who have sex with men. J Med Internet Res 2011;13(2):e38 [FREE Full text] [doi: 10.2196/jmir.1797] [Medline: 21571632]
26. Young SD, Szekeres G, Coates T. Sexual risk and HIV prevention behaviours among African-American and Latino MSM social networking users. Int J STD AIDS 2013 Aug;24(8):643-649. [doi: 10.1177/0956462413478875] [Medline: 23970575]
27. Vyas AN, Landry M, Schnider M, Rojas AM, Wood SF. Public health interventions: reaching Latino adolescents via short message service and social media. J Med Internet Res 2012;14(4):e99 [FREE Full text] [doi: 10.2196/jmir.2178] [Medline: 22789678]
28. Mutchler MG, McKay T, McDavitt B, Gordon KK. Using peer ethnography to address health disparities among young urban Black and Latino men who have sex with men. Am J Public Health 2013 May;103(5):849-852. [doi: 10.2105/AJPH.2012.300988] [Medline: 23488490]
29. Stueve A, O'Donnell LN, Duran R, San Doval A, Blome J. Time-space sampling in minority communities: results with young Latino men who have sex with men. Am J Public Health 2001 Jun;91(6):922-926. [Medline: 11392935]
30. Auerswald CL, Greene K, Minnis A, Doherty I, Ellen J, Padian N. Qualitative assessment of venues for purposive sampling of hard-to-reach youth: an illustration in a Latino community. Sex Transm Dis 2004 Feb;31(2):133-138. [doi: 10.1097/01.OLQ.0000109513.30732.B6] [Medline: 14743078]
31. McGuire WJ. Attitudes and attitude change. In: Lindzey G, Aronson E, editors. Handbook of Social Psychology. New York: Random House; 1985.
32. McKee MB, Picciano JF, Roffman RA, Swanson F, Kalichman SC. Marketing the 'Sex Check': evaluating recruitment strategies for a telephone-based HIV prevention project for gay and bisexual men. AIDS Educ Prev 2006 Apr;18(2):116-131. [doi: 10.1521/aeap.2006.18.2.116] [Medline: 16649957]
33. Wu E, El-Bassel N, Donald McVinney L, Fontaine YM, Hess L. Adaptation of a Couple-Based HIV Intervention for Methamphetamine-Involved African American Men who have Sex with Men. Open AIDS J 2010;4:123-131 [FREE Full text] [doi: 10.2174/1874613601004030123] [Medline: 20657720]

34. Wu E, El-Bassel N, McVinney LD, Hess L, Remien RH, Charania M, et al. Feasibility and promise of a couple-based HIV/STI preventive intervention for methamphetamine-using, black men who have sex with men. AIDS Behav 2011 Nov;15(8):1745-1754. [doi: 10.1007/s10461-011-9997-8] [Medline: 21766193]
35. Latinos en Pareja. URL: http://latinosenpareja.wix.com/latinosenpareja [accessed 2014-04-16] [WebCite Cache ID 6OsMwHLAc]
36. Negi NJ. Social Networks That Promote Well-Being Among Latino Migrant Day Laborers. Advances in Social Work 2013;14(1):247-259 [FREE Full text]
37. Hilfinger Messias DK, Barrington C, Lacy E. Latino social network dynamics and the Hurricane Katrina disaster. Disasters 2012 Jan;36(1):101-121. [doi: 10.1111/j.1467-7717.2011.01243.x] [Medline: 21623889]
38. Rhodes SD, Daniel J, Alonzo J, Vissman AT, Duck S, Downs M, et al. A snapshot of how latino heterosexual men promote sexual health within their social networks: process evaluation findings from an efficacious community-level intervention. AIDS Educ Prev 2012 Dec;24(6):514-526 [FREE Full text] [doi: 10.1521/aeap.2012.24.6.514] [Medline: 23206201]
39. Rhodes SD, Daniel J, Alonzo J, Duck S, García M, Downs M, et al. A systematic community-based participatory approach to refining an evidence-based community-level intervention: the HOLA intervention for Latino men who have sex with men. Health Promot Pract 2013 Jul;14(4):607-616. [doi: 10.1177/1524839912462391] [Medline: 23075504]
40. Rhodes SD, Duck S, Alonzo J, Ulloa JD, Aronson RE. Using community-based participatory research to prevent HIV disparities: assumptions and opportunities identified by the Latino partnership. J Acquir Immune Defic Syndr 2013 Jun 1;63 Suppl 1:S32-S35. [doi: 10.1097/QAI.0b013e3182920015] [Medline: 23673883]
41. Dodge B, Schnarrs PW, Reece M, Martinez O, Goncalves G, Malebranche D, et al. Sexual behaviors and experiences among behaviorally bisexual men in the midwestern United States. Arch Sex Behav 2013 Feb;42(2):247-256 [FREE Full text] [doi: 10.1007/s10508-011-9878-2] [Medline: 22187027]
42. Dodge B, Schnarrs PW, Reece M, Goncalves G, Martinez O, Nix R, et al. Community involvement among behaviourally bisexual men in the Midwestern USA: experiences and perceptions across communities. Cult Health Sex 2012 Oct;14(9):1095-1110 [FREE Full text] [doi: 10.1080/13691058.2012.721136] [Medline: 22978551]

Abbreviations

AIDS: acquired immunodeficiency syndrome
CNU: Connect ‘n Unite
HIPAA: Health Insurance Portability and Accountability Act
HIV: human immunodeficiency virus
LGBT: lesbian, gay, bisexual, and transgender
MSM: men who have sex with men
NYSPI IRB: New York State Psychiatric Institute Institutional Review Board
PI: principal investigator
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 07.02.14; peer-reviewed by B Sheoran, K Muessig; comments to author 14.03.14; revised version received 17.03.14; accepted 13.04.14; published 24.04.14.

Please cite as:
Martinez O, Wu E, Shultz AZ, Capote J, López Rios J, Sandfort T, Manusov J, Ovejero H, Carballo-Dieguez A, Chavez Baray S, Moya E, López Matos J, DelaCruz JJ, Remien RH, Rhodes SD
Still a Hard-to-Reach Population? Using Social Media to Recruit Latino Gay Couples for an HIV Intervention Adaptation Study
J Med Internet Res 2014;16(4):e113
URL: http://www.jmir.org/2014/4/e113/
doi:10.2196/jmir.3311
PMID:24763130

Original Paper

Real-Time Sharing and Expression of Migraine Headache Suffering on Twitter: A Cross-Sectional Infodemiology Study

Thiago D Nascimento[1], MS, DDS; Marcos F DosSantos[1], MS, DDS, PhD; Theodora Danciu[2], DMD, DMedSc; Misty DeBoer[1], BS; Hendrik van Holsbeeck[1], BS; Sarah R Lucas[1], BS; Christine Aiello[1]; Leen Khatib[1], BS; MaryCatherine A Bender[1], BS; UMSoD (Under)Graduate Class Of 2014[3]; Jon-Kar Zubieta[4], MD, PhD; Alexandre F DaSilva[1,4,5], DDS, DMedSc

[1]Headache and Orofacial Pain Effort (HOPE), Biologic and Materials Sciences Department, School of Dentistry, University of Michigan, Ann Arbor, MI, United States

[2]Department of Periodontics and Oral Medicine, School of Dentistry, University of Michigan, Ann Arbor, MI, United States

[3]University of Michigan School of Dentistry (UMSoD), see Acknowledgements for collaborators

[4]Translational Neuroimaging Laboratory, Molecular and Behavioral Neuroscience Institute (MBNI), University of Michigan, Ann Arbor, MI, United States

[5]Michigan Center for Oral Health Research (MCOHR), School of Dentistry, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Alexandre F DaSilva, DDS, DMedSc
Headache and Orofacial Pain Effort (HOPE), Biologic and Materials Sciences Department
School of Dentistry
University of Michigan
The Molecular and Behavioral Neuroscience Institute (MBNI), room 1021
205 Zina Pitcher Pl
Ann Arbor, MI, 48109-5720
United States
Phone: 1 734 615 9390
Fax: 1 734 647 4130
Email: adasilva@umich.edu

Abstract

Background: Although population studies have greatly improved our understanding of migraine, they have relied on retrospective self-reports that are subject to memory error and experimenter-induced bias. Furthermore, these studies also lack specifics from the actual time that attacks were occurring, and how patients express and share their ongoing suffering.

Objective: As technology and language constantly evolve, so does the way we share our suffering. We sought to evaluate the infodemiology of self-reported migraine headache suffering on Twitter.

Methods: Trained observers in an academic setting categorized the meaning of every single “migraine” tweet posted during seven consecutive days. The main outcome measures were prevalence, life-style impact, linguistic, and timeline of actual self-reported migraine headache suffering on Twitter.

Results: From a total of 21,741 migraine tweets collected, only 64.52% (14,028/21,741 collected tweets) were from users reporting their migraine headache attacks in real-time. The remainder of the posts were commercial, re-tweets, general discussion or third person’s migraine, and metaphor. The gender distribution available for the actual migraine posts was 73.47% female (10,306/14,028), 17.40% males (2441/14,028), and 0.01% transgendered (2/14,028). The personal impact of migraine headache was immediate on mood (43.91%, 6159/14,028), productivity at work (3.46%, 486/14,028), social life (3.45%, 484/14,028), and school (2.78%, 390/14,028). The most common migraine descriptor was “Worst” (14.59%, 201/1378) and profanity, the “F-word” (5.3%, 73/1378). The majority of postings occurred in the United States (58.28%, 3413/5856), peaking on weekdays at 10:00h and then gradually again at 22:00h; the weekend had a later morning peak.

Conclusions: Twitter proved to be a powerful source of knowledge for migraine research. The data in this study overlap large-scale epidemiological studies, avoiding memory bias and experimenter-induced error. Furthermore, linguistics of ongoing migraine reports on social media proved to be highly heterogeneous and colloquial in our study, suggesting that current pain questionnaires should undergo constant reformulations to keep up with modernization in the expression of pain suffering in our society. In summary, this study reveals the modern characteristics and broad impact of migraine headache suffering on patients' lives as it is spontaneously shared via social media.

(J Med Internet Res 2014;16(4):e96) doi:10.2196/jmir.3265

KEYWORDS

migraine; headache; epidemiology; social media; Twitter

Introduction

Migraine affects approximately 12% of adults in the Western world [1]. In the United States, the prevalence of migraine is approximately 18% in women and 6% in men [2-4]. About 90% of migraineurs have moderate to severe pain during the attacks, 75% have reduced ability to function, and 30% require bed rest [4-6]. Although population studies have greatly improved our understanding of migraine, they have relied on retrospective self-reports that are subject to memory error and experimenter-induced bias. Furthermore, these studies also lack specifics from the actual time that attacks were occurring, and how patients express and share their ongoing suffering. Investigators have alluded to these limitations [7-11]; up until now, there have been no practical means to evaluate these observations in a geographically diverse population.

Infodemiology is a branch of science that deals with the occurrence, distribution, and analysis of electronic information that is used to inform the public of disease patterns and discourse, and of their relationship to the health status of a population. A key feature of infodemiology is the potential to collect and analyze data in near real time [12]. In this study, we explored the use of social media to evaluate migraine experience using Twitter, an online micro-blogging system. Twitter [13] allows registered users to post short text-based announcements known as "tweets", consisting of a maximum of 140 characters, to an online public and accessible database. Tweets are instant, time-stamped, and self-reported communication from hundreds of millions of people worldwide. Tweets are usually built based on spontaneous reports with a natural self-expression, which makes social media a unique and innovative way to understand how communication and sharing of pain distress evolves. Twitter has been used as a key resource for public health surveillance, such as monitoring prescription drug abuse [14], smoking [15], and dietary behavior [16]. In addition, recent studies demonstrate that data retrieved from Twitter may be used to track dental pain [17], migraine [18], and to assess individual mood changes [19] and happiness [20], suggesting that this tool has the potential for describing universal human behaviors and patterns including emotional, social, and others [20,21]. The linguistics of suffering, as a broad context, is constantly modulated by factors such as social, cultural, and advances in technology. Additionally, the use of instant data avoids bias associated with retrospective reports, increasing accuracy and sensitivity of pain impact [11]. Nevertheless, analyses based on instant searching tools available through the Internet for social media may frequently lead to deceptive measurements due to the diversity of postings that are not all directly related to patient's suffering as a result of migraine; for example, some tweets are associated with drug advertisement, the metaphoric use of the word "migraine", and so forth [18]. To avoid these confounding factors and to estimate the instant impact of actual self-reported migraine headache suffering on the World Wide Web using Twitter, we analyzed the meaning and pattern of every single tweet message with the word "migraine" posted during an entire seven-day period.

Major aims of this study included using social media to assess migraine headache impact in real time to avoid memory bias, and to identify a set of current suffering descriptors that were not prompted by an experimenter. We report that Twitter, used as an instrument for infodemiology [12], is a rich source of information for migraine research with significant overlap of data from previously published large-scale epidemiological studies and has the potential to generate contemporary and clinically relevant results.

Methods

Study Design

A continuous cross-sectional sample of 21,741 tweets was collected between Saturday, April 30 (12:00:00 am) and Friday, May 6, 2011 (11:59:59 pm). According to the official Twitter website, a relative number of 1 billion tweets were posted weekly during that season, with an average of 200 million tweets per day from a total of 100 million active users [22,23]. In 2013, there were 200 million active users, tweeting an average of 400 million tweets per day [24]. The timeframe studied was randomly selected and it included seven consecutive days of uninterrupted posted messages (from Saturday to the following Friday). Hence, assumptions can only be drawn for this particular population that use this tool in social media. The data collected included only free, public Twitter user specific account information, which did not require any log-in data to be obtained.

Ethics Statement

This study was certified as exempt from human subjects review by the University of Michigan Committee on Human Research (reference No. HUM00054476).

Data Collection and Analysis

During the specified seven consecutive days, two investigators alternated in eight-hour shifts compiling all messages posted with the word "migraine" in the Twitter public search engine [25]. All the obtained results were then saved onto a main database. Then, in a systematic manner, three pain specialists oriented and supervised 54 undergraduates, four graduate dental students, and six research assistants on the reading, interpretation, and classification of the tweets into nine categories described below. Calibration lectures and sessions were performed with real-time samples from the Twitter webpage. Furthermore, students were divided into groups under the supervision of elected laboratory members, who answered individual questions during personal meetings and via email in case of uncertainty in the classification of a particular tweet. Participation in this project was only offered following formal lectures on primary headaches, especially migraine, and their classification guidelines based on the International Headache Society [26].

Subsequently, a coding system was used for in-depth interpretation and categorization of each tweet. The categories were: “migraine headache” (a user self-reporting having an actual migraine headache attack), “commercial” (advertising treatments or drugs), “metaphor” (the term migraine was used metaphorically), “not related” (the term migraine does not describe an actual physical experience of headache), “re-tweet” (a re-post of a previous tweet), “third person’s” migraine headache (information is related to another person’s migraine headache attack), general “discussion” (general discussion on migraine), “blanks” (missing data), and “inconclusive” (when not possible to identify the meaning of the word migraine in the tweet). In addition, when available, information about the self-reported migraine headache impact on the users’ sleep, work, social, school, mood, or debilitation was compiled using the same methods described earlier in the text. The following free and public information was also extracted: profile name, gender, and geographic location. All the acquired tweets were automatically translated to English by the Twitter website via Google Translate; however, only tweets originally written in English were used for linguistic analysis to avoid any translation bias.

We analyzed the time/date of occurrence of all the global self-reported migraine headache messages posted on Twitter using Greenwich Mean Time (GMT) for representation of the user time zone, ranging from 0h to 23h. However, the actual geographic location is discretionary information for the user and not always provided. Hence, to achieve a more comprehensive understanding of such temporal behavior, we isolated and then reported times in the United States (the largest representative group) by converting the geographic location time to the appropriate standard time, when available. For instance, an eight o’clock posting from a user on the East Coast in the United States was computed together with a similar posting from another user at local eight o’clock on the West Coast. To make sure every US tweet was corrected to the standard time zone, the US Census Bureau database [27] was used for each self-reported location, based on state and/or counties. For the states that have dual time zones, the same census database was used to estimate which time zone is more prevalent in the state, and that was used as the standard time zone. In addition, since the data was collected during a season when the majority of the United States observes Daylight Savings Time (DST), the US tweets were re-corrected for the local time when applicable. Therefore, we were able to precisely evaluate the temporal pattern of the self-reported migraine headache attacks posted on social media. The temporal pattern was based on a 24-hour time and days of the week.

All the data was compiled in Microsoft Excel, which was used to calculate basic descriptive statistics. Frequencies were reported for each category that was collected.

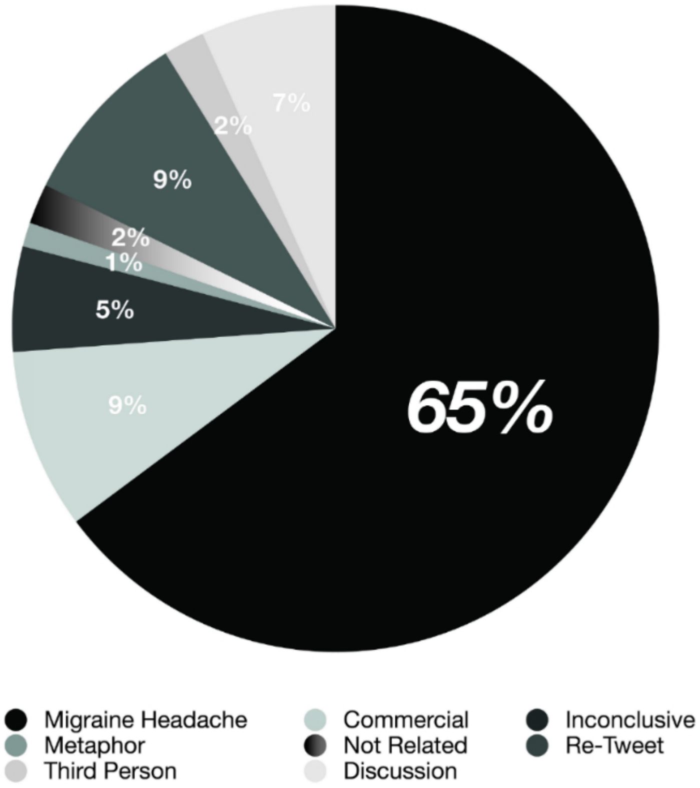
Results

Classification of Tweets in Categories

In a systematic manner, three pain specialists oriented and supervised 54 undergraduate students, four graduate students, and six research assistants on the reading, interpretation, and classification of tweets into the criteria described in the Methods section. For each posting, the following was taken into consideration: semantics, users’ demographics, impact of the attacks, geographic location, and time pattern.

Among the non-physical pain categories, advertising treatments or drugs (commercial) were the most prevalent, with 8.99% (1955 out of the 21,741 total tweets collected) prevalence. Re-tweets had a similar prevalence (8.85%, 1923/21,741 tweets), followed by general discussion (6.72%, 1462/21,741 tweets) or third person’s migraine (2.05%, 445/21,741 tweets), and metaphor (1.20%, 261/21,741 tweets). A total of 5.23% of the tweets were inconclusive (1137/21,741 tweets), 1.99% not at all related (434/21,741 tweets), and 0.44% were blanks (96/21,741 tweets). Only 64.52% of all the collected tweets (14,028/21,741) posted using the word “migraine” were an actual self-report of physical pain suffering and other migraine-related symptoms (Figure 1). Therefore, we used the 14,028 tweets (re-tweets and repeated tweets from the same account were not used) for our descriptive statistical analyses below that actually represent self-reported ongoing migraine headache (64.52% of a total of 21,741 tweets).

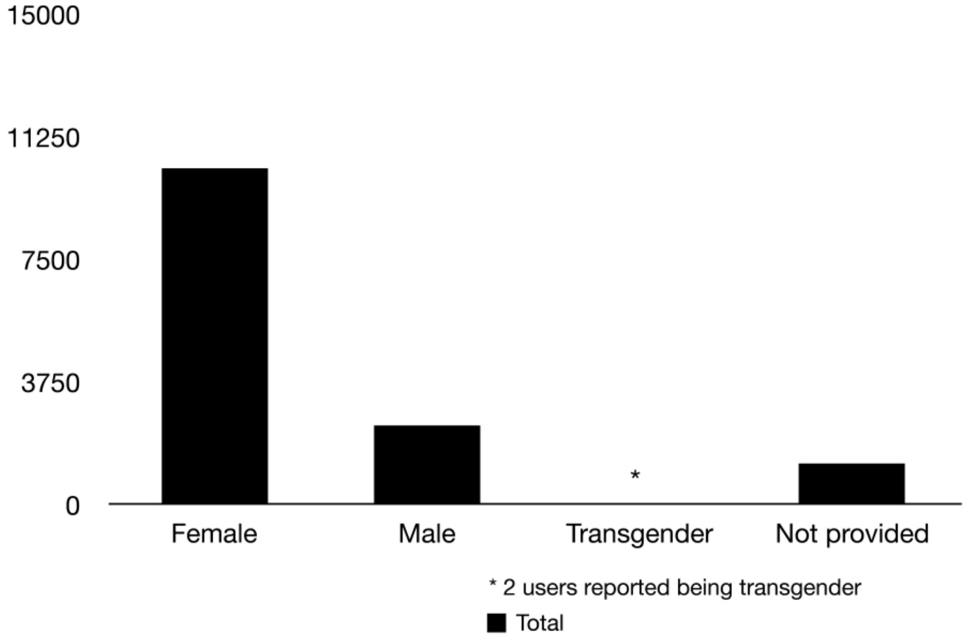
Figure 1. Classification of Tweets in Categories. Only 65% (14,028) of the 21,741 tweets were classified as self-reported migraine headache attacks.



Gender

Based on the self-reported gender, it was found that 73.47% of the 14,028 migraine headache tweets were from females (10,306 tweets), 17.40% were from males (2441 tweets), 0.01% from transgender (2 tweets), and 9.12% was not provided (1279 tweets). The presence of transgendered in the assessment reflects a new trend in research studies, where the possibility of free self-expression leads to a more accurate gender representation of our modern society and cohort (Figure 2).

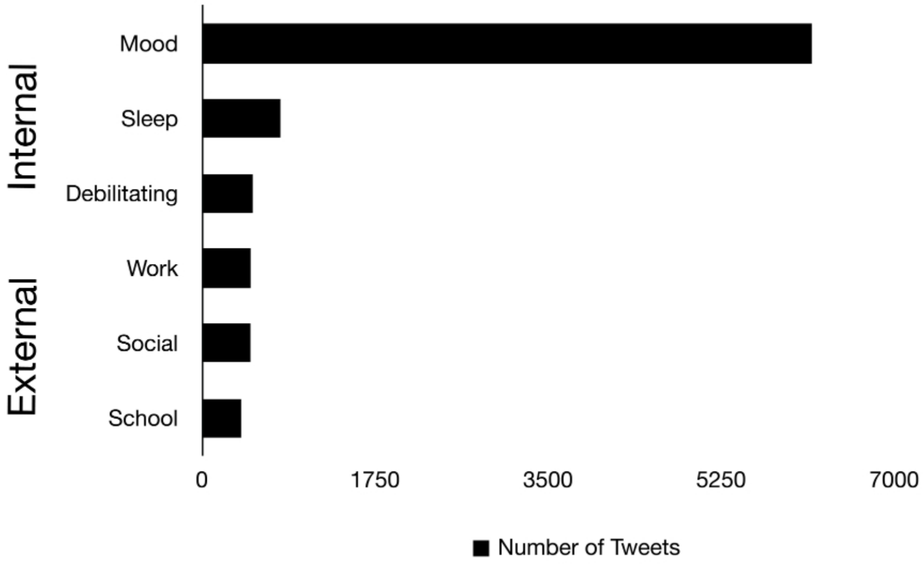
Figure 2. Gender distribution disclosed by users who reported their migraine headache attacks (n=14,028). 73.47% female (10,306 subjects), 17.40% male (2441 subjects), 0.01% (2 subjects) self-reported as transgender, and 9.12% was not provided (1279 subjects).



Migraine Impact

When possible, each interpretation of the tweets was also classified based on their migraine impact. The majority of the actual self-reported migraine headache posts generated an impact on patients' internal status: personal impact, mostly "mood" with 43.91% (defined as any changes in the natural and emotional state of mind of the individual) (6159 tweets out of the 14,028 migraine tweets), followed by an impact on "sleep" with 5.61% (meaning difficulties falling and staying sleep) (787 tweets). Another percentage related to personal impact included "debilitating" with 3.61% (defined as a physically incapacitating migraine headache) (507 tweets). The external ongoing impact of the self-reported migraine attacks (productivity impact) similarly and instantly affected "work" productivity with 3.46% (impact on work productivity and/or absenteeism) (486 tweets), "social" life with 3.45% (denoting influence and/or absenteeism in current social activities) (484 tweets), and finally "school" with 2.78% (impact on school productivity and/or absenteeism) (390 tweets). Missing data in this category was 37.18% (5215 tweets) (Figure 3).

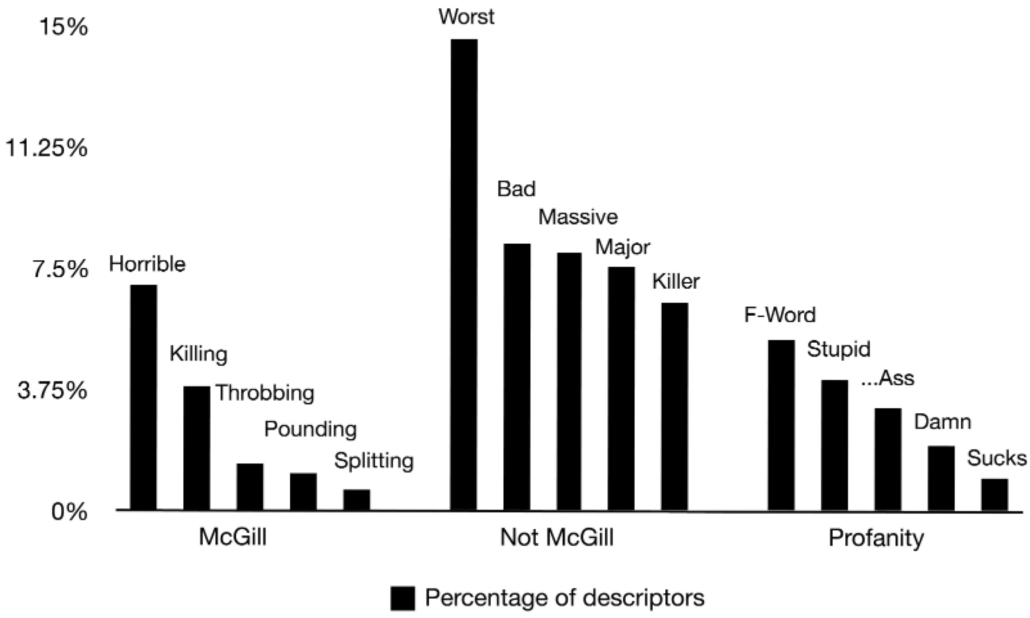
Figure 3. Impact and expression of migraine headache suffering (n=14,028). Personal impact (Internal) was predominantly on mood (43.91%, 6159 tweets). External impact was on productivity and absenteeism at work (3.46%, 486 tweets), social events (3.45%, 484 tweets), and school (2.78%, 390 tweets).



Pain Descriptors

Based on the physical tweets originally posted in English, we compared the adjectives used to describe a real ongoing migraine attack with the McGill Pain Questionnaire (MPQ) [28], one of the most widely used pain descriptor and rating questionnaires in research. In total, there were 242 descriptors used; however, there were only 45 English descriptors used a total of 1378 times. A prevalence of the word "horrible" was evident with 6.97% (used in 96 of the 1378 tweets with descriptors), followed by: "killing" (3.85%, used 53 times), "throbbing" (1.45%, 20 uses), "pounding" (1.16%, 16 uses), and "splitting" (0.65%, 9 uses). Conversely, subjects also expressed their migraine using words not included in MPQ. For classification purposes, we called these words "Not McGill". The most frequently expressed words in this category included: "worst" (14.59%; 201/1378 uses), "bad" (8.27%, 114 uses), "massive" (7.98%, 110 uses), "major" (7.55%, 104 uses), and "killer" (6.46%, 89 uses). Interestingly, profanity was also highly used to describe the ongoing migraine attack suffering, with the "F-word" being the most frequent in that category (5.30%, 73 uses). Furthermore, "stupid" (4.06%, 56 uses), "…ass" (3.19%, 44 uses), "damn" (2.03%, 28 uses), and "sucks" (1.02%, 14 uses) were also used (Figure 4).

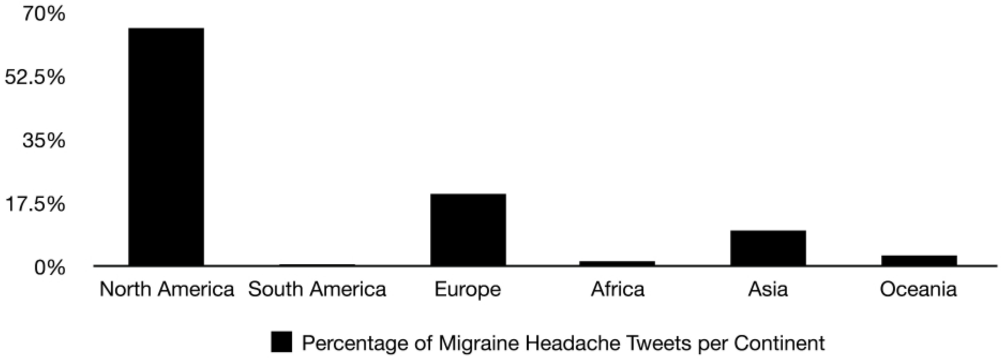
Figure 4. Most common pain descriptors used (n=1378). The most frequently used word from the McGill Pain Questionnaire (MPQ) was "horrible" (6.97%, 96 uses). Additional migraine headache adjectives (“Not McGill” words) included "worst" (14.59%, 201 uses) and profanity, the "F-word" (5.30%, 73 uses) being most frequently used.



Geographic Distribution

The collected database is geographically diverse, since it was composed of real-time tweets from all around the world. To better visualize and understand the origin of the messages, we divided the posts based on self-reported geographic location. When in doubt about the precise location, Google Maps [29] was used to help precisely locate the region using the Global Positioning System (GPS), if the user made the geographical coordinates publicly available. The vast majority of the tweets came from North America (65.57%, 3840 out of the 5856 Twitter users who reported their location), followed by Europe (19.89%, 1165 tweets), Asia (9.80%, 574 tweets), Oceania (2.92%, 171 tweets), Africa (1.33%, 78 tweets), South America (0.48%, 28 tweets), and last, Antarctica with no tweets reported. The United States alone represented 58.28% of the data (3413/5856 tweets) (Figure 5).

Figure 5. Percentage of migraine headache tweets by continent (n=5865). Most represented was North America with 65.57% (3840); United States alone represented 58.28% of overall data (3413 tweets).

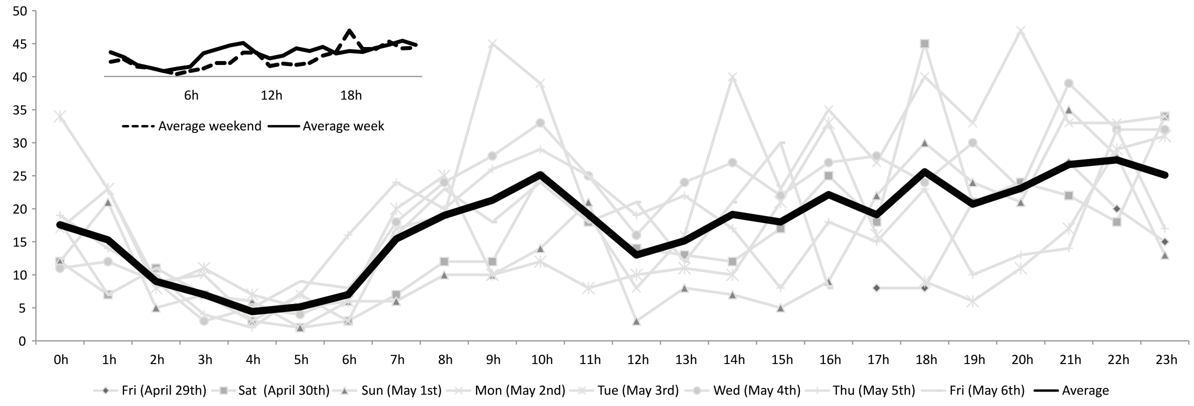


Temporal Pattern by Hour and Day of the Week

To have a better understanding of the online temporal behavior of sharing and expression of actual ongoing migraine headache suffering on the database, we organized each tweet based on the time and day when it was posted. As an initial step, we divided each global tweet by the day of the week it was posted. Our analysis demonstrated a higher global prevalence of self-reported migraine attack tweets on Tuesday (2559 of the 14,028 tweets) and Thursday (2155). It was followed by Wednesday (2074), Saturday (1933), Monday (1909), and Sunday (1752). A lower global prevalence of self-reported migraine headache tweets was observed on Friday (1646).

In an effort to improve our temporal data interpretation, the posts were divided according to the reported GMT, since it is the traditional method used by the Twitter website. By plotting the information on a timeline graphic, it was possible to observe a peak of the global migraine headache-related tweets at 14:00 GMT on Monday. This valuable information could easily lead to misleading interpretations if not adjusted for the original standard time of each specific geographic region where the tweets were generated. Consequently, we selected the United States (since it had the majority of tweets) and converted each single tweet to their particular local time. The United States observes DST during the spring season in several regions, which could also lead to erroneous interpretation if not corrected for this particular time when appropriate, during this period of the year. Ultimately, 9:00h and 20:00h on Monday across the United States were the actual peaks of prevalence when most Americans were sharing on social media the occurrence of their migraine headache attacks (Figure 6). When all weekday tweets, from Monday to Friday, were averaged based on the timeline, this first morning peak of migraine headache attack postings shifted rightward to 10:00h, and then from midday, it gradually and steadily peaked again at 22:00h. The weekend days, Saturday and Sunday, had a similar two-peak pattern of postings, though with a later morning peak at 11:00h, and an earlier and higher night peak of tweets at 18:00h (Figure 6).

Figure 6. Temporal patterns of migraine headache tweets in the United States. Tweets were converted to local times and corrected for daylight savings time. The averaged flow of migraine tweets accumulated at 10h, persisted and gradually peaked later at night (22h). During Saturday and Sunday (dashed line in the top-left graph), the highest peak occurred at 18h.



Discussion

Principal Findings

An unbiased evaluation of spontaneous sharing and expression of an ongoing migraine headache suffering is crucial for clinicians and researchers to understand the pattern of the disorder, and most importantly, the population under study. Here, we report the use of Twitter as a research tool to assess epidemiology and linguistics of migraine suffering in real-time in our modern society. Our results showed not only a significant overlap with other traditional epidemiologic studies, but also generated unique information about *who*, *what*, *how*, *where*, and *when* ongoing migraine headache suffering is shared on social media.

The methodology used here provided an effective, but laborious and time-consuming approach, to analyze reports of real-time migraine headache attacks on social media. This step was extremely important to avoid any sort of erroneous interpretation in our study. The use of current generalized algorithm search tools available on the Internet to effortlessly analyze altogether a sample of migraine tweets could inevitably lead to misleading conclusions, since they currently struggle to precisely exclude tweets that contain the word "migraine" used for commercial advertisement, general discussion, re-tweets, and metaphors [18]. In our database, only 64.52% of the migraine tweets (14,028) were an actual self-report of an ongoing migraine headache suffering. Therefore, all the analysis generated from this data was exclusively made using a sample of tweets from subjects reporting the occurrence of their own migraine attacks. Although we could not verify the accuracy of their diagnosis, in large computer-assisted telephone interview studies in the United States and abroad, individuals who call their headaches a migraine are about three times more likely to have a true migraine, based on the International Headache Society criteria [26], than those who are unaware of the type of the headaches they are suffering [30,31].

In this study, 73.47% of those who self-reported migraine headache were females (10,306 subjects of the 14,028 users who self-reported migraine) and 17.40% were males (2441 subjects). These results demonstrate and reinforce a higher prevalence of migraine headaches among females, which is consistent with population-based studies performed around the world [4,31]. Similar results have also been demonstrated in a recent study investigating the relative number of migraine searches and self-reports on Google and Twitter [18], respectively. Nonetheless, an in-depth evaluation of individual tweets was not provided in that particular study. The observed higher prevalence of migraine headaches in females could possibly be augmented by gender differences in expressing suffering or simply predilection of female users on social media. Nevertheless, real-time epidemiological studies that elaborate on point prevalence of pain/migraine and gender are lacking. Intriguingly, our study also shows that migraine headache suffering was prevalent in 0.01% (2 subjects) of Twitter users that freely self-described themselves as transgendered. Although minor in prevalence, the information for this gender group is not usually collected and investigated by current epidemiological studies, which highlights the importance of the inclusion of such gender identities in future research for better representation of migraine populations that fall outside the conventional female/male classification.

Migraine is a disabling form of primary headache [5], and the reports from patients on social media reflect what it impacts in life and society. Traditional (non-social media) epidemiological studies have emphasized the significant migraine-related impairment on productivity of routine and leisure activities [6,32]. However, the instant personal impact of migraine headache attacks freely reported in the tweets was predominantly on mood status (43.91%, 6159 of the 14,028 migraine headache tweets), which confirms the strong role of social media in affective expression [19] even when related to health issues. Second to mood status, the instant impact of migraine headache attacks was on sleep quality (dysfunctions to fall or stay asleep) on a minor level. This finding is consistent with previous studies that show a decrease in sleep quality of migraineurs [33,34]. Last, the external impact of the migraine headache attacks was immediate, as reported by migraine sufferers via tweets, and comparable on productivity and absenteeism at work, social events, and school (Figure 3). Since impact can be considered a qualitative variable, and possibly the most challenging category in this data analysis, all the students involved in this project received specific training and orientation. This was accomplished by calibration lectures based on real-time search and classification of tweets with the word migraine. All lectures in which this calibration was made were based on the classification guidelines of the International Headache Society [26]. In addition, each student was assigned a laboratory mentor in case of questions. All the training was performed prior to starting the classification, in order to ensure consistent categorization of the migraine impact.

What inevitably differentiates human from animal pain research is our ability to articulate the suffering experience and how we communicate it constantly evolves. In the case of migraine, the pulsating nature of the attacks has commonly been described in scientific literature as “throbbing”, and is even included in the International Headache Society criteria for migraine diagnosis [26]. Nonetheless, migraineurs freely defined their migraine headache attacks in multiple ways on Twitter. When restricted to the MPQ [28], a predetermined list of words and ratings widely used in medical contexts for measurement and assessment of pain, the most frequently used MPQ word in their tweets was “horrible”. It should be noted that “horrible” is used in the questionnaire to designate the pain intensity level and not pain quality. “Throbbing” was only the third MPQ word used, which was preceded by “killing”, an affective descriptive word. Additional migraine headache attack adjectives, categorized as “Not McGill” words in this study, were twice as prevalent as the MPQ words. “Worst” and “bad” were the leading migraine descriptors in this group, indicating mostly the severity differentiation by subjects of their ongoing attacks from the common ones. In addition, as the formality of doctor-patient communication is non-existent in Twitter, patients also felt understandably entitled to use profanity during actual suffering, with the “F-word” being the most frequently uttered to describe their migraine headache attacks. In summary, linguistics of ongoing migraine reports on social media proved to be highly heterogeneous and colloquial in our study, suggesting that current pain questionnaires should undergo constant reformulations to keep up with modernization in the expression of pain suffering in our society.

Real-time expression of migraine suffering occurs daily at the global level via social media. It is worth highlighting that it was at 14:00 GMT on the Monday that our planet had the highest flow of migraine headache attack postings on Twitter. The majority of those postings originated from North America, where the United States represented more than half of the total global stream. When each single tweet across the United States was converted to its particular local time, and corrected for DST during spring season when necessary, the largest point prevalence of migraine headache suffering was clustered on Monday with one peak in the morning at 9:00h and another one at 20:00h (Figure 6). As the 9:00h peak of tweets possibly reflected the onset of communication of migraine headache attacks in the early mornings [35], the second and gradual peak at night indicated that the duration of the migraine headache suffering, and its expression through social media, persisted and accumulated throughout the day. Curiously, general negative affect communications on social media also tend to rise throughout the day to a similar nighttime peak [19], and the emotional impact of migraine may potentiate the end of the working-day blues and vice-versa. In fact, when averaged with the other weekdays (Monday to Friday), the flow of migraine headache attack postings persisted and peaked even later at night. This changed during Saturday and Sunday, when people usually decompress from work-related stress and awaken later. Consequently, the initial morning peak of migraine headache attack postings during the weekend was further delayed; however, there was an earlier and high rise of ongoing migraine headache suffering postings at night, especially on Saturday. However, it is worth mentioning that the relatively short sample time may not provide the most accurate representation regarding the flow of migraine headache postings on social media.

Conclusion

This study showed that the spontaneous flow of communication on Twitter reflects multiple patterns of human interaction for sharing the ongoing suffering experience, and proved to be a rich and instant resource of knowledge regarding the actual impact of migraine attacks in our modern society.

Acknowledgments

We thank Ilkka Kristian Martikainen for early assistance in data classification, and the following professors and mentors at University of Michigan School of Dentistry (UMSoD): Robert M Bradley, Paul H Krebsbach, and William V Giannobile for valuable advice and encouragement. We would also like to acknowledge UMSoD (Under)Graduate Class of 2014 for their esteemed contribution as co-authors: A Berger, A Forest, A Green, A Petraszko, A Rohra, A Ruhlig, A Seal, B Corbett, B Jamo, B Kendziorski, B Mansfield, B Powell, D Coviak, D Fujawa, D Hammaker, D Patel, E Mencarelli, I McComb, J Birchmeier, J Harris, J Jeong, J Musselwhite, J Schell, J Sung, K Kilgore, K Swanson, L Bierwirth, L Ehardt, L Harness, L Koch, L Lungu, L McGarvey, L Persichetti, M Gariboy, M Hall, M Irani, M Lala, M Shallal-Ayzin, M Solverson, M Towler, N Poel, O Zorlu, P Dulay, P Iyer, P Kim, P Rutowski, R Abu-Zahra, R Kakar, R Mooar, S Garber, S Nematbakhsh, S Zafar, T Kwun, W Yahn (Undergraduate Students), C McWatters, B Royle, X Tang, R Wiesen (Graduate Students).

This work was supported by the following grants (DaSilva AF): National Institute of Health – National Institute of Neurological Disorders and Stroke – K23 NS062946, Dana Foundation's Brain and Immuno-Imaging Award, and the Migraine Research Foundation Research Grant Award.

Authors' Contributions

AFD, TDN, and MFD developed the study concept and design. TDN, MFD, MD, HvH, SL, CA, LK, MCB, and UMSoD (Under)Graduate Class of 2014 acquired the data. AFD and TDN analyzed and interpreted the data. AFD, TDN, and TD drafted the manuscript. JKZ, MD, HvH, SL, CA, LK, and MCB were responsible for administrative, technical, or material support. All authors had full access to all the data in the study. AFD is guarantor and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest

None declared.

References

1. Rasmussen BK. Epidemiology of migraine. Biomed Pharmacother 1995;49(10):452-455. [doi: 10.1016/0753-3322(96)82689-8] [Medline: 8746071]
2. Lipton RB, Stewart WF, Simon D. Medical consultation for migraine: results from the American Migraine Study. Headache 1998 Feb;38(2):87-96. [Medline: 9529763]
3. Lipton RB, Stewart WF, Diamond S, Diamond ML, Reed M. Prevalence and burden of migraine in the United States: data from the American Migraine Study II. Headache 2001 Aug;41(7):646-657. [Medline: 11554952]
4. Lipton RB, Bigal ME, Diamond M, Freitag F, Reed ML, Stewart WF, AMPP Advisory Group. Migraine prevalence, disease burden, and the need for preventive therapy. Neurology 2007 Jan 30;68(5):343-349. [doi: 10.1212/01.wnl.0000252808.97649.21] [Medline: 17261680]
5. Bigal ME, Serrano D, Reed M, Lipton RB. Chronic migraine in the population: burden, diagnosis, and satisfaction with treatment. Neurology 2008 Aug 19;71(8):559-566. [doi: 10.1212/01.wnl.0000323925.29520.e7] [Medline: 18711108]
6. Stewart WF, Lipton RB, Simon D. Work-related disability: results from the American migraine study. Cephalalgia 1996 Jun;16(4):231-8; discussion 215. [Medline: 8792034]
7. Kikuchi H, Yoshiuchi K, Miyasaka N, Ohashi K, Yamamoto Y, Kumano H, et al. Reliability of recalled self-report on headache intensity: investigation using ecological momentary assessment technique. Cephalalgia 2006 Nov;26(11):1335-1343. [doi: 10.1111/j.1468-2982.2006.01221.x] [Medline: 17059441]
8. Stone AA, Schwartz JE, Broderick JE, Shiffman SS. Variability of momentary pain predicts recall of weekly pain: a consequence of the peak (or salience) memory heuristic. Pers Soc Psychol Bull 2005 Oct;31(10):1340-1346. [doi: 10.1177/0146167205275615] [Medline: 16143666]
9. Erskine A, Morley S, Pearce S. Memory for pain: a review. Pain 1990 Jun;41(3):255-265. [Medline: 1697054]
10. Stone AA, Broderick JE. Real-time data collection for pain: appraisal and current status. Pain Med 2007 Oct;8 Suppl 3:S85-S93. [doi: 10.1111/j.1526-4637.2007.00372.x] [Medline: 17877531]
11. Gendreau M, Hufford MR, Stone AA. Measuring clinical pain in chronic widespread pain: selected methodological issues. Best Pract Res Clin Rheumatol 2003 Aug;17(4):575-592. [Medline: 12849713]
12. 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]
13. Twitter Inc. URL: https://twitter.com/ [accessed 2014-02-27] [WebCite Cache ID 6NhvbiSAT]
14. Hanson CL, Cannon B, Burton S, Giraud-Carrier C. An exploration of social circles and prescription drug abuse through Twitter. J Med Internet Res 2013;15(9):e189 [FREE Full text] [doi: 10.2196/jmir.2741] [Medline: 24014109]
15. Myslín M, Zhu SH, Chapman W, Conway M. Using Twitter to examine smoking behavior and perceptions of emerging tobacco products. J Med Internet Res 2013;15(8):e174 [FREE Full text] [doi: 10.2196/jmir.2534] [Medline: 23989137]
16. Hingle M, Yoon D, Fowler J, Kobourov S, Schneider ML, Falk D, et al. Collection and visualization of dietary behavior and reasons for eating using Twitter. J Med Internet Res 2013;15(6):e125 [FREE Full text] [doi: 10.2196/jmir.2613] [Medline: 23796439]

17. Heaivilin N, Gerbert B, Page JE, Gibbs JL. Public health surveillance of dental pain via Twitter. J Dent Res 2011 Sep;90(9):1047-1051 [FREE Full text] [doi: 10.1177/0022034511415273] [Medline: 21768306]
18. Linnman C, Maleki N, Becerra L, Borsook D. Migraine tweets - what can online behavior tell us about disease? Cephalalgia 2013 Jan;33(1):68-69 [FREE Full text] [doi: 10.1177/0333102412465207] [Medline: 23136208]
19. Golder SA, Macy MW. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. Science 2011 Sep 30;333(6051):1878-1881 [FREE Full text] [doi: 10.1126/science.1202775] [Medline: 21960633]
20. Dodds PS, Harris KD, Kloumann IM, Bliss CA, Danforth CM. Temporal patterns of happiness and information in a global social network: hedonometrics and Twitter. PLoS One 2011;6(12):e26752 [FREE Full text] [doi: 10.1371/journal.pone.0026752] [Medline: 22163266]
21. Bell G, Hey T, Szalay A. Computer science. Beyond the data deluge. Science 2009 Mar 6;323(5919):1297-1298. [doi: 10.1126/science.1170411] [Medline: 19265007]
22. Twitter Inc. URL: https://blog.twitter.com/2011/one-hundred-million-voices [accessed 2014-02-27] [WebCite Cache ID 6NhtroyWi]
23. Twitter Inc. URL: https://blog.twitter.com/2011/200-million-tweets-day [accessed 2014-02-27] [WebCite Cache ID 6NhuAgoZB]
24. Twitter Inc. URL: https://business.twitter.com/whos-twitter [accessed 2014-02-27] [WebCite Cache ID 6NhwGLCpH]
25. Twitter Inc. URL: https://twitter.com/search-home [accessed 2014-02-27] [WebCite Cache ID 6NhwNtSrJ]
26. Headache Classification Subcommittee of the International Headache Society. The International Classification of Headache Disorders: 2nd edition. Cephalalgia 2004;24 Suppl 1:9-160. [Medline: 14979299]
27. US Department of Commerce. US Census Bureau database. URL: http://www.census.gov/ [accessed 2014-02-25] [WebCite Cache ID 6NeOVfnY5]
28. Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. Pain 1975 Sep;1(3):277-299. [Medline: 1235985]
29. Google Maps. URL: https://www.google.com/maps/preview [accessed 2014-02-27] [WebCite Cache ID 6NhwYiQrg]
30. Lipton RB, Stewart WF, Liberman JN. Self-awareness of migraine: interpreting the labels that headache sufferers apply to their headaches. Neurology 2002 May 14;58(9 Suppl 6):S21-S26. [Medline: 12011270]
31. Stovner L, Hagen K, Jensen R, Katsarava Z, Lipton R, Scher A, et al. The global burden of headache: a documentation of headache prevalence and disability worldwide. Cephalalgia 2007 Mar;27(3):193-210. [doi: 10.1111/j.1468-2982.2007.01288.x] [Medline: 17381554]
32. Stang PE, Osterhaus JT. Impact of migraine in the United States: data from the National Health Interview Survey. Headache 1993 Jan;33(1):29-35. [Medline: 8436495]
33. Seidel S, Hartl T, Weber M, Matterey S, Paul A, Riederer F, PAMINA Study Group. Quality of sleep, fatigue and daytime sleepiness in migraine - a controlled study. Cephalalgia 2009 Jun;29(6):662-669. [doi: 10.1111/j.1468-2982.2008.01784.x] [Medline: 19210514]
34. Engstrøm M, Hagen K, Bjørk MH, Stovner LJ, Gravdahl GB, Stjern M, et al. Sleep quality, arousal and pain thresholds in migraineurs: a blinded controlled polysomnographic study. J Headache Pain 2013 Dec;14(1):12 [FREE Full text] [doi: 10.1186/1129-2377-14-12] [Medline: 23565669]
35. Fox AW, Davis RL. Migraine chronobiology. Headache 1998 Jun;38(6):436-441. [Medline: 9664747]

Abbreviations

DST: Daylight Savings Time
GMT: Greenwich Mean Time
MPQ: McGill Pain Questionnaire

Edited by G Eysenbach; submitted 22.01.14; peer-reviewed by C Linnman, N Maleki; comments to author 12.02.14; revised version received 03.03.14; accepted 13.03.14; published 03.04.14.

Please cite as:
Nascimento TD, DosSantos MF, Danciu T, DeBoer M, van Holsbeeck H, Lucas SR, Aiello C, Khatib L, Bender MA, UMSoD (Under)Graduate Class Of 2014, Zubieta JK, DaSilva AF
Real-Time Sharing and Expression of Migraine Headache Suffering on Twitter: A Cross-Sectional Infodemiology Study
J Med Internet Res 2014;16(4):e96
URL: http://www.jmir.org/2014/4/e96/
doi:10.2196/jmir.3265
PMID:24698747

Original Paper

Performance of eHealth Data Sources in Local Influenza Surveillance: A 5-Year Open Cohort Study

Toomas Timpka[1], MD, PhD; Armin Spreco[1], MSc; Örjan Dahlström[2], PhD; Olle Eriksson[3], PhD; Elin Gursky[4], PhD; Joakim Ekberg[1], PhD; Eva Blomqvist[3], PhD; Magnus Strömgren[5], PhD; David Karlsson[3], MSc; Henrik Eriksson[3], PhD; James Nyce[6], PhD; Jorma Hinkula[7], PhD; Einar Holm[5], PhD

[1]Department of Medical and Health Sciences, Linköping University, Linköping, Sweden
[2]Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden
[3]Department of Computer and Information Science, Linköping University, Linköping, Sweden
[4]National Strategies Support Directorate, ANSER/Analytic Services Inc., Arlington, VA, United States
[5]Department of Geography and Economic History, Umeå University, Umeå, Sweden
[6]Department of Anthropology, Ball State University, Muncie, IN, United States
[7]Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden

Corresponding Author:
Toomas Timpka, MD, PhD
Department of Medical and Health Sciences
Linköping University
Linköping University Hospital Campus
Linköping, SE58183
Sweden
Phone: 46 101030000
Fax: 46 101031865
Email: toomas.timpka@liu.se

Abstract

Background: There is abundant global interest in using syndromic data from population-wide health information systems—referred to as eHealth resources—to improve infectious disease surveillance. Recently, the necessity for these systems to achieve two potentially conflicting requirements has been emphasized. First, they must be evidence-based; second, they must be adjusted for the diversity of populations, lifestyles, and environments.

Objective: The primary objective was to examine correlations between data from Google Flu Trends (GFT), computer-supported telenursing centers, health service websites, and influenza case rates during seasonal and pandemic influenza outbreaks. The secondary objective was to investigate associations between eHealth data, media coverage, and the interaction between circulating influenza strain(s) and the age-related population immunity.

Methods: An open cohort design was used for a five-year study in a Swedish county (population 427,000). Syndromic eHealth data were collected from GFT, telenursing call centers, and local health service website visits at page level. Data on mass media coverage of influenza was collected from the major regional newspaper. The performance of eHealth data in surveillance was measured by correlation effect size and time lag to clinically diagnosed influenza cases.

Results: Local media coverage data and influenza case rates showed correlations with large effect sizes only for the influenza A (A) pH1N1 outbreak in 2009 (r=.74, 95% CI .42-.90; P<.001) and the severe seasonal A H3N2 outbreak in 2011-2012 (r=.79, 95% CI .42-.93; P=.001), with media coverage preceding case rates with one week. Correlations between GFT and influenza case data showed large effect sizes for all outbreaks, the largest being the seasonal A H3N2 outbreak in 2008-2009 (r=.96, 95% CI .88-.99; P<.001). The preceding time lag decreased from two weeks during the first outbreaks to one week from the 2009 A pH1N1 pandemic. Telenursing data and influenza case data showed correlations with large effect sizes for all outbreaks after the seasonal B and A H1 outbreak in 2007-2008, with a time lag decreasing from two weeks for the seasonal A H3N2 outbreak in 2008-2009 (r=.95, 95% CI .82-.98; P<.001) to none for the A p H1N1 outbreak in 2009 (r=.84, 95% CI .62-.94; P<.001). Large effect sizes were also observed between website visits and influenza case data.

Conclusions: Correlations between the eHealth data and influenza case rates in a Swedish county showed large effect sizes throughout a five-year period, while the time lag between signals in eHealth data and influenza rates changed. Further research is needed on analytic methods for adjusting eHealth surveillance systems to shifts in media coverage and to variations in age-group related immunity between virus strains. The results can be used to inform the development of alert-generating eHealth surveillance systems that can be subject for prospective evaluations in routine public health practice.

(J Med Internet Res 2014;16(4):e116) doi:10.2196/jmir.3099

KEYWORDS

influenza; infectious disease surveillance; Internet; eHealth; Google Flu Trends; telenursing call centers; website usage; open cohort design; public health

Introduction

There has been abundant global interest in the use of interactive health information technology—referred to as eHealth systems—to improve the effectiveness of infectious disease surveillance [1]. However, similar to other eHealth applications, surveillance systems based on eHealth data must live up to two potentially conflicting requirements: they must be both evidence-based and adapted to how people live within their specific environments [2]. In other words, eHealth systems' development, specification, and evaluation in any setting are dependent on the infrastructure, habits, and culture in that setting and at that point in time. This fact must be taken in regard when transferring an eHealth system for use at other locations and periods of time. An example of eHealth technology employed in infectious disease surveillance is Google Flu Trends (GFT), an Internet-based software system that uses aggregated data from the Google search engine to estimate influenza activity [3]. Early studies comparing GFT data to census region influenza-like illness (ILI) data in the United States [4] and prospectively collected sentinel data from two systems in Australia demonstrated strong correlations [5]. However, similar evaluations performed during the influenza A (A) pH1N1 circulation in 2009 in New Zealand [6], Singapore [7], and the United States [8] reported inconsistencies between the GFT and ILI data. One hypothetical reason for such inconsistencies is that mass media coverage of an influenza outbreak can influence behavior by motivating the layperson to seek additional information [9]. Another possible reason may be that because the proportion of adolescents and young adult cases varies between influenza seasons [10], age-related information technology and Internet use consequently impacts eHealth surveillance performance [11].

The purpose of this study is to support evidence-based strategies for eHealth system development in infectious disease control. The primary study objective is to examine correlations between GFT data, telenursing call data, health service webpage usage data, and influenza case rates during seasonal and pandemic influenza outbreaks. The secondary objectives are to investigate associations between eHealth data, the media coverage of influenza outbreaks, and the interaction between the circulating influenza strain(s) and age-related variations in population immunity. To avoid climate and sociogeographic factors affecting the analyses [12,13], all data were collected from one Swedish county (Östergötland, population 427,000) located in South-East Sweden. The entire county population is covered by an electronic health data repository maintained by the county council to systematically and continuously insure the quality of service [14]. The repository collects data from all health care encounters provided in the county at primary and secondary levels, as well as eHealth data from calls made by the county residents to the nation-wide telenursing service, GFT outputs relevant for the county, and data from visits at the county council website (Table 1).

Table 1. eHealth systems in Östergötland county, Sweden, investigated in the study.

eHealth systems	Description
Google Flu Trends (GFT)	The GFT service was launched by the Web search engine provider Google in 2008 to track changes in the volume of online search queries related to influenza or its symptoms [3]. For Sweden, the GFT data on Web queries are derived at country and regional levels from a pool of search terms that relate to symptoms, remedies, and complications of influenza and generate a trend that closely correlates data on ILI.
Swedish "Healthcare Direct/1177" telenursing service	Telenursing is defined as computer-supported call centers staffed by registered nurses who perform counselling and patient triage as a means of augmenting self-care support and regulating patient access to medical services [15]. The Swedish national telenursing system "Healthcare Direct" is a 24/7, nurse-led, telephone advice service with one country-wide phone number (1177). Specially-trained nurses use a computerized decision-aid program and an Electronic Health Record (EHR) system for every call. After each call, a chief complaint from a fixed-field terminology register is recorded in the EHR.
Swedish "Healthcare Direct/1177" Internet health information service	"Healthcare Direct/1177" also maintains a national Internet-based health information service, with a specific website for each participating county council. This service consists of general information pages, arranged according to topics such as symptom evaluation guidelines and disease facts and self-management information. Each website is also connected to a Web traffic analysis facility, which at the time of the study was Google Analytics (GA).

Methods

Study Design

The study used an “open cohort” design based on the total population of Östergötland county. Open cohort denotes that new cohort members were included by birth or moving into the county and other members were excluded when passing away or moving out from the county as the cohort follow-up progressed. To update the open study cohort, annual aggregated data on the sex, age, and residence of the county population were collected each year from Statistics Sweden. In accordance with Swedish legislation (SFS 2008:355), personal identifiers were removed from the records. The start and end time of an influenza outbreak was defined as 8 incident ILI cases diagnosed in the county during a floating seven-day period. The study design was approved by the Regional Research Ethics Board in Linköping (dnr. 2012/104-31).

Data Collection

Data on clinical influenza cases and eHealth data were collected between November 2007 and April 2012 using the electronic health data repository maintained by the county council. Data from the clinical laboratories were, for this study, collected during the period 2009-01-01 to 2010-09-15. Influenza cases were identified by the ICD-10 codes for influenza (J10.0, J10.1, J10.8, J11.0, J11.1, J11.8). Influenza-related telenursing calls were identified by the chief complaint codes associated with influenza symptoms: dyspnea, fever (child, adult), cough (child, adult), sore throat, lethargy, syncope, dizziness, and headache (child, adult), from the fixed-field terminology register. GFT data for the study period were collected using a Google account to download data on Google searches from Östergötland county on seasonal and pandemic (for the 2009-2010 outbreak) influenza to a database. The downloaded dataset did not consist of absolute search rate data, but consisted of influenza Web search data normalized with regard to total Web search volumes by the GFT software. Usage data from the county council webpages were collected beginning in May 2009. For technical reasons associated with a change of software providers, data could not be retrieved for the 2010-2011 influenza season. Usage data from January 2012 for the Web-based information service, measured by the numbers of visits of a certain type of page, were collected by directly accessing the Google Analytics (GA) Web traffic analysis instances and by retrieving data through its application programing interface. The Web traffic data contain information about the location of the Web user based on the IP address of the user’s computer (at the granularity of counties). Filters can also be applied based on keywords in page titles or page addresses (URLs) or by pre-selecting certain URLs. Page type refers to the kind of content the page contains, such as factual information about influenza, self-care information, frequently asked questions and answers, or news pages, respectively. Data on media coverage of influenza outbreaks were collected from the online database of the largest newspaper in the county (Östgöta Correspondenten). The database was searched for articles with the term “influenza” (influensa) for the period between November 2007 and April 2012.

Data Analysis

The influenza case data defined by clinical diagnoses were validated against case data from the microbiological laboratories for the period 2009-01-01 to 2010-09-15. In these analyses, both datasets were separately adjusted for week-day effects on care resource utilization. The correlations between the number of cases reported each day from the clinical and laboratory data were analyzed with a 0–6 day lag. Thereafter, to analyze the relative distribution of influenza cases between age groups, the Relative Illness Ratio (RIR), which is the ratio of the percentage of individuals with an influenza diagnosis in a given age group to the percentage of the general population belonging to the same age group, was computed for each age group and outbreak (circulating virus type) using the formula:

$$RIR_i = (C_i / C) / (N_i / N)$$

where C_i is the number of influenza cases in age group i, C is the number of influenza cases in total, N_i is the population in age group i, and N is the total population in Östergötland county. Further, 95% confidence intervals were calculated for each RIR, using a method based on normal approximation of the Poisson distribution.

In the main analyses of associations between eHealth and influenza case data, Pearson’s correlation coefficients (*r*) were examined to compare influenza case rates with the eHealth data sources (ie, GFT data and all possible combinations of telenursing chief complaints and website page visits with a 2-week time lag to influenza case rates). The three groupings of chief complaints and combinations of website page types, respectively, with the strongest correlation to the influenza case rate for each time lag were listed. The chief complaint grouping and website page combination with the largest correlation effect size were chosen to be used in the final analyses. Separate analyses were performed of correlations between media reports (weekly rate of articles in the regional newspaper mentioning influenza), influenza case rates, and the eHealth data sources (GFT data and all possible combinations of telenursing chief complaints and website page visits with a 2-week time lag to media reports) respectively. The level of statistical significance was set to $P<.05$. To denote the strength of correlations, limit values were applied as suggested by the Cohen Scale [16]. This scale defines small, medium, and large effect sizes as .10, .30, and .50 respectively. The analyses were performed using SPSS version 19, R Statistical Software version 2.15.2, and Minitab Statistical Software version 16.1.1.

Results

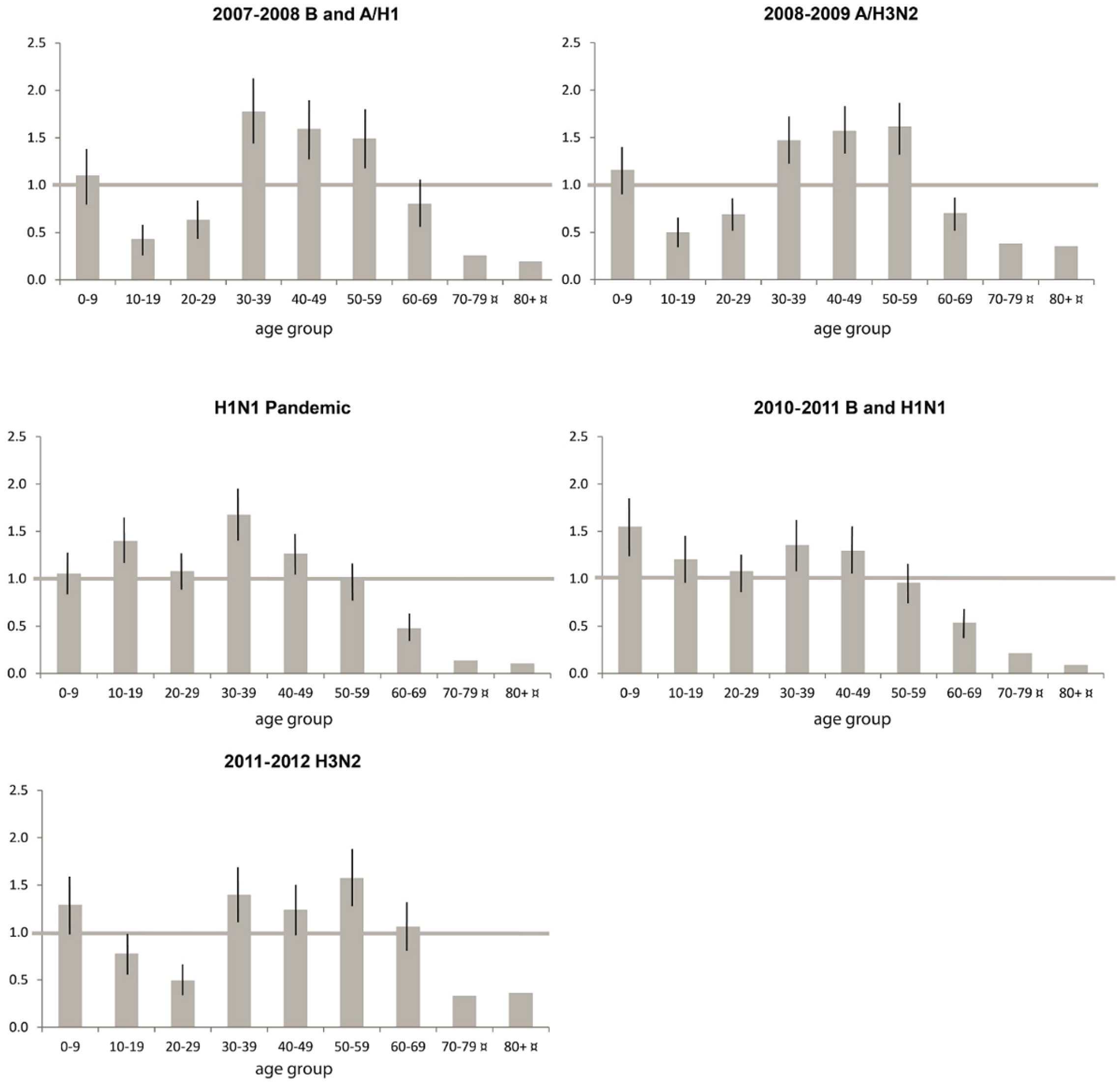
Overview

The results from the validation analyses showed correlations with large effect sizes between the number of clinically diagnosed influenza cases per day and the corresponding number of cases verified daily by microbiological analyses during the validation period. The correlation with largest effect size ($r=.63$, $P<.001$) was observed between the clinically and the microbiologically verified cases with a two-day lag.

The five-year study period covered four winter influenza seasons and one pandemic outbreak; winter influenza seasons occurred between 2008-01-21 to 2008-04-30 (B and A H1) and 2008-12-24 to 2009-03-30 (A H3N2), the pandemic outbreak lasted from 2009-08-21 to 2009-12-22 (A pH1N1), and the two winter influenza seasons occurring after the pandemic lasted from 2010-12-21 to 2011-04-21 (B and A pH1N1) and 2012-01-09 to 2012-04-14 (A H3N2).

The relative infection ratios for the different age groups and outbreaks are displayed in Figure 1. Higher-than-expected proportions of cases were distributed in the middle-aged groups (30-39 and 40-49 years) during all outbreaks, while lower-than-expected proportions of cases among adolescents and young adults (10-19 and 20-29 years) were recorded for those winter influenza seasons when the pandemic A pH1N1 virus was not circulating.

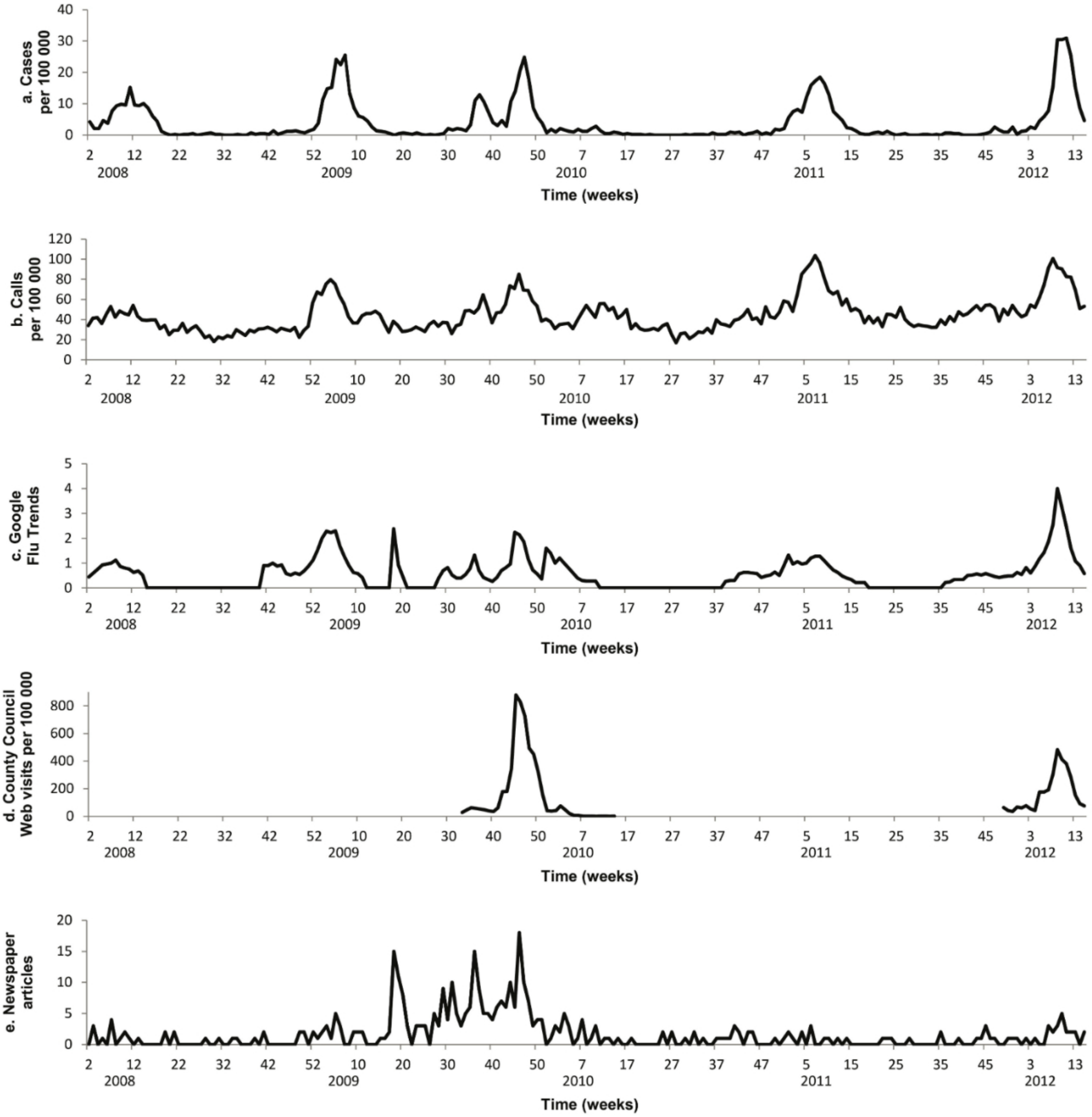
Figure 1. Relative infection ratios (RIRs) with 95% confidence intervals for influenza outbreaks between 2007 and 2012 in Östergötland county displayed by decennial age groups. ¤ Too few observations to allow statistical analysis.



Correlations Between Local Media Coverage, Influenza Rates, and eHealth Data

The correlations between local media coverage data and influenza case rates showed large effect sizes only for the A pH1N1 outbreak in 2009 (r=.74, 95% CI .42-.90; P<.001), and the severe seasonal A H3N2 outbreak in 2011-2012 (r=.79, 95% CI .42-.93; P=.001). For both outbreaks, media coverage preceded case rates by one week. In addition, media reports about influenza showed a peak for weeks 18-22 of 2009 that coincided with a sharp increase in GFT activity, but these peaks had no correspondence with influenza rates or telenursing data (Figure 2). Correlations between media coverage and GFT showed large effect sizes for the seasonal outbreak in 2008-2009 (r=.62, 95% CI .15-.86; P=.014), the A pH1N1 outbreak in 2009 (r=.69, 95% CI .35-.87; P=.001), and the seasonal A H3N2 outbreak in 2011-2012 (r=.77, 95% CI .39-.93; P=.002). The strongest correlations were found for no time lag except for the seasonal outbreak in 2012, when GFT activity preceded media coverage by one week. Neither telenursing data nor the data from health service provider webpages showed statistically significant correlations with the local media coverage data.

Figure 2. Display of (a) daily rates of influenza cases, (b) daily rates of telenursing calls for indicator chief complaints (fever and syncope), (c) Google Flu Trends output, (d) Influenza-specific website usage at local health service provider, and (e) articles mentioning influenza in major regional newspaper. All data were collected from Östergötland County, Sweden, from November 2007 to April 2012.



Correlations Between GFT and Influenza Case Data

The correlations between GFT and influenza case data showed large effect sizes for all outbreaks, varying between *r*=.69 (95% CI .22-.90), *P*=.010, for the B and A H1 outbreak in 2007-2008 to *r*=.96 (95% CI .88-.99), *P*<.001, for the seasonal A H3N2 outbreak in 2008-2009 (Table 2). The time lag between GFT and influenza case data decreased from two weeks during the first outbreaks to one week from the 2009 A pH1N1 pandemic, with GFT data preceding influenza case data.

Table 2. Associations on a weekly basis between GFT data and influenza case data displayed by the correlation coefficient *r* (95% CI), for the five influenza outbreaks observed in Östergötland county, Sweden, during the study period 2007-2012.

Outbreak time lag (weeks)	2007-2008 B and A H1 (15 weeks) *r* (95% CI)	2008-2009 A H3N2 (15 weeks) *r* (95% CI)	2009 A pH1N1 (19 weeks) *r* (95% CI)	2010-2011 B and A pH1N1 (18 weeks) *r* (95% CI)	2011-2012 A H3N2 (14 weeks) *r* (95% CI)
0	ns[c]	.66 (.23-.88) *P*=.007	.79 (.53-.92) *P*<.001	.57 (.14-.82) *P*=.013	.83 (.54-.95) *P*<.001
1[a]	ns	.86 (.61-.96) *P*<.001	.92 (.79-.97) *P*<.001	.75 (.42-.90) *P*=.001	.95 (.83-.98) *P*<.001
2[b]	.69 (.22-.90) *P*=.010	.96 (.88-.99) *P*<.001	.69 (.31-.88) *P*=.002	.81 (.53-.93) *P*<.001	.83 (.50-.95) *P*=.001

[a] Time lag 1 week=Influenza diagnoses 1-week time shift, ie, people first Google the terms “influenza” or “swine flu” and 1 week later visit the health services.

[b] Time lag 2 weeks=Influenza diagnoses 2-week time shift, ie, people first Google the terms “influenza” or “swine flu” and 2 weeks later visit the health services.

[c] ns=not statistically significant

Correlations Between Telenursing Call Data and Influenza Case Data

The correlations between telenursing data and influenza case data showed large effect sizes for all outbreaks except for the seasonal B and A H1 outbreak in 2007-2008. The preceding time lag for the optimal correlation changed from two weeks for the seasonal A H3N2 outbreak in 2008-2009 to none for the A p H1N1 outbreak in 2009 and one week for the most recent two seasonal outbreaks (Table 3). The telenursing chief complaints included in combinations showing the correlations with largest effect size to influenza case data during most seasons were fever (child, adult) and syncope.

Table 3. Associations on a weekly basis between telenursing call data and influenza case data displayed by the correlation coefficient *r* (95% CI), for the five influenza outbreaks observed in Östergötland county, Sweden, during the study period 2007-2012.

Outbreak time lag (weeks)	2007-2008 B and A H1 (15 weeks) *r* (95% CI)	2008-2009 A H3N2 (15 weeks) *r* (95% CI)	2009 A pH1N1 (19 weeks) *r* (95% CI)	2010-2011 B and A pH1N1 (18 weeks) *r* (95% CI)	2011-2012 A H3N2 (14 weeks) *r* (95% CI)
0	ns[c]	ns	.84 (.62-.94) *P*<.001	.91 (.77-.97) *P*=.001	.90 (.70-.97) *P*=.008
1[a]	ns	.81 (.48-.94) *P*=.001	.80 (.52-.92) *P*<.001	.95 (.86-.98) *P*<.001	.97 (.91-.99) *P*<.001
2[b]	ns	.95 (.82-.98) *P*<.001	ns	.88 (.69-.96) *P*=.001	.93 (.77-.98) *P*=.002

[a] Time lag 1 week=Influenza diagnoses 1-week time shift, ie, people first call Healthcare Direct/1177 and 1 week later visit the health services.

[b] Time lag 2 weeks=Influenza diagnoses 2-week time shift, ie, people first call Healthcare Direct/1177 and 2 weeks later visit the health services.

[c] ns=not statistically significant

Correlations Between Log Data From County Council Website and Influenza Case Data

During the 2009 A pH1N1 pandemic, the correlation with the largest effect size between log data from the county council website and influenza case data (*r*=.75, 95% CI .45-.90; *P*=.004) was noted for the website section “Influenza self-care and treatment” with no time lag, while the correlation with one week preceding lag (website visit followed by health care visit) for data from the section “Influenza facts” was only slightly smaller (*r*=.74, 95% CI .42-.90; *P*=.006). For the seasonal A H3N2 outbreak in 2012, the correlation with the largest effect sizes were noted for log data from the website section “Influenza self-care and treatment” with no time lag to influenza cases (*r*=.94, 95% CI .84-.98; *P*<.001) and “Influenza facts” with one week preceding lag (*r*=.94, 95% CI .84-.98; *P*<.001).

Correlations Between GFT Data, Telenursing Data, and Log Data From County Council Website

The correlations between GFT data and telenursing data showed large effect sizes for all outbreaks (Table 4). During the seasonal outbreaks in 2007-2008 and 2008-2009, the telenursing data preceded the GFT data by one week, while during the A pH1N1 in 2009 and the following seasonal outbreaks the GFT data either preceded the telenursing data by one week or corresponded in time. The correlations between GFT data and log data from the county council website showed large effect sizes for the two outbreaks for which data were available. For both the A pH1N1 outbreak in 2009 (r=.87, 95% CI .69-.95; P<.001) and the seasonal A H3N2 outbreak in 2011-2012 (r=.96, 95% CI .90-.99; P<.001), the data sources showed optimal correlation when no time lag was introduced. Also, the correlations between telenursing data and log data from county council website showed large effect sizes for the outbreaks for which data were available. For the 2009 A pH1N1 outbreak, the data sources showed optimal correlation (r=.88, 95% CI .71-.95; P<.001) when no time lag was introduced, while for the seasonal A H3N2 outbreak in 2011-2012 the optimal correlation (r=.95, 95% CI .84-.99; P<.001) was observed when telenursing data was relocated to precede the log data from county council website with one week lag.

Table 4. Associations on a weekly basis between GFT data and telenursing call data displayed by the correlation coefficient *r* (95% CI), for the five influenza outbreaks observed in Östergötland county, Sweden, during the study period 2007-2012.

Outbreak time lag	2007-2008	2008-2009	2009	2010-2011	2011-2012
	B and A H1	A H3N2	A pH1N1	B and A pH1N1	A H3N2
(weeks)	(15 weeks)	(15 weeks)	(19 weeks)	(18 weeks)	(14 weeks)
	r (95% CI)	r (95% CI)	r (95% CI)	r (95% CI)	r (95% CI)
−1[a]	.88 (0.65-0.96) P=.012	.92 (0.77-0.98) P=.001	ns	ns	.90 (0.69-0.97) P=.011
0	ns[d]	.88 (0.68-0.96) P=.008	.77 (0.49-0.91) P=.034	.85 (0.63-0.94) P=.005	.94 (0.83-0.98) P=.001
1[b]	ns	ns	.87 (0.68-0.95) P=.001	.94 (0.83-0.98) P<.001	.87 (0.60-0.96) P=.032
2[c]	ns	ns	ns	.81 (0.53-0.93) P=.016	.86 (0.56-0.96) P=.040

[a] Time −1 week=Healthcare Direct/1177 1-week time shift, ie, people first call Healthcare Direct/1177 and then use GFT one week later.

[b] Time lag 1 week=telenursing data 1-week time shift, ie, people first use GFT and then call Healthcare Direct/1177 one week later.

[c] Time lag 2 weeks=telenursing data 2-week time shift, ie, people first use GFT and then call Healthcare Direct/1177 two weeks later.

[d] ns=not statistically significant

Discussion

Principal Findings

The primary objective of this study was to examine the performance of data from GFT, telenursing call centers, and health service provider websites in influenza surveillance, while a secondary objective was to investigate associations between eHealth data, media coverage, and the interaction between circulating influenza strain(s) and age-related population immunity. We found correlations with large effect sizes between data from these eHealth sources and influenza case rates for both seasonal and pandemic outbreaks, with the exception of telenursing data during the seasonal B and A H1 outbreak in 2007-2008. A utilization study of the Swedish telenursing service reported that young adults living independently constituted a large group of callers [17]. A contributing explanation for the inferior performance of the telenursing data in 2007-2008 can therefore be that this winter influenza season comprised only a small proportion of young adults (Figure 1), which may have led to comparatively fewer individuals with influenza symptoms contacting the telenursing service. Regarding GFT, the findings are consistent with previous studies conducted at the national and state levels that have reported correlations with large effect sizes between GFT and ILI case data [18-20] and a recent study that reported large effect size correlations with ILI case rates at the local level [21]. Similar to the telenursing data, GFT showed in this study lower correlations with influenza case rates for the 2007-2008 and 2008-2009 winter seasons when comparatively fewer young adults were diagnosed with influenza. Interestingly, we also found that website usage data specified at page level (as compared to search query data) from the local health service provider showed a performance similar to GFT for the two influenza seasons for which these data were collected. Importantly, the webpages showing the best performance contained information about self-care and influenza facts, rather than general outbreak updates or materials concerning vaccination.

Although the correlations between the eHealth data and influenza case rates showed large effect sizes throughout the study period, the time lag between signals in eHealth data and increases in case rates changed. Our results thereby indicate important distinctions in the performance of eHealth systems for influenza surveillance. While the eHealth data tended to precede influenza rates by two weeks during the first two seasonal outbreaks, the time lag was reduced to one week or none from the pandemic outbreak in 2009 onwards. These findings correspond with previous studies [6,7], which concluded that eHealth data associated with infections emerge not only from personal need, but also from an associated general interest. In other words, a layperson’s interest in influenza epidemiology may be triggered by media publicity. Such differentiation is not only relevant for pandemic outbreaks. We found in this study that the local media coverage data preceded influenza case data by one week during the pandemic outbreak in 2009 and the severe winter influenza season in 2011-2012. One interpretation of these observations is that the media coverage reflected a speculative “early warning” viewpoint on the outbreak rather than reports of case rates. Nonetheless, when an early warning is ambiguous or poorly validated, the public may form misperceptions of risks that, consequently, misdirect their behavior [22]. Among the studied eHealth systems, bias from media coverage seemed to influence GFT in particular, as this was the only source that showed large effect size correlations with media coverage data. This interpretation is further supported by the fact that the local media coverage data and the GFT data displayed peaks without corresponding increases in influenza case rates in April-May 2009, a period when the “swine flu” outbreak was highlighted in international media. Additionally, during the seasonal outbreaks in 2007-2008 and 2008-2009, the telenursing data preceded the GFT data by one week, while during the A pH1N1 outbreak in 2009 and the following winter influenza seasons the GFT data either preceded the telenursing data by one week or corresponded in time. Deviances in GFT activity during seasonal outbreaks associated with shifts in media coverage have also been reported from the United States in 2012-2013 influenza season [23], when the GFT estimate for the national peak was almost double that of the CDC. This deviance was attributed to widespread media coverage. Evidently, eHealth data originating from self-care or family care needs have the highest validity when used in infectious disease surveillance. According to agenda-setting theory, mass media have an important influence on what issues the public consider to be important [24]. If a differentiation cannot be made between eHealth data driven by a widespread “lay” epidemiological interest and personal need or care of related individuals, the corresponding data sources risk losing their value in supplementing traditional infectious disease surveillance. Thus, ensuring that Internet data reflect true influenza incidence requires cross-validation with infection-specific data. One strategy is to use telenursing data for real-time validation of Internet data sources, since data on complaints such as fever and cough from telenursing services are less likely to be affected by media publicity than data reflecting Internet activity. However, telenursing data may still not be sufficient for cross-validation if the circulating virus strains mainly affect the older age groups, since the telenursing service utilization in these groups is lower. An alternative is to use data on over-the-counter (OTC) drug sales [25], but the availability of these data in real time may be limited in many countries. Additionally, eHealth data sources can be analyzed in novel ways, such as using multivariate time series methods [26], to obtain improved situational awareness and predictive performance. Nonetheless, these observations suggest that regular validation of the syndromic data sources against clinical and laboratory data is necessary when using eHealth data in influenza surveillance.

Strengths and Limitations

The strength of this study is that it compares three eHealth data streams over a five-year period, including both winter influenza seasons and a pandemic influenza outbreak, and identifies cross-correlations and time lags for the different outbreaks. The Östergötland population is fairly representative for Sweden as a whole, making it possible, although with care, to generalize the results to communities in settings with similar North-European population and geographical characteristics. Although 15% of the Swedish population is foreign-born, the immigrants have arrived mainly from European countries and are well integrated in the Swedish community. There are no reasons to assume that these immigrants’ utilization of health care or eHealth resources differ from the remaining population to an extent that would affect the use of eHealth data for influenza surveillance. Moreover, there are small differences in health care utilization between urban and rural areas in Sweden [10] and the eHealth resources evaluated in this study are evenly accessible in all Swedish counties. However, the study also has important limitations that should be considered when interpreting the results. First, influenza cases were defined by clinical diagnosis, and microbiological validation was restricted to a limited period of the study. However, the effect size of the correlation between the microbiological and clinical diagnosis rates observed in this study was large during the validation period, and similar findings have also been reported from other settings [18]. Second, the telenursing data were based on chief complaint codes defined for Sweden. Some complaints, such as fever and cough, were coded as age-specific syndromes, while other complaints had an age-neutral coding. Internationally standardized telenursing complaint codes would facilitate valid and reliable recording and comparisons between telenursing systems. Third, while data from telenursing centers and website usage data from health service providers were prospectively collected, the GFT data were downloaded in 2012. It is not known if the GFT data for 2007-2011 had been retrospectively adjusted to better correlate to recorded ILI rates. Although it has been reported that GFT algorithms are recalibrated every year [6,23], it is not evident whether or not these recalibrations influence the transformed Web query data available for download. Fourth, this study analyzed the correlations between trends in a set of eHealth data sources and influenza case data. For use in surveillance practice, algorithms need to be developed to translate the time series data into actionable alerts [27]. Finally, it should not be forgotten that different strains of the influenza virus affect different age groups and that eHealth surveillance may be less reliable during winter influenza seasons when the circulating influenza strains mainly affects the elderly in the population.

Conclusions

We found correlations with large effect sizes between eHealth data and influenza case rates in a representative Swedish county over a five-year period including both winter influenza seasons and a pandemic influenza outbreak. Both telenursing center and page-specific website usage data performed at the level of GFT. Although the study design does not allow us to draw conclusions about causal associations with media coverage, we observed that a two-week time lag between eHealth data sources and influenza rates was reduced to one week or none from the 2009 pandemic outbreak when there was parallel intense media coverage. Similarly, we found a tendency for eHealth surveillance to perform worse during winter influenza seasons when the influenza activity involved adolescents and young adults to a lesser degree. The main theoretical implications of the study are that analytic methods need to be developed that adjust eHealth surveillance system to shifts in media coverage and variations in age-group related immunity to specific virus strains. The practical inference is that further longitudinal research incorporating prospective evaluations of actionable alerts [28] is required before eHealth surveillance systems can be used in routine public health practice.

Conflicts of Interest

None declared.

References

1. 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]
2. van Gemert-Pijnen JE, Wynchank S, Covvey HD, Ossebaard HC. Improving the credibility of electronic health technologies. Bull World Health Organ 2012 May 1;90(5):323-323A [FREE Full text] [doi: 10.2471/BLT.11.099804] [Medline: 22589561]
3. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis 2009 Nov 15;49(10):1557-1564 [FREE Full text] [doi: 10.1086/630200] [Medline: 19845471]
4. 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]
5. Kelly H, Grant K. Interim analysis of pandemic influenza (H1N1) 2009 in Australia: surveillance trends, age of infection and effectiveness of seasonal vaccination. Euro Surveill 2009 Aug 6;14(31):1-5 [FREE Full text] [Medline: 19660248]
6. Wilson N, Mason K, Tobias M, Peacey M, Huang QS, Baker M. Interpreting Google flu trends data for pandemic H1N1 influenza: the New Zealand experience. Euro Surveill 2009;14(44):1-3 [FREE Full text] [Medline: 19941777]
7. Cook AR, Chen MI, Pin Lin RT. Internet search limitations and pandemic influenza, Singapore. Emerg Infect Dis 2010 Oct;16(10):1647-1649 [FREE Full text] [doi: 10.3201/eid1610.100840] [Medline: 20875307]
8. Cook S, Conrad C, Fowlkes AL, Mohebbi MH. Assessing Google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. PLoS One 2011;6(8):e23610 [FREE Full text] [doi: 10.1371/journal.pone.0023610] [Medline: 21886802]
9. Ibuka Y, Chapman GB, Meyers LA, Li M, Galvani AP. The dynamics of risk perceptions and precautionary behavior in response to 2009 (H1N1) pandemic influenza. BMC Infect Dis 2010;10:296 [FREE Full text] [doi: 10.1186/1471-2334-10-296] [Medline: 20946662]
10. Timpka T, Eriksson O, Spreco A, Gursky EA, Strömgren M, Holm E, et al. Age as a determinant for dissemination of seasonal and pandemic influenza: an open cohort study of influenza outbreaks in Östergötland County, Sweden. PLoS One 2012;7(2):e31746 [FREE Full text] [doi: 10.1371/journal.pone.0031746] [Medline: 22384066]
11. Warf B. Contemporary digital divides in the United States. Tijdschrift voor Economische en Sociale Geografie 2013;104:1-17.
12. Chowell G, Towers S, Viboud C, Fuentes R, Sotomayor V, Simonsen L, et al. The influence of climatic conditions on the transmission dynamics of the 2009 A/H1N1 influenza pandemic in Chile. BMC Infect Dis 2012;12:298 [FREE Full text] [doi: 10.1186/1471-2334-12-298] [Medline: 23148597]
13. Yu H, Alonso WJ, Feng L, Tan Y, Shu Y, Yang W, et al. Characterization of regional influenza seasonality patterns in China and implications for vaccination strategies: spatio-temporal modeling of surveillance data. PLoS Med 2013 Nov 19;10(11):e1001552. [doi: 10.1371/journal.pmed.1001552]
14. Timpka T, Eriksson H, Gursky EA, Strömgren M, Holm E, Ekberg J, et al. Requirements and design of the PROSPER protocol for implementation of information infrastructures supporting pandemic response: a Nominal Group study. PLoS One 2011;6(3):e17941 [FREE Full text] [doi: 10.1371/journal.pone.0017941] [Medline: 21464918]
15. Kumar S, Snooks H. Telenursing. London: Springer-Verlag; 2011.
16. Cohen J. Statistical power analysis for the behavioral sciences. 2nd edition. New Jersey: Lawrence Erlbaum; 1988.
17. Wahlberg AC, Wredling R. Telephone nursing: calls and caller satisfaction. Int J Nurs Pract 1999 Sep;5(3):164-170. [Medline: 10769626]
18. Ortiz JR, Zhou H, Shay DK, Neuzil KM, Fowlkes AL, Goss CH. Monitoring influenza activity in the United States: a comparison of traditional surveillance systems with Google Flu Trends. PLoS One 2011;6(4):e18687 [FREE Full text] [doi: 10.1371/journal.pone.0018687] [Medline: 21556151]
19. Malik MT, Gumel A, Thompson LH, Strome T, Mahmud SM. "Google flu trends" and emergency department triage data predicted the 2009 pandemic H1N1 waves in Manitoba. Can J Public Health 2011;102(4):294-297. [Medline: 21913587]
20. Kang M, Zhong H, He J, Rutherford S, Yang F. Using Google Trends for influenza surveillance in South China. PLoS One 2013;8(1):e55205 [FREE Full text] [doi: 10.1371/journal.pone.0055205] [Medline: 23372837]
21. Dugas AF, Hsieh YH, Levin SR, Pines JM, Mareiniss DP, Mohareb A, et al. Google Flu Trends: correlation with emergency department influenza rates and crowding metrics. Clin Infect Dis 2012 Feb 15;54(4):463-469 [FREE Full text] [doi: 10.1093/cid/cir883] [Medline: 22230244]

22. Attavanich W, McCarl BA, Bessler D. The Effect of H1N1 (Swine Flu) Media Coverage on Agricultural Commodity Markets. Applied Economic Perspectives and Policy 2011 May 18;33(2):241-259. [doi: 10.1093/aepp/ppr008]
23. Butler D. When Google got flu wrong. Nature 2013 Feb 14;494(7436):155-156. [doi: 10.1038/494155a] [Medline: 23407515]
24. McCombs ME, Shaw DL. The agenda-setting function of mass media. Public Opinion Quarterly 1972;36:176-187. [doi: 10.1086/267990]
25. Liu TY, Sanders JL, Tsui FC, Espino JU, Dato VM, Suyama J. Association of over-the-counter pharmaceutical sales with influenza-like-illnesses to patient volume in an urgent care setting. PLoS One 2013;8(3):e59273 [FREE Full text] [doi: 10.1371/journal.pone.0059273] [Medline: 23555647]
26. Lau EH, Cheng CK, Ip DK, Cowling BJ. Situational awareness of influenza activity based on multiple streams of surveillance data using multivariate dynamic linear model. PLoS One 2012;7(5):e38346 [FREE Full text] [doi: 10.1371/journal.pone.0038346] [Medline: 22675456]
27. 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 [FREE Full text] [doi: 10.2196/jmir.2102] [Medline: 23037553]
28. Milinovich GJ, Williams GM, Clements AC, Hu W. Internet-based surveillance systems for monitoring emerging infectious diseases. Lancet Infect Dis 2014 Feb;14(2):160-168. [doi: 10.1016/S1473-3099(13)70244-5] [Medline: 24290841]

Abbreviations

A: influenza A
EHR: electronic health record
GA: Google Analytics
GFT: Google Flu Trends
ILI: influenza-like illness
RIR: relative illness ratio

Edited by G Eysenbach; submitted 12.11.13; peer-reviewed by J Li, E Lau; comments to author 03.02.14; revised version received 18.02.14; accepted 24.03.14; published 28.04.14.

Please cite as:
Timpka T, Spreco A, Dahlström Ö, Eriksson O, Gursky E, Ekberg J, Blomqvist E, Strömgren M, Karlsson D, Eriksson H, Nyce J, Hinkula J, Holm E
Performance of eHealth Data Sources in Local Influenza Surveillance: A 5-Year Open Cohort Study
J Med Internet Res 2014;16(4):e116
URL: http://www.jmir.org/2014/4/e116/
doi:10.2196/jmir.3099
PMID:24776527

Review

Effectiveness and Cost-Effectiveness of eHealth Interventions in Somatic Diseases: A Systematic Review of Systematic Reviews and Meta-Analyses

Niels J Elbert[1], MD; Harmieke van Os-Medendorp[2], PhD; Wilco van Renselaar[3]; Anne G Ekeland[4,5], PhD; Leona Hakkaart-van Roijen[6,7], PhD; Hein Raat[8], MD, MBA, PhD; Tamar EC Nijsten[1], MD, PhD; Suzanne GMA Pasmans[1,9], MD, PhD

[1]Department of (Pediatric) Dermatology, Erasmus University Medical Center Rotterdam, Rotterdam, Netherlands
[2]Department of Dermatology & Allergology, University Medical Center Utrecht, Utrecht, Netherlands
[3]Patiënt1, Almere, Netherlands
[4]Norwegian Centre for Integrated Care & Telemedicine, University Hospital of North Norway, Tromsø, Norway
[5]Department of Clinical Medicine, Telemedicine & e-Health Research Group, University of Tromsø, Tromsø, Norway
[6]Institute for Medical Technology Assessment, Erasmus University Rotterdam, Rotterdam, Netherlands
[7]Institute of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, Netherlands
[8]Department of Public Health, Erasmus University Medical Center Rotterdam, Rotterdam, Netherlands
[9]Department of Pediatric Dermatology, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, Netherlands

Corresponding Author:
Suzanne GMA Pasmans, MD, PhD
Department of (Pediatric) Dermatology
Erasmus University Medical Center Rotterdam
PO Box 2040
Rotterdam, 3000 CA
Netherlands
Phone: 31 107034580
Fax: 31 107038822
Email: s.pasmans@erasmusmc.nl

Abstract

Background: eHealth potentially enhances quality of care and may reduce health care costs. However, a review of systematic reviews published in 2010 concluded that high-quality evidence on the benefits of eHealth interventions was still lacking.

Objective: We conducted a systematic review of systematic reviews and meta-analyses on the effectiveness/cost-effectiveness of eHealth interventions in patients with somatic diseases to analyze whether, and to what possible extent, the outcome of recent research supports or differs from previous conclusions.

Methods: Literature searches were performed in PubMed, EMBASE, The Cochrane Library, and Scopus for systematic reviews and meta-analyses on eHealth interventions published between August 2009 and December 2012. Articles were screened for relevance based on preset inclusion and exclusion criteria. Citations of residual articles were screened for additional literature. Included papers were critically appraised using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement before data were extracted. Based on conclusions drawn by the authors of the included articles, reviews and meta-analyses were divided into 1 of 3 groups: suitable, promising, or limited evidence on effectiveness/cost-effectiveness. Cases of uncertainty were resolved by consensus discussion. Effect sizes were extracted from papers that included a meta-analysis. To compare our results with previous findings, a trend analysis was performed.

Results: Our literature searches yielded 31 eligible reviews, of which 20 (65%) reported on costs. Seven papers (23%) concluded that eHealth is effective/cost-effective, 13 (42%) underlined that evidence is promising, and others found limited or inconsistent proof. Methodological quality of the included reviews and meta-analyses was generally considered high. Trend analysis showed a considerable accumulation of literature on eHealth. However, a similar percentage of papers concluded that eHealth is effective/cost-effective or evidence is at least promising (65% vs 62%). Reviews focusing primarily on children or family caregivers still remained scarce. Although a pooled (subgroup) analysis of aggregate data from randomized studies was performed in a higher percentage of more recently published reviews (45% vs 27%), data on economic outcome measures were less frequently reported (65% vs 85%).

Conclusions: The number of reviews and meta-analyses on eHealth interventions in patients with somatic diseases has increased considerably in recent years. Most articles show eHealth is effective/cost-effective or at least suggest evidence is promising, which is consistent with previous findings. Although many researchers advocate larger, well-designed, controlled studies, we believe attention should be given to the development and evaluation of strategies to implement effective/cost-effective eHealth initiatives in daily practice, rather than to further strengthen current evidence.

(J Med Internet Res 2014;16(4):e110) doi:10.2196/jmir.2790

KEYWORDS

eHealth; telehealth; telemedicine; review; program effectiveness; cost effectiveness

Introduction

Willem Einthoven started experiments in 1906 with remote consultations via the telephone network and this is when eHealth is likely to have seen first light [1]. It was not until the 1990s when the number of publications in this field of medicine increased dramatically [2]. This was because of the many studies that were carried out involving remote consultations through video-teleconferencing and digital images to give specialists comparable visual inspection of patients as referring doctors [3].

In modern medical practice, eHealth interventions are increasingly present. With nomenclature evolving rapidly, a significant overlap between terms such as *eHealth*, *telemedicine*, and *telehealth* has occurred. The American Telemedicine Association defines telemedicine as “the use of medical information exchanged from one site to another through electronic communications with the purpose of improving the health status of patients,” and considers *eHealth* and *telehealth* as interchangeable nouns. Both words encompass a broader definition of remote health care and also comprise related services, including nonclinical programs such as education, administration, and research [4]. However, telemedicine is a term that is generally reserved for clinical patient care applications [5].

McLean et al [6] conceptualized the definition of eHealth in a Cochrane review on telehealthcare for asthmatic patients as “the provision of personalized health care at a distance.” eHealth contains the following 3 key elements: (1) data obtained from the patient; (2) electronic transfer of data over a distance; and (3) patient-tailored feedback from a health care professional [5,6]. Therefore, communication in eHealth interventions is personalized and interactive in contrast to patient information websites on health and disease.

eHealth potentially enhances the quality of care and reduces health care costs. It may do so by providing patient education and counseling for primary prevention and early detection of disease, replacing face-to-face visits with health care professionals, collecting patient data on medical parameters remotely, among several other mechanisms [6,7]. Because eHealth interventions are considered complex interventions by the Medical Research Council, difficulty may arise in the assessment of the many interacting components of the intervention [8].

In 2010, Ekeland et al [9] published a systematic review of systematic reviews to evaluate the impact of eHealth interventions on health and health care costs. The authors concluded that high-quality evidence on health and economic benefits was still lacking despite the large number of publications. The primary objective of our review is to analyze whether, and to what possible extent, the outcome of recent research supports or differs from these previous conclusions on the effectiveness/cost-effectiveness of eHealth interventions in patients with somatic diseases.

Methods

Overview

Literature searches for systematic reviews and meta-analyses on the effectiveness/cost-effectiveness of eHealth interventions were performed in the following online databases: PubMed, EMBASE, The Cochrane Library, and Scopus. Two of the authors (NE, HO) independently screened all papers’ titles and abstracts for relevance. Citations were screened through Web of Science for additional literature.

Search Queries

Similar to Ekeland et al [9], we used the following (simplified) search query to retrieve systematic reviews and meta-analyses on the effectiveness of eHealth interventions: “[eHealth] AND [effectiveness] AND [systematic review OR meta-analysis].” To search for papers on cost-effectiveness, “AND [costs]” was added to the aforementioned syntax. Because Ekeland et al [9] took into consideration published works from 2005 to July 2009, we limited our search results to articles published between August 2009 and December 2012. Extensive search queries are presented in Tables 1 and 2.

Table 1. PubMed, EMBASE, The Cochrane Library, and Scopus search queries for systematic reviews and meta-analyses on the effectiveness of eHealth interventions (search conducted on September 12, 2013).

Database	Syntax	Hits
PubMed		
Variable 1	e health[Title/Abstract] OR ehealth[Title/Abstract] OR e consultation[Title/Abstract] OR econsultation[Title/Abstract] OR e therapy[Title/Abstract] OR e commerce[Title/Abstract] OR ecommerce[Title/Abstract] OR email consultation[Title/Abstract] OR email consultations[Title/Abstract] OR e mail consultation[Title/Abstract] OR e mail consultations[Title/Abstract] OR telemedicine[Title/Abstract] OR telecare[Title/Abstract] OR teleconsultation[Title/Abstract] OR teleconsultations[Title/Abstract] OR telehealth[Title/Abstract] OR telehomecare[Title/Abstract] OR telehealthcare[Title/Abstract] OR telemonitoring[Title/Abstract] OR telemanagement[Title/Abstract] OR internet[Title/Abstract] OR remote communication[Title/Abstract] OR remote communications[Title/Abstract] OR ict[Title/Abstract] OR web based[Title/Abstract] OR web guided[Title/Abstract]	48,993
Variable 2	effect[Title/Abstract] OR effects[Title/Abstract] OR effectiveness[Title/Abstract] OR efficiency[Title/Abstract] OR efficacy[Title/Abstract]	4,317,041
Variable 3	systematic review[Title/Abstract] OR systematic overview[Title/Abstract] OR meta-analysis[Title/Abstract]	76,640
Total	#1 AND #2 AND #3 AND 2009/01/01[PDat] : 2012/12/31[PDat]	271
EMBASE		
Variable 1	(“e health” OR ehealth OR “e consultation” OR econsultation OR “e therapy” OR “e commerce” OR ecommerce OR “email consultation” OR “email consultations” OR “e mail consultation” OR “e mail consultations” OR telemedicine OR telecare OR teleconsultation OR teleconsultations OR telehealth OR telehomecare OR telehealthcare OR telemonitoring OR telemanagement OR internet OR “remote communication” OR “remote communications” OR ict OR “web based” OR “web guided”):ab,ti	62,242
Variable 2	(effect OR effects OR effectiveness OR efficiency OR efficacy):ab,ti	5,216,580
Variable 3	(“systematic review” OR “systematic overview” OR “meta-analysis”):ab,ti	94,419
Total	#1 AND #2 AND #3 AND (2008-2012)/py	406
Cochrane Library		
Variable 1	(e health OR ehealth OR e consultation OR econsultation OR e therapy OR e commerce OR ecommerce OR email consultation OR email consultations OR e mail consultation OR e mail consultations OR telemedicine OR telecare OR teleconsultation OR teleconsultations OR telehealth OR telehomecare OR telehealthcare OR telemonitoring OR telemanagement OR internet OR remote communication OR remote communications OR ict OR web based OR web guided):ti,ab,kw	15,181
Variable 2	(effect OR effects OR effectiveness OR efficiency OR efficacy):ti,ab,kw	389,345
Variable 3	(systematic review OR systematic overview OR meta-analysis):ti,ab,kw	29,734
Total	(#1 AND #2 AND #3):ti,ab,kw, from 2009 to 2012	385
Scopus		
Variable 1	TITLE-ABS-KEY(“e health” OR ehealth OR “e consultation” OR econsultation OR “e therapy” OR “e commerce” OR ecommerce OR “email consultation” OR “email consultations” OR “e mail consultation” OR “e mail consultations” OR telemedicine OR telecare OR teleconsultation OR teleconsultations OR telehealth OR telehomecare OR telehealthcare OR telemonitoring OR telemanagement OR internet OR “remote communication” OR “remote communications” OR ict OR “web based” OR “web guided”)	365,427
Variable 2	TITLE-ABS-KEY(effect OR effects OR effectiveness OR efficiency OR efficacy)	11,090,998
Variable 3	TITLE-ABS-KEY(“systematic review” OR “systematic overview” OR “meta-analysis”)	158,694
Total	TITLE-ABS-KEY(#1 AND #2 AND #3) AND (LIMIT-TO(PUBYEAR,2012) OR LIMIT-TO(PUBYEAR,2011) OR LIMIT-TO(PUBYEAR,2010) OR LIMIT-TO(PUBYEAR,2009))	595

Table 2. PubMed, EMBASE, The Cochrane Library, and Scopus search queries for systematic reviews and meta-analyses on the cost-effectiveness of eHealth interventions (search conducted on September 12, 2013).

Database	Syntax	Hits
PubMed		
Variable 1	e health[Title/Abstract] OR ehealth[Title/Abstract] OR e consultation[Title/Abstract])OR econsultation[Title/Abstract] OR e therapy[Title/Abstract] OR e commerce[Title/Abstract] OR ecommerce[Title/Abstract] OR email consultation[Title/Abstract] OR email consultations[Title/Abstract] OR e mail consultation[Title/Abstract] OR e mail consultations[Title/Abstract] OR telemedicine[Title/Abstract] OR telecare[Title/Abstract] OR teleconsultation[Title/Abstract] OR teleconsultations[Title/Abstract] OR telehealth[Title/Abstract] OR telehomecare[Title/Abstract] OR telehealthcare[Title/Abstract] OR telemonitoring[Title/Abstract] OR telemanagement[Title/Abstract] OR internet[Title/Abstract] OR remote communication[Title/Abstract] OR remote communications[Title/Abstract] OR ict[Title/Abstract] OR web based[Title/Abstract] OR web guided[Title/Abstract]	48,993
Variable 2	effect[Title/Abstract] OR effects[Title/Abstract] OR effectiveness[Title/Abstract] OR efficiency[Title/Abstract] OR efficacy[Title/Abstract]	4,317,041
Variable 3	cost[Title/Abstract] OR costs[Title/Abstract] OR economic[Title/Abstract] OR economically[Title/Abstract]	396,949
Variable 4	systematic review[Title/Abstract] OR systematic overview[Title/Abstract] OR meta-analysis[Title/Abstract]	76,640
Total	#1 AND #2 AND #3 AND #4 AND 2009/01/01[PDat] : 2012/12/31[PDat]	76
EMBASE		
Variable 1	(“e health” OR ehealth OR “e consultation” OR econsultation OR “e therapy” OR “e commerce” OR ecommerce OR “email consultation” OR “email consultations” OR “e mail consultation” OR “e mail consultations” OR telemedicine OR telecare OR teleconsultation OR teleconsultations OR telehealth OR telehomecare OR telehealthcare OR telemonitoring OR telemanagement OR internet OR “remote communication” OR “remote communications” OR ict OR “web based” OR “web guided”):ab,ti	62,242
Variable 2	(effect OR effects OR effectiveness OR efficiency OR efficacy):ab,ti	5,216,580
Variable 3	(cost OR costs OR economic OR economically):ab,ti	502,150
Variable 4	(“systematic review” OR “systematic overview” OR “meta-analysis”):ab,ti	94,419
Total	#1 AND #2 AND #3 AND #4 AND (2008-2012)/py	113
Cochrane Library		
Variable 1	(e health OR ehealth OR e consultation OR econsultation OR e therapy OR e commerce OR ecommerce OR email consultation OR email consultations OR e mail consultation OR e mail consultations OR telemedicine OR telecare OR teleconsultation OR teleconsultations OR telehealth OR telehomecare OR telehealthcare OR telemonitoring OR telemanagement OR internet OR remote communication OR remote communications OR ict OR web based OR web guided):ti,ab,kw	15,181
Variable 2	(effect OR effects OR effectiveness OR efficiency OR efficacy):ti,ab,kw	389,345
Variable 3	(cost OR costs OR economic OR economically):ti,ab,kw	50,911
Variable 4	(systematic review OR systematic overview OR meta-analysis):ti,ab,kw	29,734
Total	(#1 AND #2 AND #3 AND #4):ti,ab,kw, from 2009 to 2012	248
Scopus		
Variable 1	TITLE-ABS-KEY(“e health” OR ehealth OR “e consultation” OR econsultation OR “e therapy” OR “e commerce” OR ecommerce OR “email consultation” OR “email consultations” OR “e mail consultation” OR “e mail consultations” OR telemedicine OR telecare OR teleconsultation OR teleconsultations OR telehealth OR telehomecare OR telehealthcare OR telemonitoring OR telemanagement OR internet OR “remote communication” OR “remote communications” OR ict OR “web based” OR “web guided”)	365,427
Variable 2	TITLE-ABS-KEY(effect OR effects OR effectiveness OR efficiency OR efficacy)	11,090,998
Variable 3	TITLE-ABS-KEY(cost OR costs OR economic OR economically)	2,205,295
Variable 4	TITLE-ABS-KEY(“systematic review” OR “systematic overview” OR “meta-analysis”)	158,694
Total	TITLE-ABS-KEY(#1 AND #2 AND #3 AND #4) AND (LIMIT-TO(PUBYEAR,2012) OR LIMIT-TO(PUBYEAR,2011) OR LIMIT-TO(PUBYEAR,2010) OR LIMIT-TO(PUBYEAR,2009))	182

Inclusion Criteria

Systematic reviews and meta-analyses on eHealth interventions in adults and/or children with somatic diseases (ie, illnesses with a physical cause, not mental), and those focusing on family caregivers were included. Interventions had to meet the following 3 criteria: (1) data were obtained from the patient or family caregiver, (2) data were electronically transferred over a distance, and (3) personalized feedback was given from a health care professional. Reviews and meta-analyses of individual studies comparing eHealth interventions to usual or no care, and those comparing different eHealth initiatives were assessed. We only accounted for papers reporting health-related outcomes, costs, patient satisfaction, and/or self-management.

Exclusion Criteria

Those eHealth interventions that were not home-based (eg, tele-ICU) or not patient or family caregiver–oriented (eg, education of medical or nursing students and health care professionals) were excluded. We excluded meta-analyses that included nonrandomized studies (eg, cohort studies) unless a subgroup analysis of randomized studies (eg, randomized controlled trials, randomized crossover trials) was performed. In addition, we did not assess papers written in languages other than English or Dutch, and those for which the full-text was not available online.

In contrast to Ekeland et al [9], we narrowed the focus of our work by excluding reviews and meta-analyses on nonsomatic disorders (eg, mental disorders such as anxiety, depression, schizophrenia, and posttraumatic stress disorder) and lifestyle changes (eg, smoking cessation and drug intervention programs) to increase the comparability of the included papers and to limit the search results.

Outcome Measures

Health-related effects (eg, morbidity, mortality, quality of life, hospitalization) and health care costs (eg, health care utilization) were defined as primary outcome measures. We considered patient satisfaction and self-management as secondary outcome measures.

Critical Appraisal

Before data were extracted, the included papers were critically appraised using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA, formerly QUOROM) Statement [10]. The PRISMA Statement provides an evidence-based 27-item checklist (eg, on objectives, methodology, and limitations) for reporting in systematic reviews and meta-analyses.

Data Extraction

Based on conclusions drawn by the authors of the included papers, all reviews and meta-analyses were divided into 1 of 3 groups: (1) suitable, (2) promising, or (3) limited evidence that eHealth is effective/cost-effective. Cases of uncertainty were resolved by consensus discussion between 2 authors of the current review (NE, HO). Effect sizes, such as standardized or weighted mean differences, relative risks, odds ratios, and *z* scores, were extracted from papers that included a pooled (subgroup) analysis of aggregate data from randomized studies. No attempt was made to contact authors for missing data. To analyze whether the results of the included papers supported or differed from previous findings by Ekeland et al [9], we performed a trend analysis using basic statistics.

Results

Search Results

The initial search yielded a total of 1657 articles, including 619 articles that reported on cost-related outcome measures (Figures 1 and 2). Following removal of duplicates and screening of the residual papers on preset inclusion and exclusion criteria, 30 eligible reviews remained [6,11-39], of which 19 reported on costs [6,13,14,16,18,19,21-24,26,28,29,31,33,34,37-39]. Subsequent citation screening through Web of Science resulted in 1 additional paper [40]. Thus, a total of 31 reviews were retrieved (Figure 1), of which 20 (65%) reported on costs (Figure 2). Three of 31 reviews (10%) reported primarily on children [28,37,38], and 1 of 31 (3%) focused on the effects of eHealth interventions on family caregivers [26].

Figure 1. Flow diagram of the literature search on the effectiveness of eHealth interventions.

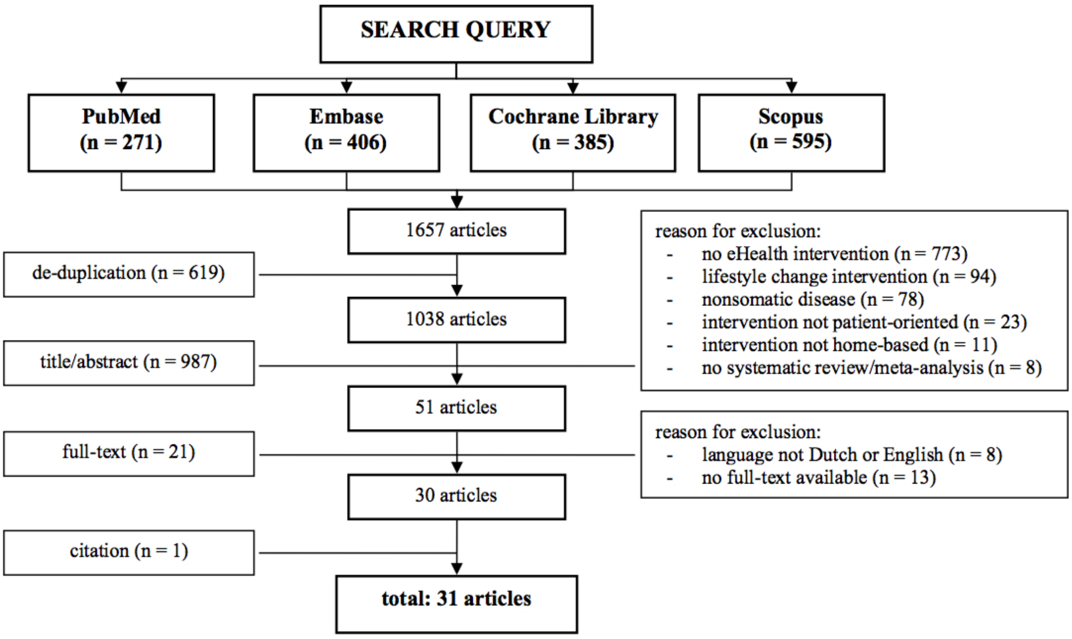
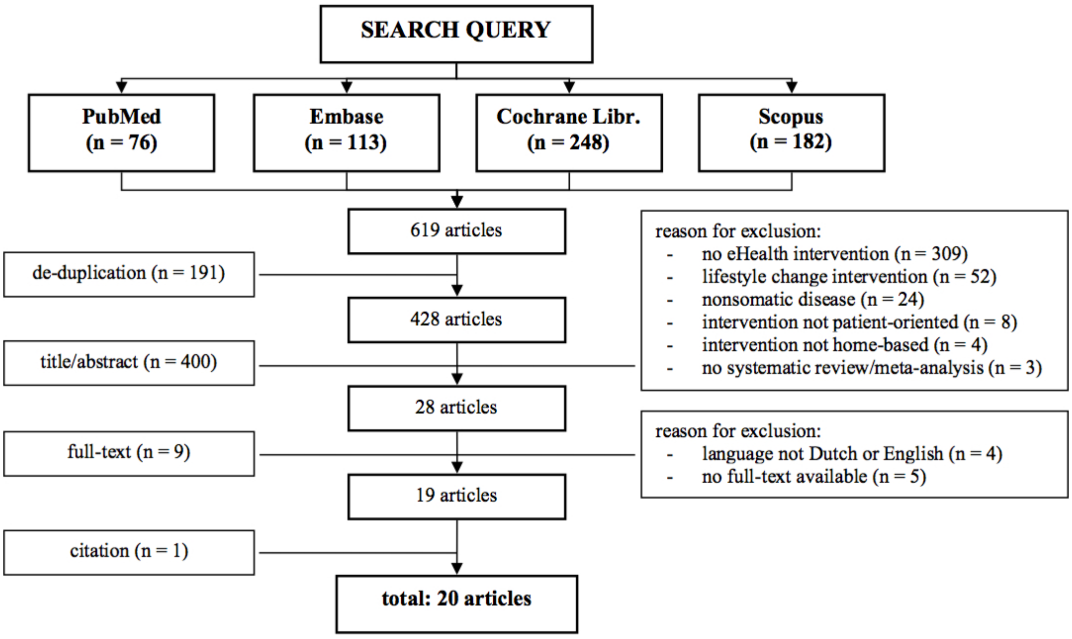


Figure 2. Flow diagram of the literature search on the cost-effectiveness of eHealth interventions.



Effects of eHealth Interventions

Results per article are summarized in 3 separate tables, 1 for systematic reviews and meta-analyses reporting eHealth interventions are effective/cost-effective (Table 3), a second table for papers showing evidence is promising (Table 4), and a third table with papers underlining evidence is lacking, limited, or inconsistent (Table 5). Table 6 demonstrates the effect sizes-among other characteristics-reported in 14 reviews in which a pooled (subgroup) analysis of aggregate data from randomized studies was performed. All tables are presented subsequently.

Effectiveness/Cost-Effectiveness of eHealth Interventions

A total of 7 reviews (23%) showed eHealth interventions are effective on either health or cost-related outcome measures (Table 3) [11-17]. Study populations consisted of patients with congestive heart failure (CHF) [13-15,17], diabetes mellitus [12,16], and hypertension [11]. Types of interventions that were effective/cost-effective comprised home telemonitoring [11,13-17], Web or mobile phone-based education [12,16], structured telephone support [14-16], and mobile phone-assisted self-management programs [16]. Patient acceptance and satisfaction were generally considered high.

Pooled analyses were performed in each of the 7 reviews and demonstrated significant reduction of all-cause mortality, all-cause hospitalization, and CHF-related hospital admissions through home telemonitoring and structured telephone support in patients with CHF [13-15,17]. Home telemonitoring also resulted in significant improvement of systolic blood pressure and nonsignificant reduction of diastolic blood pressure, antihypertensive drug use, and therapeutic inertia (ie, unchanged medication despite elevated blood pressure) in hypertensive patients [11]. Web-based education and various mobile phone interventions led to significant improvement of laboratory parameters, such as glycosylated hemoglobin (HbA1c) and low-density lipoprotein (LDL) cholesterol, in diabetic patients [12,16].

Qualitative analysis of individual studies revealed several other positive effects of eHealth interventions, including economic benefits [14,16], reduction of the number of visits to outpatient clinics [12], increase of disease-related knowledge and self-management [12,14,16], and improvement of quality of life [13,14,17].

Table 3. Systematic reviews and meta-analyses in which eHealth interventions were shown to be effective/cost-effective.

Study characteristic	Result[a]
Agarwal et al [11]	
Conditions included	Hypertension
Geographic area	Europe, North America
Service/intervention	Home telemonitoring
Outcome measures	Health
Authors' summary of results	27 RCTs included comparing home- to office-based blood pressure monitoring. Home TM was used in 7 studies and significantly improved SBP. Meta-analysis also showed reduction of DBP, higher blood pressure response rates, and lower antihypertensive drug use; however, these results were not statistically significant.
Authors' conclusions	Home blood pressure monitoring significantly improved SBP compared to office-based measurements. Reductions were even greater when TM was used.
Angeles et al [12]	
Conditions included	Type 1 and 2 diabetes mellitus
Geographic area	Europe, North America, Asia
Service/intervention	Web-based education
Outcome measures	Health, patient satisfaction, self-management
Authors' summary of results	9 RCTs included comparing Web-based education to usual care. Meta-analysis showed a significant mean difference in HbA1c after 3, 6, and 12 months and in LDL-cholesterol favoring Web-based education. No significant difference was found for HDL-cholesterol, total cholesterol, and FPG. Other benefits included better patient satisfaction, self-efficacy, and self-management, and reduced clinic visits. However, 1 study demonstrated no differences in health care visits or hospital patient days.
Authors' conclusions	Web-based education is superior to usual care in improving HbA1c and LDL-cholesterol
Clarke et al [13]	
Conditions included	Congestive heart failure
Geographic area	Not stated
Service/intervention	Home telemonitoring
Outcome measures	Health, costs, patient satisfaction, self-management
Authors' summary of results	13 RCTs included comparing home TM to usual care. Meta-analysis showed significant reduction of all-cause mortality and CHF-related hospital admissions, favoring home TM. No significant difference was found for all-cause hospital and emergency admissions. Qualitative analysis demonstrated no significant difference in hospital length, medication adherence, and costs. 6 studies showed high patient acceptance and satisfaction; 6 different studies reported a trend toward greater improvement of the quality of life. Providing knowledge to the patients allows them to take greater responsibility for their own management and increases patient empowerment.
Authors' conclusions	Patients with CHF receiving home TM lived longer without increasing their use of health care facilities. These favorable outcomes support the wider use of home TM.
Inglis et al [14]	
Conditions included	Congestive heart failure
Geographic area	Europe, North America, Asia, Oceania, Latin America
Service/intervention	Home telemonitoring, structured telephone support
Outcome measures	Health, costs, patient satisfaction, self-management
Authors' summary of results	25 RCTs included, 11 and 16 of which compared home TM and STS, respectively, to usual care. Meta-analysis showed significant reduction of all-cause mortality, favoring home TM. STS showed a similar, but nonsignificant trend. Both interventions significantly reduced all-cause and CHF-related hospitalization. Qualitative analysis demonstrated reduced costs, high patient satisfaction, and improved quality of life, patient knowledge, and functional class.
Authors' conclusions	Home TM and STS appear effective interventions in patients with CHF
Klersy et al [15]	
Conditions included	Congestive heart failure

Study characteristic	Result[a]
Geographic area	Not stated
Service/intervention	Home telemonitoring, structured telephone support
Outcome measures	Health
Authors' summary of results	32 studies (20 RCTs, 12 cohort studies) were included, comparing home TM and STS to usual care. All-cause mortality and hospital admissions were assessed separately for RCTs and cohort studies; CHF-related hospital admissions could only be assessed for RCTs. Meta-analysis showed significant reduction of all-cause mortality, and all-cause and CHF-related hospitalization, favoring home TM and STS.
Authors' conclusions	Home TM and STS confer a significant protective clinical effect, compared to usual care. Mid- and long-term cost-effectiveness of these interventions remains to be evaluated.
Liang et al [16]	
Conditions included	Type 1 and 2 diabetes mellitus
Geographic area	Not stated
Service/intervention	Mobile phone interventions, including home telemonitoring, structured telephone support, education and self-management programs
Outcome measures	Health, costs, patient satisfaction, self-management
Authors' summary of results	22 controlled studies (11 RCTs, 2 QRCTs, 2 COTs, 7 NCBAs) were included, assessing the effect of various mobile phone interventions on glycemic control. Meta-analysis showed significant reduction of HbA1c and improvement of self-management, favoring the intervention in both type 1 and 2 diabetic patients. The cost/benefit ratio of the intervention was calculated in only 5 of 22 studies; all 5 reported that the intervention was cost-effective. Most studies reported that the patients were satisfied with the intervention.
Authors' conclusions	Mobile phone interventions effectively improve glycemic control and self-management in diabetic patients, especially in patients with type 2 diabetes mellitus
Polisena et al [17]	
Conditions included	Congestive heart failure
Geographic area	Europe, North America, Asia
Service/intervention	Home telemonitoring
Outcome measures	Health, patient satisfaction
Authors' summary of results	21 studies (11 RCTs, 6 cohort studies, 4 NUBAs) included comparing home TM to usual care. Meta-analysis of the included RCTs showed significant reduction of all-cause mortality, favoring home TM. Qualitative analysis suggests home TM may lower hospitalization rates and the use of other health care services. Quality of life and patient satisfaction with home TM were similar or better than with usual care.
Authors' conclusions	Home TM is clinically effective, but the effect on health care utilization is more limited

[a] CHF: congestive heart failure; COT: randomized crossover trial; DBP: diastolic blood pressure; FPG: fasting plasma glucose; HbA1c: glycosylated hemoglobin; HDL: high-density lipoprotein; LDL: low-density lipoprotein; NCBA: nonrandomized controlled before-after study; NUBA: nonrandomized uncontrolled before-after study; QRCT: quasi-randomized controlled trial; RCT: randomized controlled trial; SBP: systolic blood pressure; STS: structured telephone support; TM: telemonitoring.

Evidence on eHealth Interventions is Promising

Thirteen reviews (42%) were less confident about the effectiveness/cost-effectiveness of eHealth interventions [18-29,40], but suggested that these initiatives are promising or bear potential (Table 4). Many of the authors claim additional research is needed to clarify efficacy and cost-related issues.

Pooled analyses were performed in 4 reviews and presented subsequently [22-24,27]. One review on chronic obstructive pulmonary disease (COPD) demonstrated the capacity of eHealth interventions to significantly reduce the number of patients with 1 or more emergency department visits or hospital admissions-due to exacerbation of pulmonary symptoms-over a 12-month period [23]. eHealth interventions did not significantly improve quality of life and all-cause mortality. Because the interventions were often part of complex interventions, the authors concluded that further investigation is required to determine the precise role of eHealth. Promising effects were also identified for Internet-based peer and clinical visit support programs—among several other eHealth interventions—in acute and chronic pain management [18,22]. Although the Internet was supportive in the treatment of pain, it remained unclear what benefits could be gained and which patients would profit most.

Table 4. Systematic reviews and meta-analyses in which promising evidence on the effectiveness/cost-effectiveness of eHealth interventions was reported.

Study characteristic	Result[a]
Bender et al [18]	
Conditions included	Acute and chronic pain
Geographic area	North America
Service/intervention	Internet-based peer and clinical visit support programs, including education and self-management programs
Outcome measures	Health, costs, patient satisfaction, self-management
Authors' summary of results	17 RCTs were included, of which 6 compared Internet-based peer and clinical visit support programs to usual care or an existing nursing website; the other 11 articles described cognitive and behavioral interventions. Meta-analysis was considered inappropriate due to substantial heterogeneity among studies. Qualitative analysis showed limited but promising evidence that Internet-based peer support programs can lead to improvements of pain intensity, activity limitations, health distress and self-management, and can reduce pain in children and adolescents. Insufficient evidence was found on the effects of Internet-based clinical support interventions. Two studies found no significant difference in health care utilization; another study showed significant difference in knowledge and patient satisfaction.
Authors' conclusions	Internet-based interventions seem promising for people in pain, but it remains unclear which patients benefit most
Eland-de Kok et al [19]	
Conditions included	Chronic diseases
Geographic area	Not stated
Service/intervention	Interactive websites with store-and-forward services, including home telemonitoring, video-teleconferencing, education, self-management programs and cardiac rehabilitation
Outcome measures	Health, costs, patient satisfaction
Authors' summary of results	12 RCTs included assessing interactive websites with store-and-forward services as an addition to or instead of usual face-to-face care. Meta-analysis was considered inappropriate due to substantial heterogeneity among studies. Qualitative analysis showed significant improvement of physical health outcomes with small to moderate effect sizes. Not all outcomes improved, and some measures showed comparable effect sizes. Costs, quality of life, and patient satisfaction were rarely assessed and showed various results.
Authors' conclusions	eHealth is a promising tool for treatment and self-management training of chronically ill patients
Hailey et al [20]	
Conditions included	Any
Geographic area	Not stated
Service/intervention	In-home telerehabilitation
Outcome measures	Health, self-management
Authors' summary of results	61 studies with various types of study designs (not further specified) were included. For a variety of populations and types of outcome, 71% of the interventions were successful, 18% were unsuccessful, and for 11% the status was unclear. The reported outcomes for 51% of the interventions appeared to be clinically significant. Success was not demonstrated in cardiac studies on improvements in self-efficacy and physical activity.
Authors' conclusions	In-home telerehabilitation shows promise in many fields, but compelling evidence of benefit is still limited
Johansson et al [21]	
Conditions included	Stroke
Geographic area	Europe, North America, Asia
Service/ intervention	In-home telerehabilitation, including telephone and videophone consulting
Outcome measures	Health, costs, patient satisfaction
Authors' summary of results	9 studies (4 RCTs, 4 case series, 1 qualitative analysis) were included. Qualitative analysis showed better SF-36 scores in stroke patients who underwent in-home telerehabilitation, but this difference was not significant. No significant differences were found in various secondary outcome measures such as the Hospital Anxiety and Depression Scale, Barthel Index, modified Ranking Scale, and the use of secondary prevention drugs since discharge. Participants reported high level of satisfaction and acceptance of the intervention. No study reported on cost-effectiveness or resource utilization.

Study characteristic	Result[a]
Authors' conclusions	In-home telerehabilitation showed promising results in poststroke care, but the quality of evidence was low
McGeary et al [22]	
Conditions included	Chronic pain
Geographic area	Not stated
Service/intervention	Technology-based interventions, including structured telephone support, video-teleconferencing, self-management programs, and outpatient telerehabilitation
Outcome measures	Health, costs, patient satisfaction, self-management
Authors' summary of results	10 studies (9 RCTs, 1 COT) were included, comparing a range of technologies to usual care or waiting-list control conditions. Meta-analysis showed a significant overall benefit of eHealth interventions over control conditions and equivalence with in-person interventions.
Authors' conclusions	eHealth interventions can result in successful pain management, but qualitative trials are lacking. Therefore, it is unclear exactly what benefits can be obtained.
McLean et al [23]	
Conditions included	COPD
Geographic area	Europe, North America, Asia
Service/intervention	Telehealthcare interventions, including home telemonitoring, structured telephone support, video-teleconferencing and self-management programs
Outcome measures	Health, costs, patient satisfaction, self-management
Authors' summary of results	10 RCTs were included, comparing a range of telehealthcare interventions to usual care or face-to-face home visits. Meta-analysis showed significant reduction of all-cause emergency department visits and hospital admissions, favoring eHealth. Quality of life and all-cause mortality did not significantly improve. Three studies reported on patient satisfaction using different invalidated scales.
Authors' conclusions	Telehealthcare interventions appear to have a possible effect on quality of life and all-cause emergency department visits and hospital admissions. Further research is needed to clarify precisely its role, since interventions were assessed as part of a complex intervention.
Omboni et al [24]	
Conditions included	Hypertension
Geographic area	Not stated
Service/intervention	Home telemonitoring, education
Outcome measures	Health, costs
Authors' summary of results	12 RCTs were included, comparing home blood pressure TM to usual care. 6 interventions incorporated an educational component. Meta-analysis showed significant improvement of blood pressure control and reduction of both SBP and DBP, favoring home TM. Home TM was associated with a modest, but significantly increased use of antihypertensive medications. Information on costs was available from merely a few studies. In 1 study, quality of life tended to be higher and costs lower in the intervention group teletransmitting blood pressure data. Another study observed lower medication and consultation costs in the intervention group, which were however offset by the cost of the telemonitoring equipment.
Authors' conclusions	Home TM may represent a useful tool to improve blood pressure control and reduce adverse cardiovascular events. Well-designed, large-scale RCTs are still needed to demonstrate its superiority and clinical usefulness.
Paré et al [25]	
Conditions included	Chronic diseases
Geographic area	Europe, North America, Asia
Service/intervention	Home telemonitoring, education
Outcome measures	Health

Study characteristic	Result[a]
Authors' summary of results	62 studies (45 RCTs, 17 nonrandomized studies that were not further specified) were included to assess the clinical effects of home TM. Meta-analysis was considered inappropriate due to substantial heterogeneity among studies. Qualitative analysis of studies on diabetes mellitus indicated a trend toward better glycemic control with home TM. In most trials on asthma, significant improvement of peak expiratory flows and quality of life, and reduction of asthma-related symptoms were found. Studies on hypertension generally demonstrated reduction of SBP and/or DBP.
Authors' conclusions	Home TM appears to be promising in patient management, but designers of future studies should consider ways to make this technology more effective as well as controlling possible mediating variables
Pron et al [40]	
Conditions included	Cardiac arrhythmia, heart failure
Geographic area	Europe, North America
Service/intervention	Internet-based device-assisted remote monitoring systems for therapeutic cardiovascular implantable electronic devices
Outcome measures	Health, costs, patient satisfaction
Authors' summary of results	23 studies (7 RCTs, 16 cohort studies) were included, comparing Internet-based device-assisted RMSs to usual care. Qualitative analysis of multiple cohort studies and 2 RCTs demonstrated feasibility and significant reduction of in-office clinic follow-ups with RMSs in the first year post implantation. Detection rates of clinically significant events were higher and the time to a clinical decision for these events was significantly shorter. Earlier detection was not associated with lower morbidity or mortality rates. Patient acceptance and satisfaction were reported to be high. The incremental cost of providing RMSs was approximately Can –$409K per year (cost savings); corresponding incremental cost per patient was Can –$98 per year.
Authors' conclusions	RMSs have the potential to improve current surveillance systems, but there is insufficient information to evaluate the overall impact to the health care system
Rietdijk et al [26]	
Conditions included	Traumatic brain injury
Geographic area	Not stated
Service/intervention	Technology-assisted training and support programs, including video-teleconferencing, education and self-management interventions
Outcome measures	Health, costs, patient satisfaction
Authors' summary of results	16 studies (7 RCTs, 4 CCTs, 5 case series) were included, qualitatively describing the effectiveness of technology-assisted training and support programs to family caregivers of patients with traumatic brain injury. Meta-analysis was considered inappropriate due to substantial heterogeneity among studies. All but 1 study reported some degree of positive results on feasibility, user satisfaction, and preliminary explorations of effectiveness. No studies reported any formal cost analysis, although 1 study that provided cost estimates noted the intervention was much less expensive than the intensive inpatient comparison condition which had similar outcomes.
Authors' conclusions	Technology-assisted programs seem a promising approach in the training and support of family members of patients with traumatic brain injury
Samoocha et al [27]	
Conditions included	Any
Geographic area	Not stated
Service/intervention	Web-based interventions, including education and self-management programs
Outcome measures	Self-management
Authors' summary of results	14 RCTs included comparing Web-based education to usual care, waiting-list control conditions, or no care in various somatic conditions. Meta-analysis showed a significant positive effect on patient empowerment measured with the Diabetes Empowerment Scale and on self-efficacy measured with disease-specific scales, both favoring Web-based tools. No effects were found for self-efficacy measured with general scales.
Authors' conclusions	Web-based education showed positive, but generally small effects. Direct and indirect impacts of these effects remain unknown
Stinson et al [28]	
Conditions included	Chronic diseases

Study characteristic	Result[a]
Geographic area	Europe, North America, Asia
Service/intervention	Internet-based self-management programs
Outcome measures	Health, costs, self-management
Authors' summary of results	9 studies (7 RCTs, 1 pilot RCT, 1 quasi-experimental study) with Internet-based self-management programs were included. Due to the limited data reported in the studies only a qualitative analysis was provided. Seven studies demonstrated significant improvement of health-related outcome measures, compared to the control group. There was conflicting evidence regarding the effect of Internet-based self-management programs on disease-specific knowledge and quality of life, whereas evidence on health care utilization was limited.
Authors' conclusions	The Internet shows great promise as a mode of delivering self-management programs for youth with health conditions
Van den Berg et al [29]	
Conditions included	Elderly people with chronic diseases
Geographic area	Europe, North America, Asia, Oceania, Latin America
Service/intervention	Telemedicine interventions, including home telemonitoring, structured telephone support, video-teleconferencing, education, self-care training and telerehabilitation
Outcome measures	Health, costs, patient satisfaction, self-management
Authors' summary of results	68 studies (56 RCTs, 12 CCTs) were included, a range of technologies to one another or to usual care in elderly people with chronic diseases. Literature shows predominantly positive results with a clear trend toward better results for telemedicine interventions, independent of the diagnosis group. 36 studies comprised an economic endpoint, of which 15 showed positive and 2 mixed results. None of the studies reported a significantly better outcome for the control group.
Authors' conclusions	The many positive examples provided in the literature indicate the considerable potential of eHealth

[a] CCT: nonrandomized controlled clinical trial; COPD: chronic obstructive pulmonary disease; COT: randomized crossover trial; DBP: diastolic blood pressure; RCT: randomized controlled trial; RMS: remote monitoring systems; SBP: systolic blood pressure; SF: short form; TM: telemonitoring.

Qualitative analysis of individual studies revealed many other promising effects of eHealth interventions, for example, Internet-based device-assisted remote monitoring systems in patients with cardiovascular implantable electronic devices [40], in-home telerehabilitation in routine care of patients with stroke and other somatic diseases [20,21], technology-assisted training and support programs for family members of patients with traumatic brain injury [26], and Web-based education to increase patient empowerment [27]. Paré et al [25] assessed the clinical effects of home telemonitoring in patients with a variety of chronic diseases. The authors highlight the fact that home telemonitoring allows for closer follow-up of individual patients' conditions and for early detection of warning signs in case of health deterioration. However, they claim larger trials are needed to confirm the clinical effects of home telemonitoring.

Evidence on eHealth Interventions Is Lacking, Limited or Inconsistent

Eleven reviews (35%) underlined that evidence on the effectiveness/cost-effectiveness of eHealth interventions is still lacking, limited, or inconsistent (Table 5) [6,30-39]. In many articles, the poor methodological quality of individual studies is criticized, and ambiguous or conflicting findings are emphasized.

McLean et al [6] conducted a Cochrane review of 21 RCTs on a range of eHealth interventions in patients with asthma. Meta-analysis did not show a clinically important improvement of disease-specific quality of life, and no significant reduction of all-cause emergency department visits over a 12-month period was found (Table 6). The authors concluded that eHealth is unlikely to result in clinically relevant improvements of health-related outcome measures in patients with relatively mild disease, but does appear to have the potential to reduce all-cause hospital admissions in those with more severe disease.

Shulman et al [37] studied the impact of eHealth interventions involving transmission of blood glucose data in youth with type 1 diabetes mellitus. Pooled analyses showed no apparent effect of the interventions on HbA1c or acute complications, such as severe hypoglycemia and diabetic ketoacidosis (Table 6). The limited data available on patient satisfaction and costs also suggested no differences between the intervention and the comparison group.

Table 5. Systematic reviews and meta-analyses in which no, limited, or inconsistent evidence on the effectiveness/cost-effectiveness of eHealth interventions was reported.

Study characteristic	Result[a]
Baron et al [30]	
Conditions included	Type 1 and 2 diabetes mellitus
Geographic area	Europe, North America, Asia
Service/ intervention	Mobile phone telemonitoring
Outcome measures	Health
Authors' summary of results	20 studies (12 RCTs, 4 NUBAs, 2 pilot COTs, 1 NCBA, 1 pilot NUBA) were included. Of the 15 controlled studies, 9 compared mobile phone telemonitoring to standard care and 6 to a different (eHealth) intervention. Qualitative analysis demonstrated mixed results of diet-focused interventions. Evidence on the effect of nondietary interventions in patients with type 1 diabetes mellitus was inconclusive. Of the 13 studies reporting on patients with type 2 diabetes mellitus, 7 found structured mobile phone support to be more effective than other eHealth interventions and standard care in reducing HbA1c.
Authors' conclusions	Evidence on the effectiveness of mobile phone telemonitoring was inconsistent and remains weak
Bolton et al [31]	
Conditions included	COPD
Geographic area	Not stated
Service/ intervention	Home telemonitoring, education
Outcome measures	Health, costs
Authors' summary of results	6 studies (2 RCTs, 2 NCBAs, 2 NUBAs) on home TM were included, of which some had an educational element. Meta-analysis was considered inappropriate, because individual studies were underpowered, had heterogeneous patient populations and had a lack of detailed intervention description. Qualitative analysis showed positive results on health and costs.
Authors' conclusions	The benefit of home TM in patients with COPD is not yet proven and further research is required before large-scale implementation
Ciere et al [32]	
Conditions included	Congestive heart failure
Geographic area	Europe, North America
Service/ intervention	Telehealthcare interventions, including home telemonitoring and education
Outcome measures	Self-management
Authors' summary of results	12 studies (9 RCTs, 1 pilot RCT, 1 CCT, 1 NUBA) were included, comparing various telehealthcare interventions to usual care or to home nurse visits. Meta-analysis was considered inappropriate due to substantial heterogeneity among studies. Qualitative analysis showed inconclusive evidence that telehealthcare interventions improve patient knowledge, self-care, or self-efficacy.
Authors' conclusions	Literature provides insufficient evidence to robustly support or disprove beneficial effects of telehealthcare interventions in patients with CHF
Franek [33]	
Conditions included	COPD
Geographic area	Not stated
Service/ intervention	Home telemonitoring, structured telephone support
Outcome measures	Health, costs, patient satisfaction, self-management
Authors' summary of results	5 studies (3 RCTs, 2 CCTs) were included, comparing home TM to usual care in patients with moderate to severe COPD; another RCT compared STS to usual care. Meta-analysis was considered inappropriate due to substantial heterogeneity among studies. Qualitative analysis showed nonsignificant or conflicting effects of home TM for all outcome measures, including health care utilization, mortality, quality of life, number of exacerbations, patient satisfaction, and safety. Low quality evidence showed significant benefit in favor of STS for self-efficacy and emergency department visits, but nonsignificant results for hospitalization and hospital length of stay. The economic impact of both interventions was uncertain.
Authors' conclusions	Low to very low quality evidence found nonsignificant or conflicting effects of home TM for all outcome measures. Low quality evidence showed significant benefit of STS for self-efficacy and emergency department visits, but nonsignificant results for hospitalization and hospital length of stay.

Study characteristic	Result[a]
McLean et al [6]	
Conditions included	Asthma
Geographic area	Europe, North America, Asia, Oceania, Latin America
Service/ intervention	Telehealthcare interventions, including home telemonitoring, structured telephone support, video-teleconferencing, education and self-management programs
Outcome measures	Health, costs, patient satisfaction
Authors' summary of results	21 RCTs were included, comparing a range of technologies to (enhanced) face-to-face usual care. Meta-analysis showed no clinically important improvement of disease-specific quality of life, and no significant reduction of all-cause emergency department visits over 12 months was found. However, all-cause hospital admissions over 12 months were significantly reduced, particularly in patients with severe asthma. Costs were favorable to continuing the intervention where hospitalization was prevented, but this was not true for all studies.
Authors' conclusions	Telehealthcare interventions are unlikely to result in clinically relevant improvements in patients with relatively mild asthma, but do appear to have the potential to reduce all-cause hospital admissions in those with more severe disease
Mistry [34]	
Conditions included	Any
Geographic area	Predominantly Europe and North America
Service/ intervention	Telemedicine and telecare interventions not otherwise specified
Outcome measures	Costs
Authors' summary of results	80 studies (38 CCAs, 18 CMAs, 15 CEAs, 7 CUAs, 2 CBAs) were included. Economic tools are increasingly being used to evaluate telemedicine and telecare interventions, but transparency in the reporting of methodologies and results is required. Literature showed no general agreement whether eHealth interventions were cost-effective, compared to conventional means.
Authors' conclusions	Literature provides no conclusive evidence on the cost-effectiveness of telemedicine and telecare interventions
Ryhänen et al [35]	
Conditions included	Breast cancer
Geographic area	North America
Service/ intervention	Internet or interactive computer-based education
Outcome measures	Health, patient satisfaction, self-management
Authors' summary of results	14 studies (9 RCTs, 2 CCTs, 3 quasi-experimental studies) were included, comparing Internet or interactive computer-based education to various control conditions, such as usual or no care, waiting-list control conditions, and discussion with counselors or physicists. Meta-analysis was considered inappropriate due to substantial heterogeneity among studies. Literature suggests that Internet or interactive computer-based education increase patients' knowledge and health information competence, and positively affect patient satisfaction.
Authors' conclusions	No clear effect on patient outcome measures could be identified, because effects differed across studies
Saksena [36]	
Conditions included	Hypertension
Geographic area	Europe, North America, Asia
Service/ intervention	Computer-based education
Outcome measures	Health, self-management
Authors' summary of results	5 studies (4 RCTs, 1 CCT) were included, assessing the effects of computer-based education on knowledge, self-management, and blood pressure control. Different control conditions were used, including usual care, pamphlet and website registration, and searching Yahoo. Meta-analysis was considered inappropriate due to substantial heterogeneity among studies. Quantitative analysis showed significant improvement of knowledge in 3 studies; self-management improved significantly in another study. Only 1 study demonstrated a significant increase of patients with controlled blood pressure before and after the intervention.
Authors' conclusions	Computer-based education is insufficient to replace provider-based education
Shulman et al [37]	

Study characteristic	Result[a]
Conditions included	Type 1 diabetes mellitus
Geographic area	Not stated
Service/ intervention	Home telemonitoring, education
Outcome measures	Health, costs, patient satisfaction
Authors’ summary of results	10 RCTs were included, comparing a range of technologies involving transmission of blood glucose data followed by unsolicited scheduled clinician feedback to usual care in youth with type 1 diabetes mellitus. Some studies incorporated an educational co-intervention. Meta-analysis showed no significant effect of the interventions on HbA1c, severe hypoglycemia or diabetic ketoacidosis. Limited data on patient satisfaction, quality of life, and costs also suggest no group differences.
Authors’ conclusions	It is unlikely that eHealth interventions have a substantial effect on glycemic control or acute complications
Welsh et al [38]	
Conditions included	Asthma
Geographic area	North America, Oceania
Service/ intervention	Home-based education
Outcome measures	Health, costs
Authors’ summary of results	12 RCTs were included, comparing home-based education to usual care in children with asthma. Meta-analysis showed no significant difference in the mean number of exacerbations requiring emergency department visits. Narrative analysis demonstrated improvement of quality of life in both groups over time. None of the studies analyzed cost-effectiveness.
Authors’ conclusions	Inconsistent evidence was found for home-based education compared to usual care
Wootton [39]	
Conditions included	Chronic diseases
Geographic area	Not stated
Service/intervention	Telemedicine interventions, including home telemonitoring, structured telephone phone support, video-teleconferencing, education and self-management programs
Outcome measures	Health, costs, self-management
Authors’ summary of results	In total, 141 RCTs were included. Meta-analysis was considered inappropriate due to substantial heterogeneity among studies. Most studies reported positive or weakly positive effects, and almost none reported negative effects. There was no significant difference in effect between diagnosis groups. There have been very few studies on cost-effectiveness.
Authors’ conclusions	The evidence base for the value of eHealth is generally weak and contradictory

[a] CBA: cost-benefit analysis.; CCA: cost-consequences analysis; CCT: nonrandomized controlled clinical trial; CEA: cost-effective analysis; CHF: congestive heart failure; CMA: cost-minimization analysis; COPD: chronic obstructive pulmonary disease; COT: randomized crossover trial; CUA: cost-utility analysis; HbA1c: glycosylated hemoglobin; NCBA: nonrandomized controlled before-after study; NUBA: nonrandomized uncontrolled before-after study; RCT: randomized controlled trial; STS: structured telephone support; TM: telemonitoring.

Table 6. Characteristics of 14 systematic reviews in which a meta-analysis was performed.

Study	Design (n)[a]	Condition[b]	Intervention[c]	Control[c]	Outcome[d]	Effect size (95% CI)[e]	Heterogeneity[f]
Meta-analyses reporting eHealth interventions are effective/cost-effective							
Agarwal et al [11]	RCT (5)	HT	Home TM	Office-based monitoring	SBP	SMD=–3.20 (–4.66, –1.73)	NR
					DBP	SMD=–1.63 (–2.47, –0.79)	NR
					Blood pressure response	RR=1.17 (1.02, 1.34)	NR
					Antihypertensive drug use	RR=2.18 (0.20, 23.68)	NR
Angeles et al [12]	RCT (9)	Type 1 and 2 DM	Web-based education	Usual care	HbA1c	3 m: WMD=–0.71 (–1.00, –0.43)/6 m: WMD=–0.52 (–0.75, –0.29)/12 m: WMD=–0.55 (–0.70, –0.39)	0/0/78
					FPG	3 m: WMD=–0.47 (–1.30, 0.35)/12 m: WMD=–0.80 (–2.81, 1.20)	0/85
					LDL-cholesterol	WMD=–0.23 (–0.28, –0.19)	11
					HDL-cholesterol	WMD=–0.00 (–0.15, 0.15)	83
					Total cholesterol	WMD=–0.14 (–0.53, 0.25)	70
Clarke et al [13]	RCT (13)	CHF	Home TM	Usual care	All-cause mortality	RR=0.77 (0.61, 0.97)	51
					All-cause hospital admissions	RR=0.99 (0.88, 1.11)	59
					CHF-related hospital admissions	RR=0.73 (0.62, 0.87)	0
					All-cause ED visits	RR=1.04 (0.86, 1.26)	82
Inglis et al [14]	RCT (11)	CHF	Home TM	Usual care	All-cause mortality	RR=0.66 (0.54, 0.81)	0
					All-cause hospital admissions	RR=0.91 (0.84, 0.99)	78
					CHF-related hospital admissions	RR=0.79 (0.67, 0.94)	39
	RCT (16)	CHF	STS	Usual care	All-cause mortality	RR=0.88 (0.76, 1.01)	0
					All-cause hospital admissions	RR=0.92 (0.85, 0.99)	24
					CHF-related hospital admissions	RR=0.77 (0.68, 0.87)	7
Klersy et al [15]	RCT (20)	CHF	Home TM, STS	Usual care	All-cause mortality	RR=0.83 (0.73, 0.95)	0
					All-cause hospital admissions	RR=0.93 (0.87, 0.99)	18
					CHF-related hospital admissions	RR=0.71 (0.64, 0.80)	2
Liang et al [16]	RCT (11)	Type 1 and 2 DM	Mobile phone interventions	NR	HbA1c	SMD=0.5 (0.2, 0.8)	NR
Polisena et al [17]	RCT (11)	CHF	Home TM	Usual care	All-cause mortality	RR=0.60 (0.45, 0.81)	0
Meta-analyses reporting promising evidence on effectiveness/cost-effectiveness of eHealth interventions							

Study	Design (n)[a]	Condition[b]	Intervention[c]	Control[c]	Outcome[d]	Effect size (95% CI)[e]	Heterogeneity[f]
McGeary et al [22]	RCT (9), COT (1)	Chronic pain	STS, VTC, self-management programs and outpatient telerehabilitation	Usual care or waiting list	Pain intensity ratings	*z*=–4.74 (–0.9, –0.4)	Q=15.73 (*P*=.07)
McLean et al [23]	RCT (10)	COPD	Home TM, STS, VTC and self-management programs	Usual care or face-to-face home visits	All-cause mortality	OR=1.05 (0.63, 1.75)	0
					All-cause hospital admissions	OR=0.46 (0.33, 0.65)	0
					All-cause ED visits	OR=0.27 (0.11, 0.66)	77
					Quality of life[g]	SMD=–6.57 (–13.62, 0.48)	51
Omboni et al [24]	RCT (12)	HT	Home TM, education	Usual care	SBP	WMD=5.64 (7.92, 3.36)	66
					DBP	WMD=2.78 (3.93, 1.62)	57
					Blood pressure control	RR=1.31 (1.06, 1.62)	78
					Antihypertensive drug use	WMD=0.22 (0.02, 0.43)	79
Samoocha et al [27]	RCT (14)	Any	Web-based education and self-management programs	Usual care, waiting list, or no care	Empowerment[h]	SMD=0.61 (0.29, 0.94)	0
					Self-efficacy (disease-specific)	SMD=0.23 (0.12, 0.33)	27
					Self-efficacy (general)	SMD=0.05 (–0.25, 0.35)	27
Meta-analyses reporting lacking, limited or inconsistent evidence on the effectiveness/cost-effectiveness of eHealth interventions							
McLean et al [6]	RCT (21)	Asthma	Home TM, STS, VTC, education and self-management programs	(Enhanced) face-to-face usual care	All-cause hospital admissions	3 m: OR=0.47 (0.01, 36.46)/12 m: OR=0.21 (0.07, 0.61)	84/0
					All-cause ED visits	OR=1.16 (0.52, 2.58)	29
					Quality of life[i]	WMD=0.08 (0.01, 0.16)	24
Shulman et al [37]	RCT (10)	Type 1 DM	Home TM, education	NR	HbA1c	WMD=–0.12 (–0.35, 0.11)	0
					Severe hypoglycemia	OR=1.42 (0.22, 9.32)	0
					Diabetic ketoacidosis	OR=1.02 (0.24, 4.23)	0

Study	Design (n)[a]	Condition[b]	Intervention[c]	Control[c]	Outcome[d]	Effect size (95% CI)[e]	Heterogeneity[f]
Welsh et al [38]	RCT (12)	Asthma	Home-based education	Usual care	Asthma-related ED visits	6 m: WMD=0.04 (–0.20, 0.27)/12-18 m: WMD=–0.32 (–0.74, 0.10)	NR/NR

[a] RCT: randomized controlled trial; COT: randomized crossover trial.

[b] HT: hypertension; CHF: congestive heart failure; DM: diabetes mellitus; COPD: chronic obstructive pulmonary disease.

[c] TM: telemonitoring; STS: structured telephone support; VTC: video-teleconferencing; NR: not reported.

[d] SBP: systolic blood pressure; DBP: diastolic blood pressure; HbA1c: glycosylated hemoglobin; FPG: fasting plasma glucose; LDL: low-density lipoprotein; HDL: high-density lipoprotein; ED: emergency department.

[e] SMD: standardized mean difference also known as Cohen’s *d*; a conventional rule is to consider a Cohen’s *d* of 0.2 as small, 0.5 as medium, and 0.8 as large [41]; RR: relative risk; WMD: weighted mean difference.

[f] Unless otherwise indicated, data are presented as a percentage of variation across studies that is due to heterogeneity (ie, I^2 statistic). An I^2 value greater than 50 is considered as substantial heterogeneity and may indicate that quantitative analysis is inappropriate [17,42].

[g] Using the St. George’s Respiratory Questionnaire.

[h] Using the Diabetes Empowerment Scale.

[i] Using the Juniper Asthma Quality of Life Questionnaire.

Methodological Quality of Reviews and Meta-Analyses

Among the systematic reviews and meta-analyses described in the current review are 4 high-quality Cochrane reviews [6,14,23,38]. Following the PRISMA Statement [10], the methodological quality of the other included papers was generally considered high. Nearly all authors provided search queries and selection criteria, described the process of data extraction, presented the results and limitations of individual studies, and demonstrated the implications of their outcome for daily practice and future research. If the authors received external funding, this was reported.

Some discrepancy between reviews was observed in terms of defining eHealth. For example, McLean et al [6] excluded Web-based tools and interventions for self-management in their Cochrane review on asthma patients because health care professionals were not actively involved with the ongoing delivery of the intervention. McGeary et al [22] chose a broader definition in their work on telehealth trials in pain management, including all studies that assessed a technology-based intervention extending care beyond the health care professional’s office.

Many authors did not conduct a meta-analysis because of important differences perceived in study populations, interventions and outcome measures [18,19,25,26,31-33,35,36,39]. Instead, they performed a qualitative analysis of their findings. Several papers presented the results of a pooled analysis or subanalysis, despite substantial heterogeneity (ie, I^2 value >50) [6,12-14,23,24]. Three studies did not report heterogeneity [11,16,38].

Trend Analysis

Since the publication by Ekeland et al in 2010 [9], the number of systematic reviews and meta-analyses on eHealth interventions in patients with somatic diseases has grown considerably (Table 7). In addition, 4 Cochrane reviews have recently been published [6,14,23,38]. However, a similar percentage of papers concluded that eHealth is effective/cost-effective or evidence is at least promising (65% vs 62%). Reviews focusing primarily on children or family caregivers still remain scarce. Between 2009 and 2012, home telemonitoring and video-teleconferencing were less frequently subject to a systematic review and/or meta-analysis on eHealth interventions, whereas educational tools and self-management programs were encountered more often. Data on economic outcome measures were less frequently reported in recent papers. Other study characteristics (eg, geographic area) barely differed between our review and the review by Ekeland et al [9].

Table 7. Trend analysis of differences in study characteristics of the current review compared with the review by Ekeland et al [9] published in 2010.

Study characteristic	Current review	Ekeland et al
Inclusion period, years	2009-2012	2005-2009
Systematic reviews, n	31	26
Meta-analyses, n (%)	14 (45)	7 (27)
Cochrane reviews, n (%)	4 (13)	0 (0)
Study population, n (%)		
Children/adolescents only	3 (10)	0 (0)
Family caregivers only	1 (3.2)	1 (3.8)
Geographic area, n (%)		
Europe	15 (48)	17 (65)
North America	18 (58)	18 (69)
Asia	11 (35)	8 (31)
Oceania	4 (13)	5 (19)
Latin America	3 (10)	4 (15)
Africa	0 (0)	0 (0)
Not explicitly stated	13 (42)	8 (31)
Authors' conclusions, n (%)		
Effective/cost-effective	7 (23)	8 (31)
Promising	13 (42)	8 (31)
Limited/inconsistent	11 (35)	10 (38)
Outcome measure, n (%)		
Health	28 (90)	24 (92)
Costs	20 (65)	22 (85)
Patient satisfaction	17 (55)	16 (62)
Self-management	16 (52)	14 (54)
Intervention components, n (%)		
Home telemonitoring	19 (61)	21 (81)
Structured telephone support	10 (32)	12 (46)
Video-teleconferencing	8 (26)	11 (42)
Education	17 (55)	12 (46)
Self-management programs	11 (35)	7 (27)
Telerehabilitation	5 (16)	4 (15)
Telemedicine (not otherwise specified)	1 (3.2)	0 (0)

Discussion

The term eHealth can be defined briefly as the delivery of personalized health care at a distance through the use of technology. It is hypothesized that this field of medicine potentially enhances the quality of health care, with simultaneous reduction of health care costs. To support this hypothesis, we undertook a systematic review of systematic reviews and meta-analyses on the effectiveness/cost-effectiveness of eHealth interventions in patients with somatic diseases. In addition, we performed a trend analysis to compare current findings with results from a systematic review by Ekeland et al published in 2010 [9].

In recent years, literature on eHealth has accumulated considerably. We found a total of 31 reviews, of which 20 (65%) concluded that eHealth interventions are effective/cost-effective or evidence is at least promising. Only 11 reviews (35%) showed no, limited, or inconsistent proof. These findings are consistent with the results from the review by Ekeland et al [9] (Table 7). Furthermore, trend analysis shows reviews focusing primarily on children or family caregivers still remain scarce. Although a pooled (subgroup) analysis of aggregate data from randomized studies was performed in a higher percentage of more recently published reviews (45% vs 27%), data on economic outcome measures were less frequently reported (65% vs 85%).

Because our review is a systematic review of systematic reviews and meta-analyses, it holds 2 important limitations. Firstly, we relied on the adequate inclusion and critical appraisal of individual studies, as well as on a correct interpretation of study results by the authors of the reviews and meta-analyses included in the current review. We did not investigate whether reviews on similar topics comprised identical studies; neither did we examine possible discrepancies in the analyses of these individual studies when included in more than one review or meta-analysis. Noteworthy, systematic reviews of systematic reviews have been conducted before in other fields of medicine, including reconstructive surgery and neuroradiology [43,44].

Secondly, reviews differed substantially in terms of study populations, intervention components, comparison groups, and outcome measures, for example. Therefore, it is difficult to identify which patients are likely to benefit from which specific intervention. Home telemonitoring and structured telephone support seemed to be effective/cost-effective in patients with CHF (Table 3), whereas evidence on both interventions seemed limited or inconsistent in patients with chronic pulmonary diseases (Table 5). Meta-analysis was often impeded because of heterogeneity among individual studies. This may have demanded careful conclusions from the authors of that particular review. In several reviews, a pooled (subgroup) analyses was presented despite substantial heterogeneity among individual studies (ie, I^2 value >50) [17,42]. Publication bias may have been the result of the exclusion of small individual studies with negative results, which could have ultimately lead to overestimation of benefits [45,46]. Noteworthy, Ciere et al [32] proposed methodological weaknesses may be partially because of artifacts of poor reporting, rather than being a reflection of poor study design or implementation.

Regarding the aforementioned methodological shortcomings, Ekeland et al [47] performed a systematic review in which they summarize methodologies used in research on eHealth interventions, discuss knowledge gaps, and postulate recommendations for methodological approaches for future research. Furthermore, we agree with recommendations made in previous reports to overcome the problem of between-study differences: researchers should adhere to and make transparent use of reporting guidelines appropriate for specific study designs. These guidelines may include Consolidated Standards of Reporting Trials (CONSORT)-EHEALTH for RCTs on eHealth interventions, Transparent Reporting of Evaluations with Nonrandomized Designs (TREND) and Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) for observational studies in general, and Workgroup for Intervention Development and Evaluation Research (WIDER) recommendations for the reporting of behavioral interventions [32,48-51].

Because pilot schemes are often limited to fewer than 100 patients, many researchers in the past decade have advocated larger RCTs with standardized study designs to provide definite proof on the effectiveness/cost-effectiveness of eHealth interventions. Results of the recent Whole System Demonstrator trial—involving 3230 patients with diabetes mellitus, COPD, and CHF—showed that eHealth interventions are associated with lower mortality and emergency admission rates [52]. In our opinion, these results should provide an important stimulus to invest in the incorporation of eHealth in daily practice. However, implementation difficulties, such as resistant or refractory behaviors of health care professionals, are an international phenomenon [53]. The Normalization Process Theory (NPT), a sociological theory that provides a framework for understanding the relationship between technology and the social environment, has been used to develop implementation tools such as the eHealth Implementation Toolkit (E-HIT) [54,55].

Although large, well-designed RCTs are likely to further support the evidence on the effectiveness/cost-effectiveness of eHealth initiatives, we believe it is more desirable to focus on overcoming the problematic gap between pilot schemes and daily practice. As proposed in both reviews by Ekeland et al [9,47], formative process assessments and complexity studies can be further explored to achieve this goal.

In conclusion, the number of reviews and meta-analyses on the effectiveness/cost-effectiveness of eHealth interventions in somatic diseases has increased considerably in recent years. The majority of these papers show eHealth is effective/cost-effective, or at least suggests evidence is promising, which is consistent with previous findings. Data on economic outcome measures were less frequently reported in articles that were published more recently. This is an interesting finding, given the importance of formal cost analyses when considering implementation of eHealth interventions in daily practice. Although many researchers advocate larger, well-designed, controlled studies, we believe attention should be given to the development and evaluation of strategies to implement effective/cost-effective eHealth initiatives, rather than to further strengthen the evidence that has already been made available.

Acknowledgments

This paper received no sources of funding or sponsorship and there is no financial disclosure.

Conflicts of Interest

None declared.

References

1. Einthoven W. Het tele-cardiogram. Ned Tijdschr Geneeskd 1906;50:1517-1547.
2. Mariani AW, Pêgo-Fernandes PM. Telemedicine: a technological revolution. Sao Paulo Med J 2012;130(5):277-278 [FREE Full text] [Medline: 23174864]
3. Rinde E, Balteskard L. Is there a future for telemedicine? Lancet 2002;359(9322):1957-1958. [doi: 10.1016/S0140-6736(02)08845-1] [Medline: 12076546]
4. American Telemedicine Association. What is telemedicine? URL: http://www.americantelemed.org/about-telemedicine/what-is-telemedicine [accessed 2014-03-31] [WebCite Cache ID 6OTveZBA7]
5. Miller EA. Solving the disjuncture between research and practice: telehealth trends in the 21st century. Health Policy 2007;82(2):133-141. [doi: 10.1016/j.healthpol.2006.09.011] [Medline: 17046097]
6. McLean S, Chandler D, Nurmatov U, Liu J, Pagliari C, Car J, et al. Telehealthcare for asthma. Cochrane Database Syst Rev 2010(10):CD007717. [doi: 10.1002/14651858.CD007717.pub2] [Medline: 20927763]
7. Finkelstein J, Friedman RH. Potential role of telecommunication technologies in the management of chronic health conditions. Dis Manag Health Out 2000;8(2):57-63. [doi: 10.2165/00115677-200008020-00001]
8. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. Int J Nurs Stud 2013;50(5):587-592. [doi: 10.1016/j.ijnurstu.2012.09.010] [Medline: 23159157]
9. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. Int J Med Inform 2010;79(11):736-771. [doi: 10.1016/j.ijmedinf.2010.08.006] [Medline: 20884286]
10. 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;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
11. Agarwal R, Bills JE, Hecht TJ, Light RP. Role of home blood pressure monitoring in overcoming therapeutic inertia and improving hypertension control: a systematic review and meta-analysis. Hypertension 2011;57(1):29-38 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.110.160911] [Medline: 21115879]
12. Angeles RN, Howard MI, Dolovich L. The effectiveness of web-based tools for improving blood glucose control in patients with diabetes mellitus: a meta-analysis. Can J Diabetes 2011;35(4):344-352. [doi: 10.1016/S1499-2671(11)54011-0]
13. Clarke M, Shah A, Sharma U. Systematic review of studies on telemonitoring of patients with congestive heart failure: a meta-analysis. J Telemed Telecare 2011;17(1):7-14. [doi: 10.1258/jtt.2010.100113] [Medline: 21097564]
14. Inglis SC, Clark RA, McAlister FA, Ball J, Lewinter C, Cullington D, et al. Structured telephone support or telemonitoring programmes for patients with chronic heart failure. Cochrane Database Syst Rev 2010(8):CD007228. [doi: 10.1002/14651858.CD007228.pub2] [Medline: 20687083]
15. Klersy C, De Silvestri A, Gabutti G, Regoli F, Auricchio A. A meta-analysis of remote monitoring of heart failure patients. J Am Coll Cardiol 2009;54(18):1683-1694. [doi: 10.1016/j.jacc.2009.08.017] [Medline: 19850208]
16. Liang X, Wang Q, Yang X, Cao J, Chen J, Mo X, et al. Effect of mobile phone intervention for diabetes on glycaemic control: a meta-analysis. Diabet Med 2011;28(4):455-463. [doi: 10.1111/j.1464-5491.2010.03180.x] [Medline: 21392066]
17. Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K, et al. Home telemonitoring for congestive heart failure: a systematic review and meta-analysis. J Telemed Telecare 2010;16(2):68-76. [doi: 10.1258/jtt.2009.090406] [Medline: 20008054]
18. Bender JL, Radhakrishnan A, Diorio C, Englesakis M, Jadad AR. Can pain be managed through the Internet? A systematic review of randomized controlled trials. Pain 2011;152(8):1740-1750. [doi: 10.1016/j.pain.2011.02.012] [Medline: 21565446]
19. Eland-de Kok P, van Os-Medendorp H, Vergouwe-Meijer A, Bruijnzeel-Koomen C, Ros W. A systematic review of the effects of e-health on chronically ill patients. J Clin Nurs 2011;20(21-22):2997-3010. [doi: 10.1111/j.1365-2702.2011.03743.x] [Medline: 21707807]
20. Hailey D, Roine R, Ohinmaa A, Dennett L. Evidence of benefit from telerehabilitation in routine care: a systematic review. J Telemed Telecare 2011;17(6):281-287. [doi: 10.1258/jtt.2011.101208] [Medline: 21844172]
21. Johansson T, Wild C. Telerehabilitation in stroke care--a systematic review. J Telemed Telecare 2011;17(1):1-6. [doi: 10.1258/jtt.2010.100105] [Medline: 21097560]
22. McGeary DD, McGeary CA, Gatchel RJ, Allison S, Hersh A. Assessment of research quality of telehealth trials in pain management: a meta-analysis. Pain Pract 2013;13(5):422-431. [doi: 10.1111/j.1533-2500.2012.00601.x] [Medline: 23017210]
23. McLean S, Nurmatov U, Liu JL, Pagliari C, Car J, Sheikh A. Telehealthcare for chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2011(7):CD007718. [doi: 10.1002/14651858.CD007718.pub2] [Medline: 21735417]
24. Omboni S, Guarda A. Impact of home blood pressure telemonitoring and blood pressure control: a meta-analysis of randomized controlled studies. Am J Hypertens 2011;24(9):989-998. [doi: 10.1038/ajh.2011.100] [Medline: 21654858]
25. Paré G, Moqadem K, Pineau G, St-Hilaire C. Clinical effects of home telemonitoring in the context of diabetes, asthma, heart failure and hypertension: a systematic review. J Med Internet Res 2010;12(2):e21 [FREE Full text] [doi: 10.2196/jmir.1357] [Medline: 20554500]
26. Rietdijk R, Togher L, Power E. Supporting family members of people with traumatic brain injury using telehealth: a systematic review. J Rehabil Med 2012;44(11):913-921 [FREE Full text] [doi: 10.2340/16501977-1058] [Medline: 23027316]

27. 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]
28. Stinson J, Wilson R, Gill N, Yamada J, Holt J. A systematic review of internet-based self-management interventions for youth with health conditions. J Pediatr Psychol 2009;34(5):495-510 [FREE Full text] [doi: 10.1093/jpepsy/jsn115] [Medline: 19029142]
29. van den Berg N, Schumann M, Kraft K, Hoffmann W. Telemedicine and telecare for older patients--a systematic review. Maturitas 2012;73(2):94-114. [doi: 10.1016/j.maturitas.2012.06.010] [Medline: 22809497]
30. Baron J, McBain H, Newman S. The impact of mobile monitoring technologies on glycosylated hemoglobin in diabetes: a systematic review. J Diabetes Sci Technol 2012;6(5):1185-1196 [FREE Full text] [Medline: 23063046]
31. Bolton CE, Waters CS, Peirce S, Elwyn G, EPSRCMRC Grand Challenge Team. Insufficient evidence of benefit: a systematic review of home telemonitoring for COPD. J Eval Clin Pract 2011;17(6):1216-1222. [doi: 10.1111/j.1365-2753.2010.01536.x] [Medline: 20846317]
32. Ciere Y, Cartwright M, Newman SP. A systematic review of the mediating role of knowledge, self-efficacy and self-care behaviour in telehealth patients with heart failure. J Telemed Telecare 2012;18(7):384-391. [doi: 10.1258/jtt.2012.111009] [Medline: 23019605]
33. Franek J. Home telehealth for patients with chronic obstructive pulmonary disease (COPD): an evidence-based analysis. Ont Health Technol Assess Ser 2012;12(11):1-58 [FREE Full text] [Medline: 23074421]
34. Mistry H. Systematic review of studies of the cost-effectiveness of telemedicine and telecare. Changes in the economic evidence over twenty years. J Telemed Telecare 2012;18(1):1-6. [doi: 10.1258/jtt.2011.110505] [Medline: 22101609]
35. Ryhänen AM, Siekkinen M, Rankinen S, Korvenranta H, Leino-Kilpi H. The effects of Internet or interactive computer-based patient education in the field of breast cancer: a systematic literature review. Patient Educ Couns 2010;79(1):5-13. [doi: 10.1016/j.pec.2009.08.005] [Medline: 19744817]
36. Saksena A. Computer-based education for patients with hypertension: a systematic review. Health Educ J 2010;69(3):236-245. [doi: 10.1177/0017896910364889]
37. Shulman RM, O'Gorman CS, Palmert MR. The impact of telemedicine interventions involving routine transmission of blood glucose data with clinician feedback on metabolic control in youth with type 1 diabetes: a systematic review and meta-analysis. Int J Pediatr Endocrinol 2010;2010:pii: 536957 [FREE Full text] [doi: 10.1155/2010/536957] [Medline: 20886054]
38. Welsh EJ, Hasan M, Li P. Home-based educational interventions for children with asthma. Cochrane Database Syst Rev 2011(10):CD008469. [doi: 10.1002/14651858.CD008469.pub2] [Medline: 21975783]
39. Wootton R. Twenty years of telemedicine in chronic disease management--an evidence synthesis. J Telemed Telecare 2012;18(4):211-220 [FREE Full text] [doi: 10.1258/jtt.2012.120219] [Medline: 22674020]
40. Pron G, Ieraci L, Kaulback K, Medical Advisory Secretariat, Health Quality Ontario. Internet-based device-assisted remote monitoring of cardiovascular implantable electronic devices: an evidence-based analysis. Ont Health Technol Assess Ser 2012;12(1):1-86 [FREE Full text] [Medline: 23074419]
41. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: L Erlbaum Associates; 1988.
42. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21(11):1539-1558. [doi: 10.1002/sim.1186] [Medline: 12111919]
43. Thoma A, Veltri K, Haines T, Duku E. A systematic review of reviews comparing the effectiveness of endoscopic and open carpal tunnel decompression. Plast Reconstr Surg 2004;113(4):1184-1191. [Medline: 15083019]
44. Clarke M. Systematic review of reviews of risk factors for intracranial aneurysms. Neuroradiology 2008;50(8):653-664. [doi: 10.1007/s00234-008-0411-9] [Medline: 18560819]
45. Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA 2004;291(20):2457-2465. [doi: 10.1001/jama.291.20.2457] [Medline: 15161896]
46. Eysenbach G. Tackling publication bias and selective reporting in health informatics research: register your eHealth trials in the International eHealth Studies Registry. J Med Internet Res 2004;6(3):e35 [FREE Full text] [doi: 10.2196/jmir.6.3.e35] [Medline: 15471761]
47. Ekeland AG, Bowes A, Flottorp S. Methodologies for assessing telemedicine: a systematic review of reviews. Int J Med Inform 2012;81(1):1-11. [doi: 10.1016/j.ijmedinf.2011.10.009] [Medline: 22104370]
48. 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]
49. Des Jarlais DC, Lyles C, Crepaz N, TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health 2004;94(3):361-366. [Medline: 14998794]

50. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Bull World Health Organ 2007;85(11):867-872 [FREE Full text] [Medline: 18038077]
51. WIDER Group. Implementation Science. Recommendations to improve reporting of the content of behaviour change interventions URL: http://www.implementationscience.com/content/supplementary/1748-5908-7-70-s4.pdf [accessed 2014-03-31] [WebCite Cache ID 6OTy35W5h]
52. Steventon A, Bardsley M, Billings J, Dixon J, Doll H, Hirani S, Whole System Demonstrator Evaluation Team. Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. BMJ 2012;344:e3874 [FREE Full text] [Medline: 22723612]
53. Murray E, Burns J, May C, Finch T, O'Donnell C, Wallace P, et al. Why is it difficult to implement e-health initiatives? A qualitative study. Implement Sci 2011;6:6 [FREE Full text] [doi: 10.1186/1748-5908-6-6] [Medline: 21244714]
54. May C, Finch T. Implementing, embedding, and integrating practices: an outline of Normalization Process Theory. Sociology 2009;43(3):535-554. [doi: 10.1177/0038038509103208]
55. MacFarlane A, Clerkin P, Murray E, Heaney DJ, Wakeling M, Pesola UM, et al. The e-Health Implementation Toolkit: qualitative evaluation across four European countries. Implement Sci 2011;6:122 [FREE Full text] [doi: 10.1186/1748-5908-6-122] [Medline: 22098945]

Abbreviations

CHF: congestive heart failure
CONSORT: Consolidated Standards of Reporting Trials
COPD: chronic obstructive pulmonary disease
E-HIT: eHealth Implementation Toolkit
HbA1c: glycosylated hemoglobin
ICU: intensive care unit
LDL: low-density lipoprotein
NPT: Normalization Process Theory
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QUOROM: Quality of Reporting of Meta-analyses
RCT: randomized controlled trial
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
TREND: Transparent Reporting of Evaluations with Nonrandomized Designs
WIDER: Workgroup for Intervention Development and Evaluation Research

Edited by G Eysenbach; submitted 24.06.13; peer-reviewed by A Straten, R Choubisa; comments to author 12.07.13; revised version received 13.02.14; accepted 13.03.14; published 16.04.14.

Please cite as:
Elbert NJ, van Os-Medendorp H, van Renselaar W, Ekeland AG, Hakkaart-van Roijen L, Raat H, Nijsten TEC, Pasmans SGMA
Effectiveness and Cost-Effectiveness of eHealth Interventions in Somatic Diseases: A Systematic Review of Systematic Reviews and Meta-Analyses
J Med Internet Res 2014;16(4):e110
URL: http://www.jmir.org/2014/4/e110/
doi:10.2196/jmir.2790
PMID:24739471

Original Paper

Electronic Versus Paper-Based Assessment of Health-Related Quality of Life Specific to HIV Disease: Reliability Study of the PROQOL-HIV Questionnaire

Martin Duracinsky[1,2], MD, PhD; Christophe Lalanne[1,3], PhD; Cécile Goujard[2,4], MD, PhD; Susan Herrmann[5], PhD; Christian Cheung-Lung[2], BPharm; Jean-Paul Brosseau[2,6], MD; Yannick Schwartz[7,8], PhD; Olivier Chassany[1,9], MD, PhD

[1]Université Paris-Diderot, Sorbonne Paris-Cité, Unité de Méthodologie des critères d'évaluation (Patient-Reported Outcomes), EA Recherche Clinique Coordonnée Ville-Hôpital, Méthodologies et Société (ED 393), Paris cedex 10, France

[2]Hopital Universitaire de Bicetre, Service de Médecine Interne et de Maladies Infectieuses, Assistance Publique - Hopitaux de Paris (AP-HP), Le Kremlin-Bicetre, France

[3]Inserm Unit UMR-SO 669, University Paris Sud, Paris Descartes, Paris, France

[4]Paris Descartes, University Paris Sud, Le Kremlin-Bicêtre, France

[5]Institute for Immunology and Infectious Diseases, Murdoch University, Western Australia, Perth, Australia

[6]Fondation Maison des champs, Service ACT du Val de Marne, Val de Marne, France

[7]Parietal Team, Inria Saclay–Ile de France, Saclay, France

[8]Neurospin, CEA/DSV/I2BM, Gif-sur-Yvette, France

[9]Department of Clinical Research, Saint-Louis Hospital, AP-HP, Paris, France

Corresponding Author:
Martin Duracinsky, MD, PhD
Université Paris-Diderot, Sorbonne Paris-Cité
Unité de Méthodologie des critères d'évaluation (Patient-Reported Outcomes)
EA Recherche Clinique Coordonnée Ville-Hôpital, Méthodologies et Société (ED 393)
Carré historique de l'Hôpital Saint-Louis, porte 23
1 avenue Claude Vellefaux
Paris cedex 10, 75475
France
Phone: 33 1 40 27 57 48
Fax: 33 1 44 84 17 88
Email: duracinsky.m@gmail.com

Abstract

Background: Electronic patient-reported outcomes (PRO) provide quick and usually reliable assessments of patients' health-related quality of life (HRQL).

Objective: An electronic version of the Patient-Reported Outcomes Quality of Life-human immunodeficiency virus (PROQOL-HIV) questionnaire was developed, and its face validity and reliability were assessed using standard psychometric methods.

Methods: A sample of 80 French outpatients (66% male, 52/79; mean age 46.7 years, SD 10.9) were recruited. Paper-based and electronic questionnaires were completed in a randomized crossover design (2-7 day interval). Biomedical data were collected. Questionnaire version and order effects were tested on full-scale scores in a 2-way ANOVA with patients as random effects. Test-retest reliability was evaluated using Pearson and intraclass correlation coefficients (ICC, with 95% confidence interval) for each dimension. Usability testing was carried out from patients' survey reports, specifically, general satisfaction, ease of completion, quality and clarity of user interface, and motivation to participate in follow-up PROQOL-HIV electronic assessments.

Results: Questionnaire version and administration order effects (N=59 complete cases) were not significant at the 5% level, and no interaction was found between these 2 factors (P=.94). Reliability indexes were acceptable, with Pearson correlations greater than .7 and ICCs ranging from .708 to .939; scores were not statistically different between the two versions. A total of 63 (79%) complete patients' survey reports were available, and 55% of patients (30/55) reported being satisfied and interested in electronic assessment of their HRQL in clinical follow-up. Individual ratings of PROQOL-HIV user interface (85%-100% of positive responses) confirmed user interface clarity and usability.

Conclusions: The electronic PROQOL-HIV introduces minor modifications to the original paper-based version, following International Society for Pharmacoeconomics and Outcomes Research (ISPOR) ePRO Task Force guidelines, and shows good reliability and face validity. Patients can complete the computerized PROQOL-HIV questionnaire and the scores from the paper or electronic versions share comparable accuracy and interpretation.

(J Med Internet Res 2014;16(4):e115) doi:10.2196/jmir.3330

KEYWORDS

HIV; quality of life; patient-reported outcomes; electronic records; reliability

Introduction

Patient reports are useful to “recall observations, to inform others, to instruct students, to gain knowledge, to monitor performance, and to justify interventions” [1] and form part of clinical decision-making. With the advent of modern technologies, electronic health records (eg, biological measures, treatments, imaging results) are used increasingly, especially in managing chronic disease [2]. Information such as pain, fatigue, depression, and health-related quality of life (HRQL) are also gathered through self-reported measures, and are known as patient-reported outcomes (PRO) [3]. Gathering this information electronically is becoming increasingly common because electronic assessment provides users with direct feedback including secure storage in databases supporting access controls and role-based permissions, lower administrative costs, and easier follow-up of patients’ records over time. Moreover, electronic diaries or electronic PRO (ePRO) measurement provide quick, convenient, and reliable assessment of patients’ HRQL, and improve compliance with self-assessed HRQL [4] or at least help to understand reasons for noncompliance [5]. Because responses can be enforced (eg, patients are not able to submit a Web form or possibly skip to the next questionnaire if some of the questions were not answered), ePROs help to reduce the problem of missing data [6].

There is compelling evidence that electronic and paper-and-pencil PROs deliver equivalent measures [7], and sometimes electronic ones are more reliable [8,9], although some discrepancies between paper and electronic versions of the same questionnaire have been reported [10]. However, the Internet is used increasingly to seek information related to symptoms, HRQL, drug adverse events, or simply share self-experience with chronic disease. Human immunodeficiency virus (HIV) disease is now considered a chronic disease with costly treatment, and researchers are currently seeking solutions to optimize HRQL assessments not only in clinical trials, but also in clinical routine care. For example, intervention studies have demonstrated that adherence could be improved when monitored via mobile phones [11,12], suggesting that electronic health records and care management systems are promising approaches to improving HIV care.

Demonstrating equivalence between electronic and paper versions of PRO measurement is an essential step when a validated paper instrument is migrated to an electronic format, especially when both versions are to be used interchangeably. According to recent International Society for Pharmacoeconomics and Outcomes Research (ISPOR) guidelines [13], an ePRO questionnaire should deliver comparable or better data compared to a paper-and-pencil questionnaire, and the measurement of difference between the 2 data gathering methods is an essential feature of validation.

An electronic version of the Patient-Reported Outcomes Quality of Life-HIV (PROQOL-HIV) questionnaire [14,15] was developed to meet the challenges of the electronic health measures era. The aim of this study was to study the psychometric properties, especially face validity and reliability, of the electronic version of the PROQOL-HIV questionnaire and to suggest further refinements to the Web interface based on participant input and the technical issues encountered during the validation study. This study provides details about the participants, study setup, and principal results for HRQL data as well as users’ feedback.

Methods

Recruitment

The study was conducted in 2 centers in France: the Kremlin Bicêtre hospital (Assistance Publique-Hôpitaux de Paris) and the institutional apartments for people living with HIV, *Service ACT du Val de Marne, Fondation Maison des champs*.

Informed consent was obtained from all participants. The study was approved by a local ethics committee. Storage of individual patient data on a dedicated server was also approved by an independent French administrative authority in charge of personal data registration and protection (CNIL record #1566050). Inclusion criteria were French-speaking, HIV-seropositive outpatients receiving routine HIV clinical care, aged 18 years or older. People were excluded from the study if they attended the hospital for urgent care or were hospitalized with HIV-related illness.

Measurement Instrument

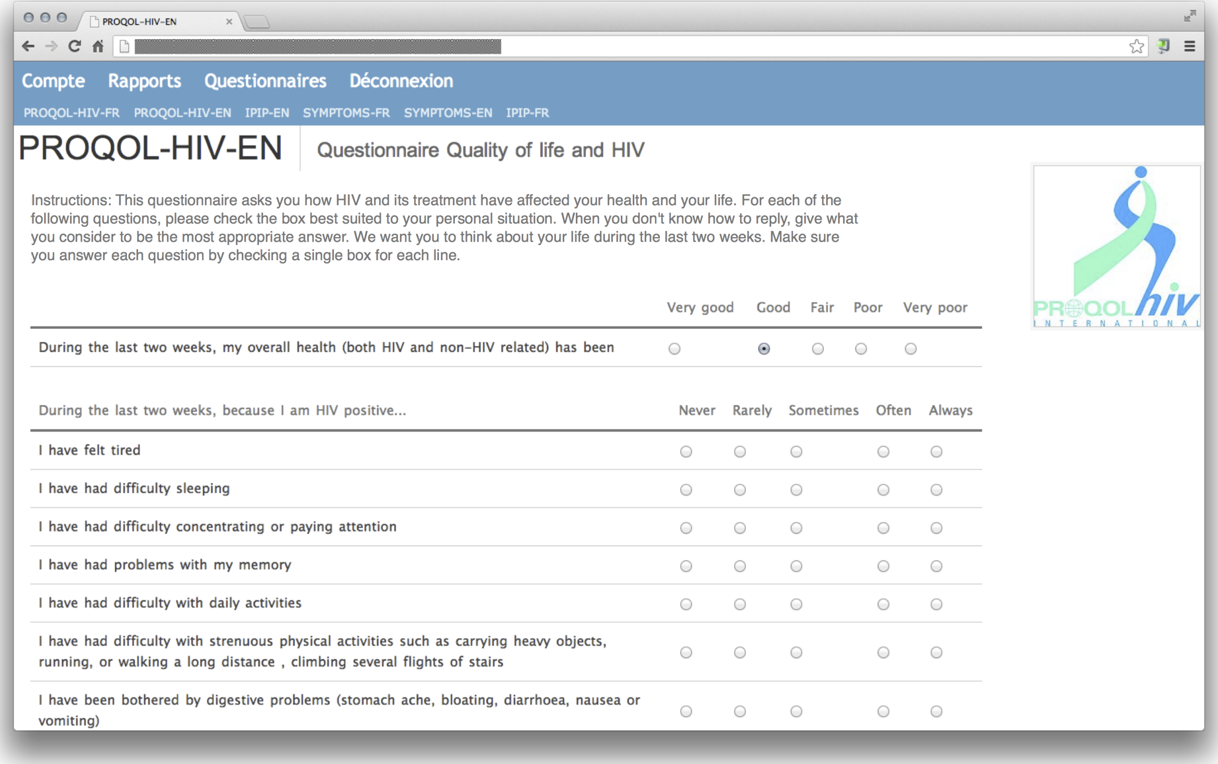
The PROQOL-HIV questionnaire [14,15] is composed of 43 Likert-type items (5-point scale ranging from 0=never to 4=always), including 39 items targeting 8 domains of HRQL and general health: physical health and symptoms (9 items), treatment impact (10 items), emotional distress (4 items), health concerns (4 items), body change (4 items), intimate relationships (3 items), social relationships (2 items), and stigma (2 items). Four extra items dealing with religious beliefs, finance, having children, and satisfaction with care are not part of the scoring scheme, but are used to gather additional information from the respondent. This questionnaire has been shown to exhibit good reliability with Cronbach alpha ranging from .772 to .885 and intraclass correlations greater than .7 for all dimensions with more than 2 items. Responses to items were totaled for each dimension and standardized on a scale from 0 to 100 points, in which higher values indicate a better health state.

The psychometric validation of the PROQOL-HIV questionnaire [15] included 123 French participants. In this sample, the estimated standard deviation for the physical health and symptoms scale was 20.1 points, with a Cronbach alpha of .885, suggesting that the standard error of measurement approached 6.8 points. This shows individual scores with a half-width confidence interval (or margin of error) of 13 points.

Web Interface

The Web interface was developed using the Python programming language and data were stored in a PostgreSQL database. The system makes use of dynamic Hypertext Markup Language (HTML; JQuery) to save answers instantly or highlight missing responses when validating entries. The Web version of the PROQOL-HIV questionnaire was developed to replicate the paper-based questionnaire as closely as possible. Compared with the original paper-based version, only minor modifications were made: the entire questionnaire was presented on a single HTML page (rather than 2 separate sheets of paper), and response options were presented as radio buttons on a horizontal grid (instead of checkboxes) with headings aligned on top of each section of the questionnaire. A sample screenshot of the Web questionnaire is provided (Figure 1). Before completing the Web form, users registered with a personal username and password on a log-in page. If no activity was detected after 5 minutes, the session was timed out to ensure data was protected. The uncompleted questionnaires were saved, but were not used in the analysis. Responses, timestamp, and username were saved in a secure database. Individual timestamps were kept for each selected response options, not simply for the questionnaire as a whole. Scale scores were computed directly on the server and were visually presented to the user at the end of the session through bar charts and a numerical table. For the purpose of this validation study, anonymized individual data were extracted from the database.

Figure 1. Screenshot of the Web interface for the English version of the PROQOL-HIV questionnaire.



PROQOL-HIV-EN Questionnaire Quality of life and HIV

Instructions: This questionnaire asks you how HIV and its treatment have affected your health and your life. For each of the following questions, please check the box best suited to your personal situation. When you don't know how to reply, give what you consider to be the most appropriate answer. We want you to think about your life during the last two weeks. Make sure you answer each question by checking a single box for each line.

	Very good	Good	Fair	Poor	Very poor
During the last two weeks, my overall health (both HIV and non-HIV related) has been	○	●	○	○	○

During the last two weeks, because I am HIV positive...	Never	Rarely	Sometimes	Often	Always
I have felt tired	○	○	○	○	○
I have had difficulty sleeping	○	○	○	○	○
I have had difficulty concentrating or paying attention	○	○	○	○	○
I have had problems with my memory	○	○	○	○	○
I have had difficulty with daily activities	○	○	○	○	○
I have had difficulty with strenuous physical activities such as carrying heavy objects, running, or walking a long distance , climbing several flights of stairs	○	○	○	○	○
I have been bothered by digestive problems (stomach ache, bloating, diarrhoea, nausea or vomiting)	○	○	○	○	○

Administration

Each patient received an information sheet describing the aim of the study, detailed instructions on how to connect to the website with a secure password, and a brief questionnaire to fill in about basic demographic and clinical information. Biological variables, such as viral load or lymphocyte CD4 count, were retrieved from patients' files. The paper and electronic HRQL questionnaires were completed in a randomized crossover design, with a 2- to 7-day interval. At the first meeting, patients completed the clinical and demographic questionnaire with the help of the nursing staff if needed. They were then asked to complete either the PROQOL-HIV paper version or the electronic version on a dedicated computer at the hospital, depending on the sequence to which they were randomized. A randomization list was established using a computer program (shuffled balanced binomial draws, as in 2-arm treatment allocation) before the beginning of the study. Each new patient was allocated to the next sequence in the randomization list. Patients unable to complete the electronic questionnaire at home (eg, lack of Internet access) were switched from the electronic to the paper group. All patients were then recalled to complete the second questionnaire at home from the next day up to 1 week later. Satisfaction with the electronic questionnaire and general impressions regarding the user interface were assessed directly after completing the Web version of the PROQOL-HIV questionnaire. Each patient also provided their general impression and preference through a separate survey on a paper sheet at the end of the study period, and was asked to report any clinical events in the intervening period. This was returned by the prepaid envelope. They were also asked to rate the clarity of presentation, readability of questions and response options, and ease of use of the website. Participants randomized to the paper-based HRQL assessment in the second round mailed their completed questionnaires to the researcher. Cognitive debriefing was carried out on 10 patients who were administered the electronic version at the hospital.

Statistical Analysis

Statistical Tests

Biomedical and demographic data were summarized using classical descriptive statistical indicators of central location and dispersion. Group comparisons were performed using Wilcoxon-Mann-Whitney and Pearson chi-square tests for numerical and categorical variables, respectively, to assess heterogeneity between centers. Comparisons of scale scores for the paper and electronic versions were performed using nonparametric Wilcoxon test for paired samples. The effects of form (electronic vs paper) and order of presentation or sequence (electronic first or paper first) were tested on physical health and symptoms scores (9 items with total score expressed on a scale of 0-100 points) in a 2-way ANOVA with patients as random effects. The physical health and symptoms dimension was chosen because it has a high number of items and was shown to explain more of the total variance in factor analysis when validating this questionnaire [15]. However, similar analyses were also carried out on full-scale scores (39 items, 0-100 points). Alternate form and test-retest reliability were evaluated using Pearson and intraclass correlations for each dimension. A bootstrap procedure (B=500 replicates) was used to compute the 95% confidence interval for the ICC. This last measure of consistency, or temporal stability, was considered to determine the number of participants in a preliminary power analysis.

Usability testing used data from patients' survey reports (ease of completion, quality, and clarity of user interface, and motivation for on-going monitoring of HRQL). In addition, analysis of response time per item was carried out based on available individual timestamps. All statistical analyses were done using the R statistical software, version 2.15 (The R Project for Statistical Computing).

Power

From the validation study of the PROQOL-HIV questionnaire, which included a test-retest analysis on 34 French patients (average interval=52 days), the intraclass correlation was estimated at .859 (95% CI .710-.960) for full-scale scores [15]. Considering a theoretical reliability of 0.8, a sample size of 65 individuals is sufficient to verify if the ICC is greater than .75 with a 95% confidence interval [16].

Results

Participants

A sample of 80 outpatients (male: 52/79, 66%; female: 27/79, 34%) with a mean age of 47 years (SD 10.9) were enrolled in this study. Most participants were enrolled from the hospital (70/79, 88%). A flowchart demonstrating the randomization procedure is provided (Figure 2). Patients were randomized to 1 of 2 groups (paper version first or electronic version first) when they entered the study. However, because some patients reported having no Internet connection available at home, there were 10 switches in the order of administration (patients allocated to the group paper version first were given the electronic version first).

Overall, two-thirds of the participants were native French speakers (51/80, 64%). The main characteristics of the patients were stratified by center and summarized (Table 1). Participants from the institutional apartments were more likely to be living alone and to be without a professional activity. Their HIV-related immune decline was more advanced (56%, 5/9) stage C, average CD4 counts <500 cells/mm^3) compared with patients enrolled at the hospital. Overall, 89% (70/79) of patients were treated with an antiretroviral treatment (ART). Regarding viral hepatitis co-infection, 18% (14/79) of patients were seropositive for hepatitis C and 6% (5/79) for hepatitis B. Antidepressants were the most common concomitant treatment (14%,11/79 overall). None of the participants reported clinical events in the interval between the 2 study time points.

Table 1. Demographic and biomedical information on study participants.

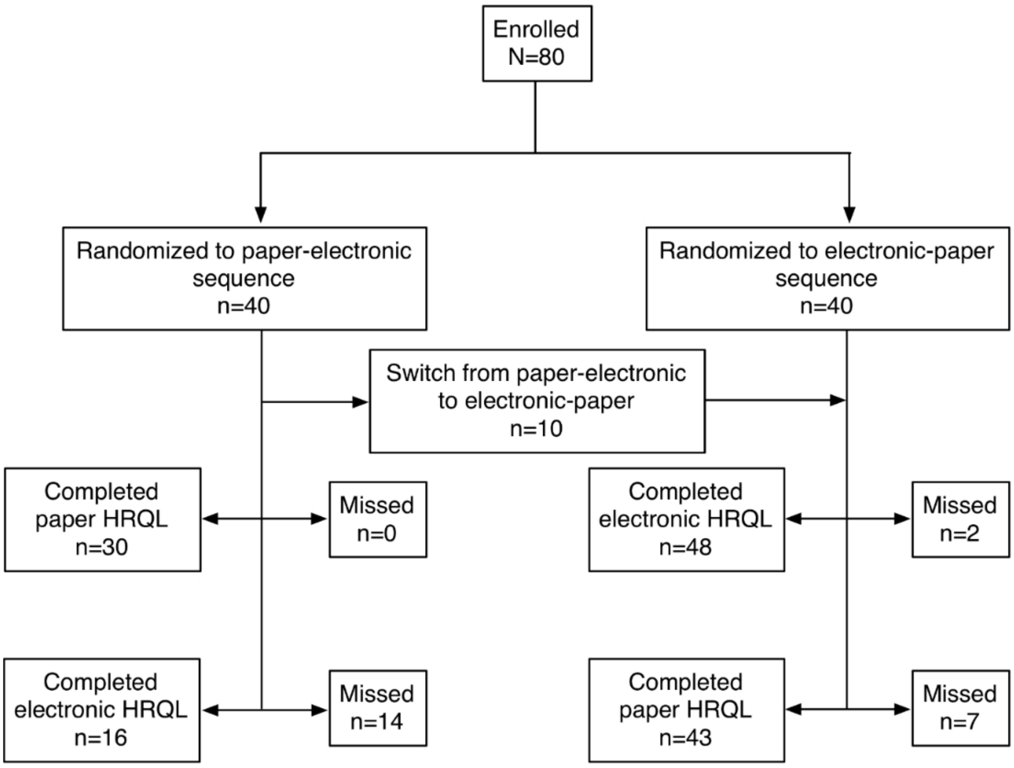
Variable	N	Hospital (n=70)	Institution (n=10)	All patients	*P* [b]
Age (years)	79				.58
Mean (SD)		46.9 (11.2)	44.4 (8.9)	46.7 (10.9)	
IQR		39.5-53.0	37.0-51.0	39.0-53.0	
Gender, n (%)	79				.12
Male		44 (63)	8 (89)	52 (66)	
Female		26 (37)	1 (11)	27 (34)	
Not currently working	78	15 (21)	6 (75)	21 (27)	.005
Education level (university), n (%)	78	29 (41)	4 (50)	33 (42)	.93
Marital status (single), n (%)	77	33 (48)	5 (62)	38 (49)	.68
Living alone, n (%)	78	31 (44)	8 (100)	39 (50)	.009
Comorbidities, n (%)					
Depression	78	9 (13)	3 (38)	12 (15)	.07
Psychiatric disorder	78	1 (1)	0 (0)	1 (1)	.73
Cardiovascular disease	78	9 (13)	3 (38)	12 (15)	.07
Diabetes	78	6 (9)	1 (12)	7 (9)	.71
Other comorbidities	78	5 (7)	0 (0)	5 (6)	.43
Lipodystrophy	79	15 (21)	2 (22)	17 (22)	.96
Current treatment, n (%)	78				
Prophylaxis[c]		3 (4)	3 (38)	6 (8)	<.001
Antidepressant		8 (11)	3 (38)	11 (14)	.045
Lipid-lowering		7 (10)	0 (0)	7 (9)	.35
Year of diagnosis	78				.19
Mean (SD)		1998 (8)	2002 (7)	1999 (8)	
IQR		1990-2005	2000-2008	1991-2006	
HAART status, n (%)	79	61 (87)	9 (90)	70 (89)	.99
CDC stage, n (%)	79				
A (asymptomatic)		41 (59)	2 (22)	43 (54)	.07
B (symptomatic conditions)		13 (19)	2 (22)	15 (19)	
C (AIDS-indicator conditions)		16 (23)	5 (56)	21 (27)	
Year of first HAART	66				.02
Mean (SD)		2002 (6)	2008 (3)	2003 (6)	
IQR		1996-2007	2007-2009	1997-2007	
Viral co-infection, n (%)					
Hepatitis C	79	14 (20)	0 (0)	14 (18)	.31
Hepatitis B	79	4 (6)	1 (10)	5 (6)	.99
CD4					
CD4 counts (cells/mm^3), mean (SD)	79	623 (438)	407 (191)	598 (422)	.04
IQR		441-700	213-527	424-694	
CD4 %, mean (SD)	74	30.3 (10.2)	20.2 (9.8)	29.6 (10.4)	—
IQR		24-37	19-24	23-37	
Viral load (undetectable)	21	1 (8)	8 (89)	9 (43)	.001

[a] HAART: highly active antiretroviral therapy; CD4: cluster of differentiation 4, T helper cells playing a major role in the human immune system; CDC: Centers for Disease Control and Prevention (classification system for HIV-infected adults and adolescents).

[b] Wilcoxon-Mann-Whitney test for 2 independent samples and Pearson chi-square.

[c] Toxoplasmosis, pneumocystis.

Figure 2. Participant flowchart.



Health-Related Quality of Life Data

A total of 64 of 80 electronic surveys were available at the end of the study. Of these, 50 participants completed all 43 items. In all, 9 participants (11%) were naive to ART; therefore, they were not required to complete the 10 treatment-related items. Three patients filled in only part of the questionnaire and were excluded from further analyses. Regarding the paper version, there were 73 records, including 49 full records, 10 ART-naive patients, and 14 partially filled records. A total of 59 electronic and paper questionnaires were available for test-retest analysis (Figure 2).There was a greater number of missing questionnaires for the paper version (n=14) compared to the electronic version (n=3, 2-sample test for equality of proportions with continuity correction, *P*=.02). This might be explained by the fact that people forgot to fill the questionnaire at home or there were problems with mailing the questionnaires. Regarding missing responses in the paper-based questionnaires, there were no more than 5% (2/43) of missing items for very few cases (n=3), and they were imputed by individual mean scores for each dimension.

The distribution of individual standardized scale scores on the physical health and symptoms dimension (0-100 points, where higher values reflect a better HRQL state) are illustrated as a 2-way scatter display (Figure 3). Summary statistics for all paper and electronic scale scores were computed for each dimension (Table 2). The lower level of agreement for the general health scale (Spearman ρ=.72, ICC=.714) may be because there are only 5 possible scores for this single item. In the present study, standard deviations for the physical health and symptoms scale (paper: SD 20.8; electronic: SD 20.3) were close to what was observed previously.

Table 2. Reliability analysis.

Domain (n of items)	Paper		Electronic		r (ρ)	ICC (95% CI)	P [a]
	Mean (SD)	IQR	Mean (SD)	IQR			
All items[b] (38)	68.3 (16.5)	54.1-82.1	69.5 (16.2)	56.7-84.4	.868 (.851)	.868 (.815-.931)	.21
Body concerns (4)	78.0 (26.1)	54.7-100.0	76.9 (26.0)	56.2-100.0	.827 (.777)	.827 (.708-.955)	.30
Emotional distress (4)	68.8 (26.9)	50.0-93.8	69.4 (26.0)	50.0-93.8	.842 (.874)	.842 (.772-.929)	.97
General health (1)	23.5 (20.7)	0.0-25.0	21.4 (21.7)	0.0-25.0	.715 (.790)	.714 (.550-.881)	.46
Health concerns (4)	51.6 (28.3)	25.0-81.2	55.6 (29.4)	31.2-81.2	.785 (.767)	.785 (.680-.884)	.46
Intimate relationships (3)	57.4 (33.8)	33.3-85.4	63.4 (31.1)	39.6-91.7	.782 (.782)	.779 (.625-.909)	.05
Physical health and symptoms (9)	76.3 (20.8)	63.9-94.4	75.6 (20.3)	63.9-91.7	.940 (.923)	.939 (.905-.979)	.63
Social relationships (2)	80.8 (27.1)	75.0-100.0	80.4 (28.9)	62.5-100.0	.824 (.803)	.822 (.725-.941)	.99
Stigma (2)	33.3 (33.8)	0.0-50.0	36.4 (34.5)	0.0-50.0	.712 (.741)	.712 (.543-.893)	.25
Treatment impact (10)	71.7 (21.2)	57.5-90.0	72.2 (20.9)	61.2-90.0	.708 (.783)	.708 (.444-.982)	.94

[a] Using Wilcoxon signed rank test.

[b] Full-scale score was calculated following the exclusion of 4 extra items and the general health item.

A random-effects 2-way ANOVA was used to assess the effects of the type of questionnaire (electronic or paper) and the order of administration. A total of 59 complete cases (74% of participants) were available for this analysis (Figure 2). It is worth noting that the order of administration was not balanced because 28% of participants ended up filling out the paper version first then the electronic version. No interaction between the type of questionnaire and administration order was found ($F_{1,55}$=0.098, P=.76). Likewise, there was no effect of the type of questionnaire ($F_{1,55}$=0.529, P=.47) or administration order ($F_{1,76}$=0.942, P=.34). These results indicate that scores obtained on either the electronic- or paper-based version were not statistically different, independent of the order of administration. Analysis based on full-scale scores yielded similar results (data not shown).

Reliability indexes (Table 2) were in the acceptable range, with Pearson correlations greater than .7 and intraclass correlation ranging from .708 (treatment impact) to .939 (physical state and symptoms). Mean scores for each dimension were not significantly different according to Wilcoxon signed rank tests for paired samples, even without considering correction for multiple testing (Bonferroni method). This suggests that, on average, this sample demonstrates comparable HRQL on all dimensions. The joint distribution of individual scores obtained from electronic and paper versions of the questionnaire for the physical and health symptoms dimension was analyzed separately (Figure 3). It is worth noting that although mean scores were slightly different between the paper (76.3 points) and electronic (75.6 points) versions, this difference was not significant (P=.63; Table 2) and would not be considered as clinically relevant anyway. A Bland-Altman chart is provided in Figure 4 with limits of agreement computed as ±1.96×SD, where SD is the standard deviation for the difference between individual scores of the two versions. In both cases, it can be seen that scores are generally close to one another, with the exception of 3 participants who had higher physical health and symptoms scores on the electronic version compared to the paper version.

Figure 3. Individual scores (N=59) on electronic and paper versions of the PROQOL-HIV questionnaire for the physical health and symptoms (PHS) dimension. Individual points are displayed with alpha transparency so that darker symbols indicate a higher number of identical pairs of scores. The straight line represents the ordinary least squares regression line.

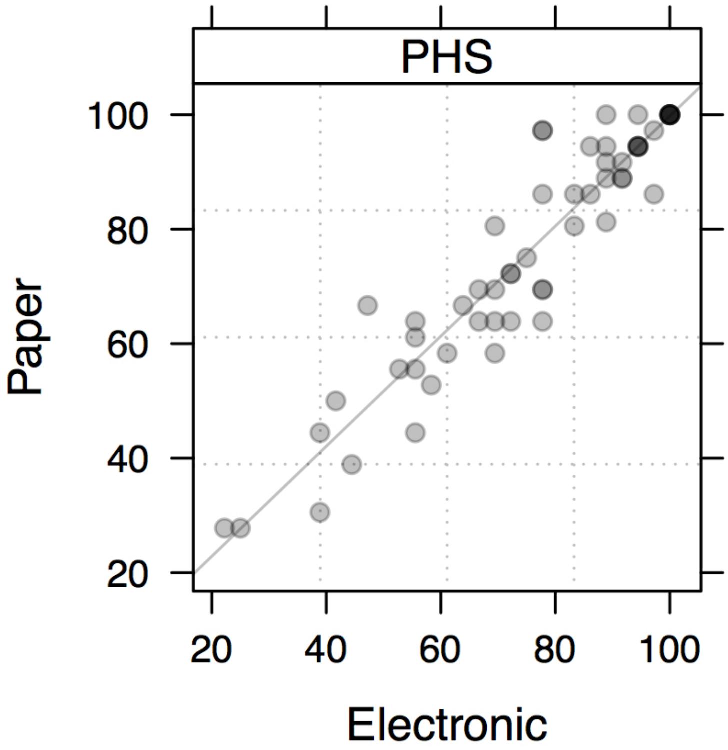
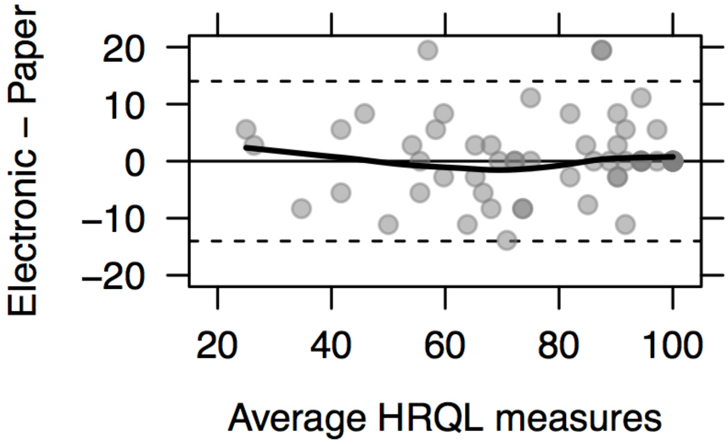


Figure 4. Bland-Altman plot showing the difference between scores of the electronic and paper versions as a function of average physical health and symptoms scores. The upper and lower dashed horizontal lines represent limits of agreement. A Lowess smoother (solid thicker black line) has been added to differentiate local irregularities on the distribution of scores.



Usability Analysis

Results on the usability and satisfaction survey were analyzed separately (Table 3). Of the 79% (63/80) of surveys that were completed, two-thirds of participants (65%, 36/60) reported that they would be interested in electronic assessment of their HRQL by using an electronic version of PROQOL-HIV in routine clinical care. Overall, 55% of the participants indicated a preference toward the electronic questionnaire compared to the paper-based questionnaire. Only 9% (5/55) of participants preferred the paper version, and 36% (20/55) were indifferent. Regarding the version preferred by patients, there were no significant associations with age (*P*=.12), gender (*P*=.60), marital status (*P*=.39), living mode (alone, *P*=.48), education level (*P*=.18), professional activity (*P*=.59), and diagnosis date (*P*=.39). Individual ratings of the PROQOL-HIV user interface (80% to 100% of positive responses) confirmed the clarity and usability of the Web version, although the visual display of summary scores might need further improvement as confirmed by 10 cognitive debriefings that were carried out with patients.

Table 3. Participants' survey responses.

Question	N	n (%)	Mean (SD)	IQR
Display on screen is comprehensible and easy to follow	63	63 (100)		
Font size looks appropriate	62	61 (98)		
Single page design is satisfactory	62	62 (100)		
Visualization of results is an interesting option	61	58 (95)		
Display of results is comprehensible	60	48 (80)		
Ease of input responses	63		4.6 (0.7)	4-5
Questions readability	63		4.6 (0.8)	4-5
Scores readability	61		4.0 (1.0)	4-5
Interested in longitudinal follow-up of personal scores	60		4.4 (0.9)	4-5
Difficulty with computing material	60	8 (13)		
Ease of filling (electronic vs paper)	55		4.0 (1.0)	3-5
Preference	55			
Indifferent		20 (36)		
Paper		5 (9)		
Electronic		30 (55)		

Analysis of Response Time

The median time to complete the electronic version of PROQOL-HIV was 7 minutes, with 50% of the times between 4.8 and 11.1 minutes. In 2 cases, participants took more than 20 minutes to complete the electronic questionnaire; in 16 cases, participants took less than 5 minutes. The former was explained by disconnection from the hospital network. In the latter case, respondents had completed the paper questionnaire previously and were familiar with the items. Mean individual response time (trimmed to 5%) per item was 9.1 seconds on average (range 3.7-20.1). Only 17 of 54 (31%) participants completed the questionnaire in the order presented. For those participants who provided answers in a different order, it should be noted that filling order was usually altered for one question, but not for more than 6 questions in total. This included participants who delayed completing the questionnaire for a short period of time or those who forgot to answer some of the questions before submitting the Web form. In the latter case, missing answers were highlighted by the system and participants had to complete them again.

Discussion

Principal Results

According to international guidelines, the validation of an electronic version of an existing questionnaire requires the demonstration of equivalent measurement properties between the original paper-based and the electronic versions of the questionnaire. This is achieved through statistical measures of correlation and mean differences between the 2 series of individual scores which can be summarized by the intraclass correlation. However, several threats to ensuring equivalent measures have been reported. In particular, substantial changes to the presentation of items or questions to accommodate screen limitations and poor experience with computer use can alter validity or reliability of scores [7].

Typically, cognitive debriefing is carried out with a small number of participants to verify that the content of the electronic questionnaire is perceived in the same way as that of the paper-based version; test-retest studies are restricted to cases in which significant changes were introduced in the electronic version (eg, questionnaire layout, response options). This study goes beyond standard recommendations (cognitive debriefing) and it provides a quantitative assessment of both test-retest reliability and users' self-perception of the Web version of the PROQOL-HIV questionnaire.

The demographics of the participants in this study are representative of the population of people living with HIV in this French metropolitan area. In 2011, this population was composed of 67% men and 33% women, and 24% of the patients were non-European Union residents. Among them, 17% originated from sub-Saharan Africa. In the Vespa 2 study, 41% of the participants were not currently working and only 37% considered their financial situation as good [17]. When introducing new technologies to health care, it is important that all populations, including immigrants and people with low income, can benefit from such programs. A heterogeneous sample of patients participated in this study, providing they could read French and they could use the electronic questionnaire even if they did not have Internet at home, because it was available at the hospital directly.

The present findings suggest that assessing health-related quality of life specific of HIV disease on a Web-based platform is easy and reliable, and that the electronic version of the PROQOL-HIV questionnaire fulfills the criteria for migrating a paper-based questionnaire to a computer-based mode of administration. Following their analysis of 46 single studies (of 65 eligible case reports) relying mainly on computer-based assessment (n=31, 67%), Gwaltney and collaborators [7] reported an average correlation between paper-based and electronic assessment of .90 (95% CI .87-.92, n=32) without significant differences from studies relying on intraclass correlation or weighted kappa. Our results suggest that the PROQOL-HIV questionnaire can achieve good test-retest reliability, as measured by an intraclass correlation greater than .8 for the principal dimension of the questionnaire or the full-scale score. Moreover, it demonstrated good face validity according to respondents' self-perception collected at the end of the study, with more than 80% with a positive opinion toward PROQOL-HIV usability and clarity when assessing HRQL specific to HIV disease. Interestingly, only 9% of the patients indicated that they preferred the paper version at the end of the study. However, this might be a biased indicator because the objective of the study was to validate the electronic version. One of the important findings was the high interest in longitudinal follow-up of personal HRQL scores, suggesting that electronic assessment may be of value in routine clinical care for HIV. The electronic version of PROQOL-HIV has been in use by some French patient associations with positive feedback from the local coordinators. Further studies will determine the value of electronic longitudinal follow-up of self-reported HRQL.

The use of electronic PRO measures in HIV care can offer important implications. First, physicians could benefit from a direct and contemporaneous assessment of a patient's HRQL at the time of the consultation, which will enhance clinical observation and decisions around treatments. Second, there is a growing interest in patient-centered care, which has been shown to improve perceived health outcomes because patients feel engaged in their health care management. Consequently, the opportunity for patients to record their HRQL at home and to have access to their results, together with their physicians, should enhance the relationship between care providers and the patients.

Limitations

The analysis of open-ended satisfaction questions highlighted critical issues with using PROQOL-HIV as an electronic HRQL questionnaire. Because the questionnaire was given on a single HTML page, response headings for some of the items were not always visible depending on screen size and length of that section. Occasional breakdown of the Internet connection was reported in one center; hence, patients' responses were not taken into account and they had to fill in the questionnaire again. Display of HRQL summary scores should be complemented by a brief overview of the patients' health state. Another limitation of the study was the imbalance in the order of administration, although this did not affect the validity of the present findings because there was enough data to analyze the temporal stability of HRQL scores and to compute reliable indicators of participants' impressions regarding the Web version of PROQOL-HIV. Finally, no systematic pattern was detected for patients who did not follow the order of the questions, suggesting that this does not affect the structure of the PROQOL-HIV questionnaire. It should be noted that patients could also fill in the PRO paper-based questionnaire in random order, but this could obviously not be detected. The absence of follow-up data to study the responsiveness of the electronic PROQOL-HIV and to collect information on how to best display summary scores for easy monitoring of personal data will be assessed in a forthcoming study.

Conclusions

The PROQOL-HIV instrument has been adapted in a way that would be classified as minor according to ISPOR ePRO Task Force guidelines [12]. The new electronic version shows good reliability and face validity, and scores obtained from paper or electronic modes share comparable accuracy and interpretation. An interesting finding was that few patients (9%), including patients having no Internet at home, preferred the paper-based version of PROQOL-HIV. The desire of participants to have access to the scores of this instrument as a way of tracking themselves over time also shows their interest in Web-based assessments in clinical health care.

Acknowledgments

This study was funded by Sidaction. At Bicetre Hospital, we wish to thank Katia Bourdic, Martine Mole, Marie-Josee Dulucq, Sandrine Pottez, Patrizia Assal, Dr Yann Quertainmont, Dr Imane Amri, Dr Erwan Fourn, Dr Delphine Peretti, Dr Olivier Segeral, Dr Arnaud Blanc, Dr Elina Teicher, Pr Olivier Lambotte, Dr Nicolas Noel, and the nurses of the ambulatory center for HIV disease. We wish to thank the staff at *Fondation Maison des champs*, *Service ACT du Val de Marne*, and Association Tipi.

Conflicts of Interest

None declared.

References

1. Reiser SJ. The clinical record in medicine. Part 1: Learning from cases. Ann Intern Med 1991 May 15;114(10):902-907. [Medline: 2014953]
2. Tang PC, LaRosa MP, Gorden SM. Use of computer-based records, completeness of documentation, and appropriateness of documented clinical decisions. J Am Med Inform Assoc 1999 Jun;6(3):245-251 [FREE Full text] [Medline: 10332657]
3. US Department of Health and Human Services, Food and Drug Administration. Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Silver Spring, MD: Center for Drug

Evaluation and Research, Food and Drug Administration; 2009. URL: http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf [accessed 2014-02-12] [WebCite Cache ID 6NKQY89wp]

4. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient non-compliance with paper diaries. BMJ 2002 May 18;324(7347):1193-1194 [FREE Full text] [Medline: 12016186]
5. Deshpande PR, Rajan S, Sudeepthi BL, Abdul Nazir CP. Patient-reported outcomes: A new era in clinical research. Perspect Clin Res 2011 Oct;2(4):137-144 [FREE Full text] [doi: 10.4103/2229-3485.86879] [Medline: 22145124]
6. Bernhard J, Cella DF, Coates AS, Fallowfield L, Ganz PA, Moinpour CM, et al. Missing quality of life data in cancer clinical trials: serious problems and challenges. Stat Med 1998;17(5-7):517-532. [Medline: 9549801]
7. 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 Apr;11(2):322-333. [doi: 10.1111/j.1524-4733.2007.00231.x] [Medline: 18380645]
8. McKenzie S, Paty J, Grogan D, Rosano M, Curry L, Sciarappa K, et al. Applied Clinical Trials. 2004. Proving the eDiary dividend URL: http://www.appliedclinicaltrialsonline.com/appliedclinicaltrials/article/articleDetail.jsp?id=98374 [accessed 2014-02-12] [WebCite Cache ID 6NKSDLr46]
9. Ganser AL, Raymond SA, Pearson JD. In: Byrom B, Tiplady B, editors. Epro: Electronic Solutions for Patient-Reported Data. Oxford: Gower Pub Co; 2010.
10. 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]
11. Miller CW, Himelhoch S. Acceptability of Mobile Phone Technology for Medication Adherence Interventions among HIV-Positive Patients at an Urban Clinic. AIDS Res Treat 2013;2013:670525 [FREE Full text] [doi: 10.1155/2013/670525] [Medline: 23997948]
12. Swendeman D, Rotheram-Borus MJ. Innovation in sexually transmitted disease and HIV prevention: internet and mobile phone delivery vehicles for global diffusion. Curr Opin Psychiatry 2010 Mar;23(2):139-144 [FREE Full text] [doi: 10.1097/YCO.0b013e328336656a] [Medline: 20087189]
13. Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, ISPOR ePRO Task Force. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value Health 2009 Jun;12(4):419-429. [doi: 10.1111/j.1524-4733.2008.00470.x] [Medline: 19900250]
14. Duracinsky M, Herrmann S, Berzins B, Armstrong AR, Kohli R, Le Coeur S, et al. The development of PROQOL-HIV: an international instrument to assess the health-related quality of life of persons living with HIV/AIDS. J Acquir Immune Defic Syndr 2012 Apr 15;59(5):498-505. [doi: 10.1097/QAI.0b013e318245cafe] [Medline: 22205438]
15. Duracinsky M, Lalanne C, Le Coeur S, Herrmann S, Berzins B, Armstrong AR, et al. Psychometric validation of the PROQOL-HIV questionnaire, a new health-related quality of life instrument-specific to HIV disease. J Acquir Immune Defic Syndr 2012 Apr 15;59(5):506-515. [doi: 10.1097/QAI.0b013e31824be3f2] [Medline: 22293550]
16. Streiner D, Norman G. Health Measurement Scales: A Practical Guide to Their Development and Use. Oxford: Oxford University Press; 2008.
17. Dray-Spira R, Wilson d'Almeida K, Aubriere C. Living With HIV: First Results of ANRS-VESPA2 Study. 2013. Health status of people living with HIV followed at hospital in metropolitan France in 2011 and characteristics of those recently diagnosed URL: http://www.invs.sante.fr/beh/2013/27/2013_26-27_1.html [accessed 2014-02-12] [WebCite Cache ID 6NKQKpdtz]

Abbreviations

ART: antiretroviral therapy
HAART: highly active antiretroviral therapy
HIV: human immunodeficiency virus
HRQL: health-related quality of life
HTML: Hypertext Markup Language
ICC: intraclass correlation coefficient
ISPOR: International Society for Pharmacoeconomics and Outcomes Research
PRO: patient-reported outcomes
PROQOL-HIV: Patient-Reported Outcomes Quality of Life-HIV questionnaire

Edited by G Eysenbach; submitted 16.02.14; peer-reviewed by S Raymond, R Harding; comments to author 10.03.14; revised version received 24.03.14; accepted 24.03.14; published 25.04.14.

Please cite as:
Duracinsky M, Lalanne C, Goujard C, Herrmann S, Cheung-Lung C, Brosseau JP, Schwartz Y, Chassany O
Electronic Versus Paper-Based Assessment of Health-Related Quality of Life Specific to HIV Disease: Reliability Study of the PROQOL-HIV Questionnaire
J Med Internet Res 2014;16(4):e115
URL: http://www.jmir.org/2014/4/e115/
doi:10.2196/jmir.3330
PMID:24769643

Original Paper

User-Generated Online Health Content: A Survey of Internet Users in the United Kingdom

Braden O'Neill[1], BSc; Sue Ziebland[1], MSc; Jose Valderas[1,2], MPH, MD, PhD; Francisco Lupiáñez-Villanueva[3], PhD

[1]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

[2]Medical School, University of Exeter, Exeter, United Kingdom

[3]Department of Information and Communication Science, Universitat Oberta de Catalunya, Barcelona, Spain

Corresponding Author:
Braden O'Neill, BSc
Nuffield Department of Primary Care Health Sciences
University of Oxford
Radcliffe Observatory Quarter
Woodstock Road
Oxford, OX2 6GG
United Kingdom
Phone: 44 (0)1865289414
Fax: 44 (0)1865289287
Email: braden.oneill@phc.ox.ac.uk

Abstract

Background: The production of health information has begun to shift from commercial organizations to health care users themselves. People increasingly go online to share their own health and illness experiences and to access information others have posted, but this behavior has not been investigated at a population level in the United Kingdom.

Objective: This study aims to explore access and production of user-generated health content among UK Internet users and to investigate relationships between frequency of use and other variables.

Methods: We undertook an online survey of 1000 UK Internet users. Descriptive and multivariate statistical analyses were used to interpret the data.

Results: Nearly one-quarter of respondents (23.7%, 237/1000) reported accessing and sharing user-generated health content online, whereas more than 20% (22.2%, 222/1000) were unaware that it was possible to do this. Respondents could be divided into 3 groups based on frequency of use: rare users (78.7%, 612/778) who accessed and shared content less than weekly, users (13.9%, 108/778) who did so weekly, and superusers (7.5%, 58/778) who did so on a daily basis. Superusers were more likely to be male ($P<.001$) and to be employed ($P<.001$), but there were no differences between the groups with respect to educational level ($P=.99$) or health status ($P=.63$). They were more likely to use the Internet for varied purposes such as banking and shopping ($P<.001$).

Conclusions: Although this study found reasonably widespread access of user-generated online health content, only a minority of respondents reported doing so frequently. As this type of content proliferates, superusers are likely to shape the health information that others access. Further research should assess the effect of user-generated online content on health outcomes and use of health services by Internet users.

(J Med Internet Res 2014;16(4):e118) doi:10.2196/jmir.3187

KEYWORDS

eHealth; consumer health information; patient education; health education; health promotion; social media; superusers

Introduction

Background

There can be little doubt that the Internet has changed the way that people experience health and illness. People routinely use the Web to learn about the meaning of symptoms, tests, diagnoses and treatments, and to find out how others have rated doctors and hospitals [1]. In the United Kingdom, more than two-thirds of residents use the Internet to obtain health information [2], and the UK Department of Health considers availability of high-quality information a key policy objective [3]. The Internet once provided information that could be accessed and viewed, but not easily modified. Patients increasingly go online to share information and to access information others have posted on the Internet, including their own experiences. Although this is a widespread practice, especially for long-term conditions [4], relatively little is known about the characteristics of people who share and access user-generated content (UGC) online. This is in contrast to general online health information-seeking behavior which has been extensively studied [1,3,4]. In this paper, we draw on a major European survey to explore who shares and accesses user-generated health content online in the United Kingdom.

User-Generated Health Content

User-generated health content is presented in many different ways across different types of websites [5]. Traditional, commercially produced online health content (developed by businesses or health systems or governments) has predominantly consisted of facts and figures, such as numerical data related to the diagnosis or prognosis of a condition. User-generated health information, by virtue of its distributed, often informal development and dissemination, is a broad concept and incorporates anything that a patient or carer posts online for the purpose of others seeing it. Entwistle et al [6] noted that online health information can be broadly divided into 2 categories: general facts and personal experiences. Although these categories are not mutually exclusive, the differences have important implications. General facts consist of information about particular conditions, treatments, or services, and are intended to be broadly applicable to many people interested in the same topic. Personal experiences may also provide information about conditions, treatments, or services, but are often in narrative form and portray an individual's experience of undergoing a particular treatment or of living with a long-term condition. People tend to draw on both general facts and personal experiences, depending on the issue involved [6,7]. It has been noted that adequate information is an important aspect of health care delivery that matters to patients [8].

General health information sites often provide UGC in the form of curated experiences selected by the site's developers and the ability for site users to post their own comments. There are sites exclusively devoted to interview-based research on health experiences; for example, the UK's Healthtalkonline has been available for more than a decade and now has many counterparts in other countries [9]. Blogs enable people to post their stories over time and provide the facility to embed video and links to other content, including peer-reviewed journal articles where applicable. Social networking sites enable patients to post content (both facts and experiences) and have pages devoted to specific conditions and treatments. Online forums host pages for specific conditions and facilitate patients coming together who have similar diseases or have undergone similar treatments. Some sites include crowdsourcing: "an approach to accomplishing a task by opening up its completion to broad sections of the public" [10]. For example, the website PatientsLikeMe aggregates data provided by site users. Patients can compare their data with aggregated input from other site users, and data obtained through crowdsourcing can be used for research [11]. Finally, there are those sites that enable a health service user to write and read reviews of specific hospitals, doctors, or health systems [12]. Increasingly, websites include several different ways of incorporating content from patients and the public. What all these disparate sources share is that their content depends on people posting, sharing, and comparing health information, whether about themselves or those for whom they care.

Use and Nonuse

One of the predominant places where UGC is accessed and shared online is in the context of virtual support groups for particular conditions as well as individual treatments. The heterogeneity of online support group membership has been reported in several publications; these groups consist of some people who post occasionally and some who are frequent prolific posters. Those who post have been described as key players [13], active users [14], and caretakers [15]. Most people who visit support groups do not actively post. In a study of people accessing an online smoking cessation group, 84.7% of those who registered and accessed the resource at least once never posted [16]. To explain this distribution, the "1% rule" has been adapted from marketing literature to the use of health social networks, suggesting that 90% of actors observe but do not participate, 9% rarely contribute, and 1% create most of the content [17]. This is contextual, and people's behavior may vary between sites if they visit several different sites. Adams [18] has suggested that people may be more motivated to add their own commentary if they feel that their experiences are not covered by others accounts (eg, adverse drug effects). Because the content on these websites is generated by a small proportion of the people who use it, the concept of the "superuser" [17,19] of user-generated online health information—the minority of people who post information online—has received recent attention. This has not been characterized within the United Kingdom.

The Aim of Our Research

The purpose of this paper is twofold. First, we identify characteristics of UK Internet users who access and share UGC online. In this study, we use UGC to denote 4 specific behaviors about which questions were asked in the survey: (1) participating in an online support group for people who are concerned about the same health or medical issue, (2) participating in social networking sites talking about health and wellness, (3) describing a medical condition or problem online to get advice from other online users, and (4) disclosing medical information on social networking sites.

Second, we investigate the frequency with which respondents engage in the aforementioned study behaviors, paying particular attention to those who do so most frequently because these users predominantly generate the content that others view. By grouping respondents according to how frequently they go online to access and share UGC, we aim to characterize the differences between these groups. We aim to determine areas for further research to support the effective and beneficial use of online health information.

Methods

Survey Instrument and Ethics

We analyzed United Kingdom data from the Citizens and Information Communication Technology for Health survey, a project undertaken in 2011 by the Institute for Prospective Technology Studies of the European Commission's Joint Research Centre. This online survey was developed from a theoretical framework of the social determinants of information and communication technology (ICT) for health, translated into native languages in 14 European Union (EU) member countries. The survey was developed to understand and characterize European citizens' use of ICT for health. It included questions on a range of triggers, motivations, and behaviors. This analysis examines a subset of questions related to user-generated online health content. The full questionnaire is available in Multimedia Appendix 1 [20].

Technical, methodological, and legal considerations were carefully addressed in the context of designing and implementing the survey. These considerations ensured anonymity and confidentiality of individual responses [21-23]. The survey was conducted in accordance with European Society for Opinion and Marketing Research ethical guidelines [24]. At the time the survey was carried out, one of the authors (FLV) was employed by the European Commission.

Sample and Data Collection

The sample consisted of 1000 respondents from the United Kingdom who completed an online survey in 2011 covering a variety of domains regarding the use of ICT for health. We randomly sampled people aged 16 to 74 years who had used the Internet in the 3 months before the survey to obtain the sample.

An online invitation to participate in the survey was sent to 7291 individuals. The data were collected between July 20 and August 1, 2011. Simple random sampling was carried out by the software provider who administered the survey (Cint), then quota sampling was used to accurately reflect the demographic composition of UK residents [25]. We obtained 2410 responses (4881 nonresponses) of which 1320 were out of quota and 90 were excluded because of missing data. The average length of time to complete the questionnaire was 20.5 minutes.

More than half (57.00%, 570/1000) of respondents were employed and nearly all (90.00%, 900/1000) had completed upper secondary or tertiary education (Table 1). This is broadly comparable to the characteristics of the UK population, although it should be noted that this sample is more highly educated than the general UK population [25].

Table 1. Demographic composition of sample.

Age group (years)	Female, n (%)	Male, n (%)	Total, n (%)
16-24	103 (10.3)	106 (10.6)	209 (20.9)
25-54	297 (29.7)	295 (29.5)	592 (59.2)
55-74	98 (9.8)	101 (10.1)	199 (19.9)
Total	498 (49.8)	502 (50.2)	1000 (100.0)

Statistical Analysis

Initially, we calculated descriptive statistics about the characteristics of the study participants. After this, to determine the extent to which the 4 behaviors studied represent an underlying composite variable, we employed principal components analysis (PCA). In carrying out PCA, we a priori defined components explaining adequate amounts of variability in the data as those that have an eigenvalue greater than 1.00 [26]. Because the results of the PCA suggested that all 4 behaviors could be explained by one underlying component, for further analysis we treated this as a single composite variable, which we called "accessing and sharing user-generated health content online."

We then used nonhierarchical *k*-means cluster analysis to characterize respondents based on their reported frequency of accessing and sharing UGC online. We determined differences between the frequency groups by carrying out ANOVA and chi-square tests for categorical and continuous variables, respectively. Data were examined to determine if they met assumptions for the univariate and multivariate statistical tests carried out, and all assumptions were met. Statistical analyses were performed using SPSS version 20.0 (IBM Corp, Armonk, NY, USA).

Results

Frequency of Engaging in Individual Study Behaviors

We found that approximately one-quarter of all respondents participated in online support groups (27.80%, 278/1000) and health-related social networking sites (23.80%, 238/1000) (Table 2). A similar proportion reported describing a medical condition or problem online to get help from other online users (23.10%, 231/1000) and disclosing medical information on social networking sites (16.70%, 167/1000).

Less than 10% of respondents (4.80%, 48/1000 to 7.80%, 78/1000) reported engaging in any of the study behaviors at least or more than once a week. Approximately 20% of respondents (22.20%, 222/1000) reported being unaware that it was possible to engage in these behaviors online.

Table 2. User-generated content behaviors.

Regarding health, wellness, and the Internet, how often have you...	Frequency, n (%)					
	Never	Less than once a month	At least once a month (but not every week)	At least once a week (but not every day)	Every day or almost every day	Not aware of it
Participated in an online support group for people who are concerned about the same health or medical issue	588 (58.8)	145 (14.5)	66 (6.6)	46 (4.6)	21 (2.1)	134 (13.4)
Participated in social networking sites talking about health and wellness	629 (62.9)	106 (10.6)	54 (5.4)	53 (5.3)	25 (2.5)	133 (13.3)
Described a medical condition or problem online in order to get advice from other online users	628 (62.8)	119 (11.9)	64 (6.4)	25 (2.5)	23 (2.3)	141 (14.1)
Disclosed medical information on social networking sites	660 (66.0)	72 (7.2)	40 (4.0)	33 (3.3)	22 (2.2)	173 (17.3)

Characterizing Predictors of Accessing and Sharing User-Generated Content

Applying cluster analysis to the “accessing and sharing user-generated health content online” composite variable resulted in a 3-cluster solution. After excluding those who did not know that these resources were available (excluded because they have different reasons for nonuse than those who are aware of the possibility and choose to not engage) [26], we were left with 778 responses for further analysis.

Most respondents (78.7%, 612/778) were rare users who reported that they never or very rarely accessed and shared user-generated health content. Users who went online and engaged in these behaviors infrequently (ie, weekly or monthly but not daily) accounted for more than 10% (13.9%, 108/778) of respondents. A further minority (7.5%, 58/778) were clustered as superusers who accessed and/or generated UGC daily.

Sociodemographic Characteristics

Comparing the 3 groups, superusers were more likely to be male than users or rare users (Table 3). Those in the 16-24 years age group were more likely to be superusers than users or rare users, whereas those in the 55-74 years age group were generally rare users, as opposed to superusers or users. Superusers were more likely to be employed (57.8%, 33/58) than unemployed (9.3%, 5/58), students (19.0%, 11/58), or retired/inactive in the labor force (24.0%, 13/58). The groups did not differ on level of education.

Regarding household composition, respondents with children (1-parent, 2-parent, or extended family households) were less likely to be superusers than they were to be users or rare users. There were no differences between the groups based on the number of children younger than 16 years in the household. More rare users and superusers reported that they lived in intermediate areas, as opposed to densely or thinly populated areas, whereas users were more likely to report living in densely populated areas.

Table 3. Sociodemographic characteristics of respondents (N=778).

Sociodemographic characteristics	Rare users, n (%) n=612	Users, n (%) n=108	Superusers, n (%) n=58	*P*
Gender				.02
Male	301 (49.3)	61 (56.5)	38 (67.2)	
Female	311 (50.7)	47 (43.5)	20 (32.8)	
Age (years)				<.001
16-24	89 (14.7)	39 (36.1)	25 (43.1)	
25-54	370 (60.6)	63 (58.3)	32 (55.2)	
55-74	151 (24.7)	6 (5.6)	1 (1.7)	
Level of education completed				.99
Primary or lower secondary education (ISCED 0-2)[a]	70 (11.4)	14 (13)	7 (12.1)	
Upper secondary education (ISCED 3 or 4)[a]	261 (42.6)	45 (41.7)	26 (44.8)	
Tertiary education (ISCED 5 or 6)[a]	281 (45.9)	49 (45.4)	25 (43.1)	
Occupation				<.001
Employed or self-employed (including family workers)	354 (57.8)	72 (66.7)	31 (53.4)	
Unemployed	57 (9.3)	6 (5.6)	11 (19.0)	
Student (not in the labor force)	54 (8.8)	21 (19.4)	11 (19.0)	
Other not in the labor force (eg, retired, inactive)	147 (24.0)	9 (8.3)	5 (8.6)	
Type of locality				<.001
Densely populated area (cities and large towns)	186 (30.4)	52 (48.1)	18 (31.0)	
Intermediate area (towns)	308 (50.3)	43 (39.8)	36 (62.1)	
Thinly populated area (village and rural)	118 (19.3)	13 (12)	4 (6.9)	
Number of members in the household				.13
1	81(13.2)	11 (10.2)	7 (12.1)	
2	224 (36.6)	36 (33.3)	12 (20.7)	
3	123 (20.1)	22 (20.4)	19 (32.8)	
≥4	184 (30.1)	39 (36.1)	20 (34.5)	
Number of children under 16 years				.08
None	436 (71.2)	67 (62)	32 (55.2)	
1	92 (14.9)	23 (21.3)	12 (22.4)	
2	74 (12.1)	14 (13.0)	10 (17.2)	
>2	11 (1.8)	4 (3.7)	3 (5.2)	
Household composition				.02
Single person ≥65 years	14 (2.3)	1 (0.9)	0 (0)	
Single person <65 years	67 (10.9)	10 (9.3)	8 (13.8)	
2 persons, ≥1 aged ≥65 years	69 (11.3)	6 (5.6)	2 (3.4)	
2 persons, both <65 years	148 (24.2)	26 (24.1)	8 (13.8)	
Single person with child(ren) <16 years	7 (1.1)	5 (4.6)	1 (1.7)	
2 persons with child(ren) <16 years	161 (26.3)	34 (31.5)	20 (34.5)	
2 persons ≥65 years	27 (4.4)	2 (1.9)	3 (5.2)	
Extended family (<16 + 16-64 + <65 years)	6 (1.0)	3 (2.8)	3 (5.2)	
≥3 adults <65	113 (18.5)	21 (19.4)	13 (22.4)	

[a]ISCED: International Standard Classification of Education [27].

Health Characteristics

We did not identify any differences between groups for health status, undergoing long-term medical treatment, or number of visits to the doctor in the past 12 months (Table 4). Among superusers who reported having someone close to them currently experiencing long-term illness or disability, 64.3% (37/58) reported taking care of them.

Table 4. Health characteristics of respondents (N=778).

Health characteristics	Rare users, n (%) n=612	Users, n (%) n=108	Superusers, n (%) n=58	*P*
Health status reported				.63
Very bad	3 (0.5)	1 (0.9)	1 (1.7)	
Bad	45 (7.4)	7 (6.5)	6 (10.3)	
Neither good or bad	115 (18.8)	24 (22.2)	6 (10.3)	
Good	297 (48.5)	53 (49.1)	28 (48.3)	
Very good	152 (24.8)	23 (21.3)	17 (29.3)	
Long-standing illness or health problem reported				.57
Yes	263 (42.9)	41 (37.5)	25 (43.9)	
No	349 (57.1)	68 (62.5)	33 (56.1)	
Undergoing a long-term medical treatment				.74
Yes	215 (35.1)	35 (32.1)	18 (31.6)	
No	397 (64.9)	73 (67.9)	40 (68.4)	
Visit a doctor during the past 12 months				.29
Yes	512 (83.7)	95 (88.0)	52 (89.7)	
No	100 (16.3)	13 (12.0)	6 (10.3)	
How many times did you visit a doctor during the last 12 months?				.64
None	100 (16.3)	13 (12)	6 (10.3)	
1-2 visits	216 (35.3)	31 (28.7)	21 (36.2)	
3-4 visits	118 (19.3)	27 (25.0)	11 (19.0))	
5-6 visits	70 (11.4)	15 (13.9)	8 (13.8)	
>6 visits	108 (17.6)	22 (20.4)	12 (20.7)	
Is someone close to you currently experiencing long-term illness or disability?				.01
Yes	227 (37.1)	56 (51.5)	29 (50.0)	
No	385 (62.9)	52 (48.5)	29 (50.0)	
Are you taking care of such a person?				.02
Yes	234 (38.2)	39 (35.8)	37 (64.3)	
No	378 (61.8)	69 (64.2)	21 (35.7)	

Online Health Information and Health Behavior

Superusers were more likely than users or rare users to report that the information they accessed affected the way they care for themselves and the way they eat or exercise (Table 5). They were also more likely to report that after accessing health information on the Internet, they subsequently talked to a doctor or nurse about what they found. Finally, superusers were more likely to report accessing information for their children than users or rare users.

Table 5. Online health information and health behavior of respondents (N=778).

Online health information and health behavior	Rare users, n (%) n=612	Users, n (%) n=108	Superusers, n (%) n=58	*P*
Looking for health and/or wellness information for...				
Yourself	531 (86.7)	102 (94.4)	56 (96.6)	.01
Your child	152 (24.9)	32 (29.6)	26 (44.8)	.004
Parent	158 (25.8)	48 (44.4)	27 (46.6)	<.001
Another relative	194 (31.7)	41 (38.0)	24 (41.4)	.19
Some else	155 (25.3)	39 (36.1)	24 (41.4)	.005
Did you later talk to a doctor or nurse about the information you got online?				<.001
Yes	234 (38.3)	65 (59.8)	43 (74.1)	
No	378 (61.7)	43 (40.2)	15 (25.9)	
Overall, how useful was the health information you got online?				.02
Not at all useful	15 (2.5)	2 (1.9)	0 (0)	
Not too useful	55 (9.0)	5 (4.6)	1 (1.7)	
Somewhat useful	400 (65.3)	71 (65.7)	33 (56.9)	
Very useful	142 (23.2)	30 (27.8)	24 (41.4)	
Did the information you got online affect any of your decisions about health treatments or the way you take care of yourself?				<.001
Yes	201 (32.8)	56 (51.9)	38 (65.5)	
No	411 (67.2)	52 (48.1)	20 (34.5)	
Did the information you got online affect the way you eat or exercise?				<.001
Yes	171 (28.0)	56 (52.0)	40 (69.8)	
No	441 (72.0)	52 (48.0)	18 (30.2)	

Internet Activity

Almost all superusers (94.8%, 55/58) reported engaging in 11 or more Internet activities (eg, online banking, keeping a blog, and looking for travel information), whereas approximately one-quarter of rare users (26.6%, 163/612) reported the same engagement with other Internet activities (Table 6). Similarly, nearly three-quarters of superusers (74.1%, 43/58) reported accessing the Internet through 3 or more devices (eg, a home computer, work computer, and smartphone), which contrasts with less than one-third (29.4%, 180/612) of rare users doing the same.

Table 6. Internet activities and devices of respondents (N=778).

Internet activities and devices	Rare users, n (%) n=612	Users, n (%) n=108	Superusers, n (%) n=58	*P*
Breadth of Internet use (number of Internet activities)				<.001
1-5	109 (17.8)	2 (1.9)	1 (1.7)	
6-10	340 (55.6)	25 (23.1)	2 (3.4)	
≥11	163 (26.6)	81 (75)	55 (94.8)	
Breadth of Internet access (number of Internet devices)				<.001
1	203 (33.2)	15 (13.9)	7 (12.1)	
2	229 (37.4)	27 (25.0)	8 (13.8)	
≥3	180 (29.4)	66 (61.1)	43 (74.1)	

Discussion

Principal Findings

Within the past few years, a rapid expansion of technologies that people can use to generate their own Internet content has provided novel opportunities for members of the general population to share and access health information. Social media have facilitated this ever-growing shift from the production of health information online being in the hands of commercial enterprise and health systems, to being led by users themselves. This study, the first to investigate user-generated health content on the Internet in the United Kingdom, sought to understand this growing trend through people's responses to survey questions. One of the more striking findings in this study is that one-quarter of this sample of UK Internet users reported accessing and/or posting user-generated health information online. Because this is the first study of its kind in the United Kingdom, there are no equivalent data from which to study trends.

As has been repeatedly shown in studies of online health support groups, most participants do not actively post their own content, although they read what others have written [13-17]. Although it is not possible to directly measure the activity of these "lurkers," they represent such a large proportion of those engaged with UGC that we should see this as "normal" online behavior. Further exploration of the motivations and usage of UGC by these passively engaged individuals is required. By grouping responses to this survey according to the frequency with which people reported accessing and sharing UGC, we sought to understand the characteristics of people who are highly engaged in these behaviors—the superusers. Although our data do not provide information on the effects of UGC, our findings characterizing the 7.5% of respondents classified as superusers is valuable in describing the group who may be generating the content that others are viewing.

These superusers are predominantly male. Respondents in the youngest age category (16-24 years) are more likely to be superusers, whereas those in the oldest age category (55-74 years) are generally rare users. Superusers carry out more varied activities on the Internet, such as online banking and booking travel, than the users and rare users. Despite these differences, our results suggest that there are no differences in health status or health service use between the superusers and the other 2 groups. This suggests that people who are well may be as responsible for producing our online health content as people who are ill. To our knowledge, this is the first explication that UGC does not appear to be associated with self-reported use of health services. This study did not distinguish between the use of information that provides guidance for a diagnosis or treatment of a health condition, and information shared or accessed by people who are healthy and either seeking or offering lifestyle advice. Several possible mechanisms have been identified by which UGC and, in particular, others' experiences may affect health, such as finding information, feeling supported, and experiencing health services [5]. Therefore, our results are congruent with the contingent model of health information use, which suggests that health information is one important component of a health care experience, rather than as something that displaces or reduces use of health services [28].

It was surprising that about 20% of participants were unaware of the availability of UGC on the Internet, particularly as this was a sample of Internet users. The outcomes of UGC are unclear; therefore, we would not necessarily recommend interventions to increase access. Further research should focus on reasons why these individuals, who have access to the Internet and are motivated to complete a health survey, are unaware of the availability of this increasingly ubiquitous content.

Previous studies of health-seeking behavior on the Internet have found that females more commonly access health information on the Internet than males [1,4]. However, in the present study, superusers were more likely to be male. It has been suggested that men rate their self-assessed online skills higher than women do, which is one possible explanation for higher reported use of these complex online resources by men [29]. Although our data did not provide an opportunity to examine this issue further, it is an important area for future research as it guides content development and implementation.

Limitations

The survey used in this study was administered online. Administering a survey in this way enables rapid data collection, reduces administrative burden, and is cost-effective. It may cause selection bias because those who do not have Internet access as well as adequate digital skills are excluded [30]. In this study, rare users may have been excluded because of a lack of digital skills. Those with chronic illness may also have been less inclined to participate in an online survey or excluded because of fatigue or disability. Another limitation is the use of quota sampling (done to facilitate international comparisons in the overall survey), which reduces the generalizability of the results of this analysis to the overall population when compared with a true random sample. Although the sample was selected to be representative of UK Internet users, subsequent studies with different designs might sample purposively for superusers to explore their behavior and motivations more fully. Further research focusing on smaller age categories and going deeper into some of the life-cycle variables that may affect use of online health information (eg, parenting or retirement) would be useful as well. Furthermore, this study reports a secondary analysis of a larger survey on people's use of the Internet for health. Some potential areas of inquiry relevant to UGC were not included in the original survey, such as its use to obtain social support and to select health services based on others' experiences [5]. Finally, this study reports cross-sectional data, which precludes determining causal relationships. Because this is the first study of its kind in the United Kingdom, it is exploratory in nature and does not evaluate effects of UGC.

Conclusions

This study reports results of the first representative sample of UK Internet users that investigates accessing and sharing user-generated health content online. Within the context of available surveys from other countries, our results suggest that

UGC may be increasingly popular among those who are healthy and who have chronic conditions alike, and that a minority of people who frequently access and share may be primarily responsible for generating the majority of content that others view.

The potential benefits of ICT for health are vast, but it is likely that some online resources are effective and desirable for some people and they are not for others. Understanding the contexts in which they are helpful is important to be able to support individual patients, public health initiatives, and to develop information policy and strategy in clinics and health systems. Through understanding characteristics of those who already participate in accessing and sharing user-generated health content online, as well as the differences between groups of users based on the frequency with which they do so, the results of this study bring clarity to this important issue regarding use of ICT for health information. These findings inform an agenda for further research to identify why people access and share UGC, what the impact is on health behaviors and outcomes, and if expanding engagement with user-generated health content online is likely to be beneficial.

Acknowledgments

We would like to thank those who participated in the survey, as well as Ioannis Maghiros and Fabienne Abadie for their work on the original survey. We are grateful to Dr Brian Nicholson for his comments on earlier drafts of the paper. Braden O'Neill would like to thank the Rhodes Trust for their support of his studentship. This research was funded by The Institute for Prospective Technological Studies (IPTS), 1 of the 7 scientific institutes of the European Commission's Joint Research Centre (JRC), and by the European Commission Directorate General for Communications Networks, Content and Technology (DG Connect) as a part of the Strategic Intelligence Monitor on Personal Health Systems II project (SIMPHS II). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The views expressed are purely those of the authors and may not in any circumstances be regarded as stating an official position of the European Commission.

Conflicts of Interest

None declared.

Multimedia Appendix 1

"Citizens and Information Communication Technology for Health" survey.

[PDF File (Adobe PDF File), 6MB - jmir_v16i4e118_app1.pdf]

References

1. Fox S. The social life of health information. Washington, DC: Pew Internet & American Life Project; 2011 May 12. URL: http://www.pewinternet.org/Reports/2011/Social-Life-of-Health-Info.aspx [accessed 2013-12-10] [WebCite Cache ID 6LlbyKPHN]
2. Dutton W, Blank G. Next Generation Users: The Internet in Britain. Oxford: Oxford Internet Institute; 2011. URL: http://oxis.oii.ox.ac.uk/sites/oxis.oii.ox.ac.uk/files/content/files/publications/oxis2011_report.pdf [accessed 2013-12-10] [WebCite Cache ID 6LlcAtE27]
3. United Kingdom Department of Health. The Power of Information: Putting all of us in control of the health and care information we need. 2012. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213689/dh_134205.pdf [accessed 2013-12-10] [WebCite Cache ID 6LlcFrwgh]
4. Fox S. Peer-to-peer healthcare. Washington, DC: Pew Internet & American Life Project; 2011 Feb 28. URL: http://www.pewinternet.org/Reports/2011/P2PHealthcare.aspx [accessed 2013-12-10] [WebCite Cache ID 6LlcOT6Bd]
5. Ziebland S, Wyke S. Health and illness in a connected world: how might sharing experiences on the internet affect people's health? Milbank Q 2012 Jun;90(2):219-249 [FREE Full text] [doi: 10.1111/j.1468-0009.2012.00662.x] [Medline: 22709387]
6. Entwistle VA, France EF, Wyke S, Jepson R, Hunt K, Ziebland S, et al. How information about other people's personal experiences can help with healthcare decision-making: a qualitative study. Patient Educ Couns 2011 Dec;85(3):e291-e298. [doi: 10.1016/j.pec.2011.05.014] [Medline: 21652162]
7. Kamel Boulos MN, Wheeler S. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. Health Info Libr J 2007 Mar;24(1):2-23. [doi: 10.1111/j.1471-1842.2007.00701.x] [Medline: 17331140]
8. Entwistle V, Firnigl D, Ryan M, Francis J, Kinghorn P. Which experiences of health care delivery matter to service users and why? A critical interpretive synthesis and conceptual map. J Health Serv Res Policy 2012 Apr;17(2):70-78 [FREE Full text] [doi: 10.1258/jhsrp.2011.011029] [Medline: 21967821]
9. Herxheimer A, McPherson A, Miller R, Shepperd S, Yaphe J, Ziebland S. Database of patients' experiences (DIPEx): a multi-media approach to sharing experiences and information. Lancet 2000 Apr 29;355(9214):1540-1543. [doi: 10.1016/S0140-6736(00)02174-7] [Medline: 10801187]

10. Ranard BL, Ha YP, Meisel ZF, Asch DA, Hill SS, Becker LB, et al. Crowdsourcing--harnessing the masses to advance health and medicine, a systematic review. J Gen Intern Med 2014 Jan;29(1):187-203. [doi: 10.1007/s11606-013-2536-8] [Medline: 23843021]
11. 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]
12. Adams E, Boulton M, Watson E. The information needs of partners and family members of cancer patients: a systematic literature review. Patient Educ Couns 2009 Nov;77(2):179-186. [doi: 10.1016/j.pec.2009.03.027] [Medline: 19406609]
13. Cobb NK, Graham AL, Abrams DB. Social network structure of a large online community for smoking cessation. Am J Public Health 2010 Jul;100(7):1282-1289 [FREE Full text] [doi: 10.2105/AJPH.2009.165449] [Medline: 20466971]
14. Cunningham JA, van Mierlo T, Fournier R. An online support group for problem drinkers: AlcoholHelpCenter.net. Patient Educ Couns 2008 Feb;70(2):193-198. [doi: 10.1016/j.pec.2007.10.003] [Medline: 18022340]
15. Jones R, Sharkey S, Smithson J, Ford T, Emmens T, Hewis E, et al. Using metrics to describe the participative stances of members within discussion forums. J Med Internet Res 2011;13(1):e3 [FREE Full text] [doi: 10.2196/jmir.1591] [Medline: 21239373]
16. Selby P, van Mierlo T, Voci SC, Parent D, Cunningham JA. Online social and professional support for smokers trying to quit: an exploration of first time posts from 2562 members. J Med Internet Res 2010;12(3):e34 [FREE Full text] [doi: 10.2196/jmir.1340] [Medline: 20719739]
17. van Mierlo T. The 1% rule in four digital health social networks: an observational study. J Med Internet Res 2014;16(2):e33 [FREE Full text] [doi: 10.2196/jmir.2966] [Medline: 24496109]
18. Adams S. Sourcing the crowd in healthcare and the hierarchy of pharmaceutical knowledge. Presented at: The Co-Production of Knowledge; July 18-20, 2012; York, UK.
19. van Mierlo T, Voci S, Lee S, Fournier R, Selby P. Superusers in social networks for smoking cessation: analysis of demographic characteristics and posting behavior from the Canadian Cancer Society's smokers' helpline online and StopSmokingCenter.net. J Med Internet Res 2012;14(3):e66 [FREE Full text] [doi: 10.2196/jmir.1854] [Medline: 22732103]
20. Lupiáñez-Villanueva F, Maghiros I, Abadie F. Citizens and ICT for Health in 14 EU Countries: Results from an Online Panel. Seville, Spain: European Commission, Joint Research Centre, Institute for Prospective Technological Studies; 2012. URL: http://ftp.jrc.es/EURdoc/JRC71142.pdf [accessed 2013-12-10] [WebCite Cache ID 6LlW8T34h]
21. Eysenbach G, Wyatt J. Using the Internet for surveys and health research. J Med Internet Res 2002 Nov;4(2):E13 [FREE Full text] [doi: 10.2196/jmir.4.2.e13] [Medline: 12554560]
22. Dillman D, Smyth JD, Christian LM. Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method. Hoboken, NJ: Wiley & Sons; 2009.
23. Braithwaite D, Emery J, De Lusignan S, Sutton S. Using the Internet to conduct surveys of health professionals: a valid alternative? Fam Pract 2003 Oct;20(5):545-551 [FREE Full text] [Medline: 14507796]
24. Eurostat Statistical Office of the European Union. 2012. European social statistics URL: http://www.esomar.org/knowledge-and-standards/codes-and-guidelines/guideline-for-online-research.php [accessed 2014-01-15] [WebCite Cache ID 6Me9pOP0P]
25. European Social Statistics. Luxembourg: Eurostat Statistical Office of the European Union; 2012. URL: http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-31-12-666/EN/KS-31-12-666-EN.PDF [accessed 2013-12-10] [WebCite Cache ID 6LlfawO3P]
26. Jackson DA. Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. Ecology 1993:2204-2214.
27. United Nations Educational, Scientific and Cultural Organization Institute for Statistics. 1997. International standard classification of education URL: http://www.uis.unesco.org/Education/ISCEDMappings/Pages/default.aspx [accessed 2013-12-10] [WebCite Cache ID 6Llg4HqPk]
28. Heaney D, Wyke S, Wilson P, Elton R, Rutledge P. Assessment of impact of information booklets on use of healthcare services: randomised controlled trial. BMJ 2001 May 19;322(7296):1218-1221 [FREE Full text] [Medline: 11358776]
29. Hargittai E, Shafer S. Differences in actual and perceived online skills: the role of gender. Soc Sci Quart 2006;87(2):432-448.
30. Whitehead LC. Methodological and ethical issues in Internet-mediated research in the field of health: an integrated review of the literature. Soc Sci Med 2007 Aug;65(4):782-791. [doi: 10.1016/j.socscimed.2007.03.005] [Medline: 17512105]

Abbreviations

EU: European Union
ICT: information and communication technology
PCA: principal components analysis
UGC: user-generated content

Edited by G Eysenbach; submitted 13.12.13; peer-reviewed by T van Mierlo, M Broom; comments to author 12.01.14; revised version received 21.01.14; accepted 23.02.14; published 30.04.14.

Please cite as:
O'Neill B, Ziebland S, Valderas J, Lupiáñez-Villanueva F
User-Generated Online Health Content: A Survey of Internet Users in the United Kingdom
J Med Internet Res 2014;16(4):e118
URL: http://www.jmir.org/2014/4/e118/
doi:10.2196/jmir.3187
PMID:24784798

Original Paper

The Impact of Search Engine Selection and Sorting Criteria on Vaccination Beliefs and Attitudes: Two Experiments Manipulating Google Output

Ahmed Allam[1], MSc; Peter Johannes Schulz[1], PhD; Kent Nakamoto[2], PhD

[1]Institute of Communication and Health, Faculty of Communication Sciences, University of Lugano (Università della Svizzera italiana), Lugano, Switzerland

[2]Marketing Department, Pamplin College of Business, Virginia Tech, Blacksburg, VA, United States

Corresponding Author:
Ahmed Allam, MSc
Institute of Communication and Health
Faculty of Communication Sciences
University of Lugano (Università della Svizzera italiana)
Blue Building, 1st floor
13 G Buffi street
Lugano, 6900
Switzerland
Phone: 41 58 666 4821
Fax: 41 58 666 4647
Email: ahmed.allam@usi.ch

Abstract

Background: During the past 2 decades, the Internet has evolved to become a necessity in our daily lives. The selection and sorting algorithms of search engines exert tremendous influence over the global spread of information and other communication processes.

Objective: This study is concerned with demonstrating the influence of selection and sorting/ranking criteria operating in search engines on users' knowledge, beliefs, and attitudes of websites about vaccination. In particular, it is to compare the effects of search engines that deliver websites emphasizing on the pro side of vaccination with those focusing on the con side and with normal Google as a control group.

Method: We conducted 2 online experiments using manipulated search engines. A pilot study was to verify the existence of dangerous health literacy in connection with searching and using health information on the Internet by exploring the effect of 2 manipulated search engines that yielded either pro or con vaccination sites only, with a group receiving normal Google as control. A pre-post test design was used; participants were American marketing students enrolled in a study-abroad program in Lugano, Switzerland. The second experiment manipulated the search engine by applying different ratios of con versus pro vaccination webpages displayed in the search results. Participants were recruited from Amazon's Mechanical Turk platform where it was published as a human intelligence task (HIT).

Results: Both experiments showed knowledge highest in the group offered only pro vaccination sites (Z=–2.088, P=.03; Kruskal-Wallis H test [H_5]=11.30, P=.04). They acknowledged the importance/benefits (Z=–2.326, P=.02; H_5=11.34, P=.04) and effectiveness (Z=–2.230, P=.03) of vaccination more, whereas groups offered antivaccination sites only showed increased concern about effects (Z=–2.582, P=.01; H_5=16.88, P=.005) and harmful health outcomes (Z=–2.200, P=.02) of vaccination. Normal Google users perceived information quality to be positive despite a small effect on knowledge and a negative effect on their beliefs and attitudes toward vaccination and willingness to recommend the information (χ^2_5=14.1, P=.01). More exposure to antivaccination websites lowered participants' knowledge (J=4783.5, z=−2.142, P=.03) increased their fear of side effects (J=6496, z=2.724, P=.006), and lowered their acknowledgment of benefits (J=4805, z=–2.067, P=.03).

Conclusion: The selection and sorting/ranking criteria of search engines play a vital role in online health information seeking. Search engines delivering websites containing credible and evidence-based medical information impact positively Internet users seeking health information. Whereas sites retrieved by biased search engines create some opinion change in users. These effects are apparently independent of users' site credibility and evaluation judgments. Users are affected beneficially or detrimentally but are unaware, suggesting they are not consciously perceptive of indicators that steer them toward the credible sources or away from the dangerous ones. In this sense, the online health information seeker is flying blind.

(J Med Internet Res 2014;16(4):e100) doi:10.2196/jmir.2642

KEYWORDS

consumer health information; search engine; searching behavior; Internet; information storage and retrieval; online systems; public health informatics; vaccination; health communication

Introduction

Background

In 2012, more than 2 billion people worldwide used the Internet [1]. In the United States, 81% of adults reported Internet usage [2], whereas 73% of households in the United Kingdom had access to Internet [3]. One of the main purposes of Internet usage is seeking health information [2-6]. In fact, several studies reported that looking for medical information on the Internet was the first resort by individuals [2,4,5,7,8]. Despite the abundance of medical portals and health-related websites, many studies showed the main gateway for seeking health information was through search engines and, in particular, general search engines [2,9-11]. According to one study, 8 in 10 online health information seekers started with a search engine (Google, Bing, or Yahoo) [2].

In 2012, Google, the most popular search engine, processed as many as 1.2 trillion search queries in 146 different languages [12]. This number illustrates the importance of search engines for information seeking on the World Wide Web. A recent study [13] using "breast cancer" as a keyword found that results retrieved by 4 widely used search engines overlapped considerably, but each engine had a unique way of sorting/ranking (and thus emphasizing) various types of content. Because most searches produce many more results than anybody would be willing to read, people concentrate on the first results [10,14]. This gives the sorting/ranking algorithms of the search engines tremendous influence over the global spread of information and other communication processes. This study is concerned with demonstrating the influence of selection and sorting/ranking criteria operating in search engines on users' knowledge, beliefs, and attitudes.

The promise of the Internet as a decision support system that makes available to patients not only a vast array of information, but also advice, product promotion, services, and even decision aids for health care and health maintenance [15] is seriously marred by the fact that the quality of health information on the Internet varies tremendously [16-24]. Many challenges arise from this variance, especially from the low-quality information content for people who might be literate enough to find, understand, and process such information and store it in their memories, but not literate enough to recognize it for what it is (false, irrelevant, or fraudulent). This is why, in the context of false or misleading health information, we speak of bad (or dangerous) health literacy [15], meaning the presence of the ability to understand medical information turned sour by the simultaneous absence of the ability to recognize it as false.

Quality criteria for health websites were developed in scholarly literature [17-19,24,25-31], but they cannot be expected to affect the everyday information-searching behavior of Internet users. At least 3 reasons exist for quality deficits making the Internet a potentially dangerous decision support system in matters of health. First, some Internet users might be incapable of telling high- from low-quality information. This problem was there before the Internet, but the sheer amount and diversity of health information available on the Web enlarges it considerably. Second, they might mistake the Internet as an authoritative source and not see the necessity of assessing the quality of the information. This is also not a new problem, and is likely increased by the specific nature of the Internet, including the easiness with which information can be retrieved. Third, users might mistake a search engine's ranking of results as a quality ranking. This means they trust the search engine to provide the best websites and transfer this trust to the sites and the information they offer.

Quality deficits might result from knowledge deficits in the communicator because anybody can post health information on the Web. Quality deficits may also originate from material interests of communicators, such as businesses in the health sector that operate websites primarily as advertising for their products or services. Also, quality deficits may be caused by advocacy. As with commercial interests, strong advocacy in a controversial issue may mar the perception of facts relevant to the issue.

Previous Studies

Early studies [16,20-24] discussed the challenges that arise from the variability of the quality of health information. Information overload, searching difficulties, and lack of organization, regulation, quality, and accurate information [23] are some of these challenges. Many initiatives and ideas were proposed to solve or reduce this problem. These initiatives included proposing guidelines or quality indicators for users [27-29,31], suggesting a seal of approval or code of conduct [20,27,28,30], promoting an eHealth code of ethics [32], evaluating website design features, and factors for boosting websites' credibility [33-35]. Other studies suggested the use of rating systems, by giving the physicians and medical societies a role in applying filtering and labeling technologies [20].

The common methods used to assess the quality of the websites were to ask participants, patients, or physicians to rate a set of websites retrieved by multiple search engines [19,21,36-40]. Tools like DISCERN [41] and its brief version [42] were developed to assess patients' written information with respect to treatment options; they were modified later for validating health information on the Internet. Many other studies have developed questionnaires that serve as measurements of the credibility and the quality of websites [34,35]. Google PageRank [43] may be a potential quality indicator after having medium correlation with ratings from DISCERN and the evidence-based quality of content [37]. Additionally, the readability level was also assessed for health-related websites [19,36], which showed that the majority of health-related websites required high school level or above [19,44]. This presented another challenge to understanding the online health material, especially for people with low literacy [44,45].

Not only was the quality of websites investigated, studies also evaluated the efficiency of search engines and the relevance of retrieved websites among multiple search engines. These studies proceeded by measuring the coverage and accuracy of the retrieved content [19] or by measuring the share of relevant content among the retrieved search results [38-40] from multiple search engines. Additional methods looked into the log files of search engines to understand what cancer-related queries people search for [8]. Another study combined transactional log analysis with a complementary pilot study to understand users' online behavior and navigational trends at ClinicalTrials.gov [11]. In addition, keyword effectiveness indexing was applied and explored to estimate the ability of search engines to retrieve relevant results [46].

Other experiments observed participants in the laboratory during their online health information seeking [10,44,47-50]. Most of the studies started with a hypothetical scenario and asked the participants to perform a related search either for themselves or for others. The participants were audio/video recorded and their computer screen was also recorded. Additional log files were used for further analysis. Most of these studies [10,44,47-49] used a think-aloud protocol [51] in which participants spoke while they were searching. Some of the studies contained in-depth interviews [10,52,49] and/or focus groups [10,48].

Other studies looked at eHealth literacy of college students aiming at a career in a health profession, by measuring their research skills [50,53]. These studies used the Research Readiness Self-Assessment (RRSA) based on the Information Literacy Competency Standards for Higher Education. The RRSA measures proficiency in obtaining health information, evaluating the quality of health information, and understanding plagiarism [50].

To our knowledge, few or none of the previous studies manipulated a widely known search engine, such as Google, with the goal of systematically studying the effect of its sorting/ranking and selecting algorithm on users' knowledge, beliefs, and attitudes toward a controversial health topic. By manipulating the search engine, we aimed at ascertaining the effects of using normal (unmanipulated) Google against a manipulated search on the beliefs and attitudes toward a particular health topic.

Hypotheses

The subject for our study was the medical controversy around vaccination. Vaccination is one of the most important and influential medical discoveries, protecting and saving millions of lives [54-56]. In spite of its benefits, it has come into criticism and controversy instigated by antivaccination activists and organizations that have a strong online presence. The vaccination issue is one that plots medical evidence against lay skepticism or resistance [57]. The medical evidence considers the benefits of vaccination to clearly outweigh its risks. That means that the position against vaccination is only tenable if pertinent medical knowledge is disregarded or misrepresented. Therefore, in our view, website quality and position on the issue are intertwined and cannot be treated separately. We compared a search engine that delivered search results from high-quality provaccination websites with another that yielded lower-quality antivaccination websites.

Because websites and health interventions are known to increase knowledge [52,58-61] and high-quality sites contain more correct information on vaccination, we expect that users offered information (webpages) from high-quality websites by their search engine will gain more knowledge than users offered information from low-quality websites (hypothesis 1).

The effect of message tendency on recipient opinion may be the oldest theme in communication effects study. In light of this tradition, we expect that the higher the share of webpages (retrieved search results) from antivaccination websites offered by a search engine, the more critical users' beliefs and attitudes on vaccination will become (hypothesis 2).

Favorable message and source assessment can be considered a prerequisite for communication effects [62-64]. There is much to say about source and message assessment than cannot be tested here (eg, how they are affected by pre-experimental attitude or how they change in the experiment). The concern here is with recognition of the website quality. We assume that people, even if they do not apply the whole catalog of scholarly quality criteria, can recognize the quality of websites (or the lack thereof). Consequently, users should assess high-quality provaccination websites more favorably than low-quality antivaccination sites (hypothesis 3).

Methods

Overview

We conducted 2 experiments to test our hypotheses. The independent variable in both experiments was the proportion of webpages (retrieved search results) from high-quality provaccination and lower-quality antivaccination websites, achieved by manipulating the search engine's search space and sorting/ranking criteria. Dependent variables in both experiments were users' knowledge, beliefs, and attitudes, as well as their assessment of the websites. Both studies were conducted as an experiment which set an information-seeking task to participants, directed them to customized search engines (without being aware of it), left their choice of search terms and their selection of sites (from among those offered by the search engines) uncontrolled, and measured the dependent variables immediately after.

Experiment 1

Design

The first experiment was a pilot study to verify the existence of dangerous literacy in connection with searching and using health information on the Internet. Moreover, the goal was to explore the effect of 2 extreme manipulations of search engines that retrieved search results from either high-quality pro sites only or low-quality con sites only. A pre-post test design was employed to assess change, using the same questions and exact wording.

Experimental Conditions

Participants were allocated randomly to 3 experimental groups. Group 1 used normal Google, with its search coverage being the whole Web. Group 2 used Google configured to search for information from a set of websites certified by the Health on the Net (HON) code [30], which aims to provide health information of better quality and trustworthiness on the Web. The set additionally included websites from the World Health Organization (WHO), Public Health Agency of Canada, and governmental health agencies in the United States, such as the National Institutes of Health (NIH), Centers for Disease Control and Prevention (CDC), Food and Drug Administration (FDA), and other similar trustworthy credible websites. Group 3 used an engine configured to search for information from websites, blogs, and forums that discourage vaccination or were run by antivaccination activists and movements. Participants were unaware of the search engine manipulation and of which experimental group they belonged to.

Manipulating the Search Engine

The customization of the search engines was achieved by limiting their search coverage to different predefined sets of websites. This manipulation was realized by configuring the context and annotation files of Google custom search engine [65]. Both files were written in Extensible Markup Language (XML) in which the context file described the features and the settings of the search engine and the annotation file listed the set of websites the search engine covered in addition to weights and indicators instructing the search engine how information (webpages) retrieved from these sites would show or rank in the search results. Moreover, it allowed the possibility to widen the search space by including all webpages and documents found under the same domain. The first annotation file (for the search engine used by group 2) contained the HON-certified websites plus similar trustworthy and credible websites mentioned previously. The second annotation (for the search engine used by group 3) contained a set of websites that supported the con side of vaccination or discouraged people from vaccination. The set was assembled by running several regular searches on Google and other search engines for vaccination using negative keywords. Examples of the terms employed are vaccination and autism, side effects of vaccination, dangerous vaccines, antivaccination movements, antivaccination groups, bad vaccines, vaccine efficacy and skepticism, stop vaccination, etc. The sites yielded were read and informally classified as antivaccination or not. The search ended when the set of antivaccination websites contained 88 items. For the control group 1, we used normal Google (unmanipulated) to search the whole Web.

Procedure

The participants were informed that they were going to take part in an experiment about seeking online health information conducted by the University of Lugano in Switzerland. They were asked to inform themselves about vaccination and were given 10 minutes to do so. The search engines were embedded in a webpage that used the regular search layout provided by Google. We randomly distributed the instruction sheets of the experiment in front of the computers in the laboratory. Participants entered and chose a workstation as they wished. Each instruction contained a link to 1 of 3 webpages, which the participant was to open to begin the study. After answering a pretest questionnaire, participants were asked to search for information about vaccination for 10 minutes. Participants were free to search with any keyword they wanted and with as many as they wanted during the 10-minute time frame.

The retrieved search results were presented as a set of 10 search result pages; each page contained 10 results, adding up to a total of 100 websites/pages. A pager from 1 to 10 was displayed at the bottom of each page.

Strong emphasis was placed on the fact that this search engine was powered by Google and results were retrieved in a way so that participants were not alerted to any manipulation. After the search phase ended, they were redirected to answer a posttest questionnaire.

Recruitment and Participants

Marketing students from Virginia (USA) enrolled in a study-abroad program in Lugano, Switzerland, were asked to participate in the experiment. It ran in the computer laboratory of the University of Lugano. The sample of students was N=39 (group 1: n=12, group 2: n=14, group 3: n=13); 21% (8/39) were males and 79% (31/39) females.

Measures

Students had to fill out a pre-post questionnaire. The pretest questionnaire included the following measures used in the ensuing analyses:

1. Vaccination knowledge (or vaccination literacy): Battery of 14 true/false items combined to produce a knowledge index.
2. Attitude and beliefs toward vaccination: Set of 10 items that were presented as statements about vaccination, which participants responded to on a Likert scale (1=completely disagree to 7=completely agree). An additional 5 items measured the level of perceived side effects and benefits of vaccination for adults and children.
3. Sources and assessment of health information: A total of 7 items identified the participants' main sources of health information and their trust in a set of predefined sources, measured on a Likert scale.
4. Sociodemographic items for gender, nationality, year of birth, level of education, type of work, and frequency of Internet usage. Two additional items were included if the participant had experience or worked in a medical

environment and if the participant knew anyone who had a negative experience or side effect(s) from vaccines.

For the posttest questionnaire, students were asked to answer the following items:

1. Vaccination knowledge (or vaccination literacy) and attitude and belief toward vaccination: same as described previously.
2. Trust in retrieved information and consulted websites: A total of 7 items measuring summarily the credibility, satisfaction, trustworthiness, and relevance of the information retrieved and the websites visited.
3. Persuasion measure: A total of 5 categorical items measuring participants' self-perceived persuasion by the sites visited.

Data Analysis

Wilcoxon signed rank test was used for testing significance within each group in a before/after experiment for the knowledge index and the different attitude and belief items. In addition, 1-way ANOVA and the Kruskal-Wallis H test were used to analyze if there was a difference in information quality assessment among the 3 groups.

Experiment 2

Design

The second experiment was developed to leverage on the first and replicate it with a larger sample. It also increased the number of experimental conditions allowing for different ratios of pro and con sites in the mix provided by the search engine. Participants were again allocated randomly to one of the experimental conditions, and the procedure was similar to experiment 1; however, a posttest-only design was used.

Experimental Conditions

Experiment 2 compared 5 experimental groups and 1 control group who again used normal Google (group 1). The experimental groups differed in the ratio of con versus pro vaccination retrieved webpages offered by the customized search engine. Con and pro webpages were offered in the ratio of 0:10, 4:6, 6:4, 8:2, and 10:0 to groups 2 to 6, respectively. Group 2 (ratio 0:10) corresponds to group 2 in experiment 1, and group 6 (ratio 10:0) is similar to group 3 in experiment 1.

Manipulating the Search Engine

To manipulate the search engine for experiment 2, we started by writing context files and annotation files as in experiment 1, one pair describing the search engine restricted to a con vaccination set of websites and another pair to a pro vaccination set of websites (HON-certified and similar trustworthy credible sources reported in experiment 1). By using JavaScript and the Google custom search application programming interface (API) [65] without manipulating the ranking or the searching algorithm of Google, we programmatically controlled the search execution of both search engines and the display of the search results delivered from both [66]. The first search engine was called ConVaccineSearcher and the second ProVaccineSearcher. As the participant entered a search query in the provided search box, the 2 engines were launched with the same query and the results retrieved were displayed according to the participant's experimental condition. For example, to achieve a ratio of 4:6, the first 4 results retrieved from ConVaccineSearcher were added on the top of the first 6 results retrieved from ProVaccineSearcher to build a search result page with 10 results.

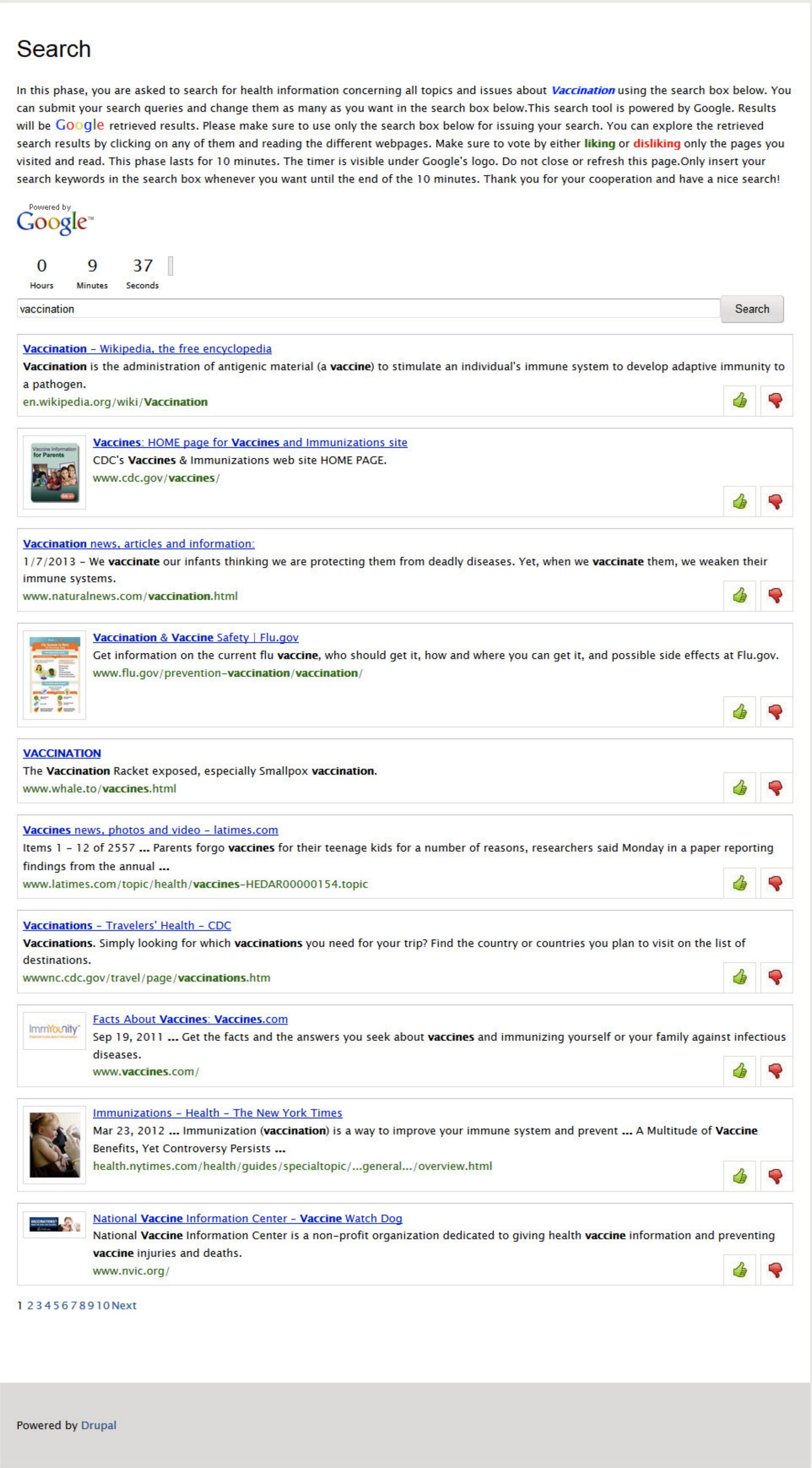
Procedure

Experiment 2 was presented as a website that proceeded in several steps similar to experiment 1. Figure 1 represents the search page presented to participants during the experiment. An additional like/dislike button was attached to every result (webpage). By clicking on these buttons, participants were able to rate the pages they read. As a visual indication for the rating action, the corresponding like/dislike button disappeared (Multimedia Appendix 1) once it was clicked. All the retrieved results (webpages), the like/dislike button, and the pager that represented a link to each of the 10 search result pages were attached to event handlers that updated the database using asynchronous JavaScript (AJAX), once they were clicked. This system was developed with the intention to record the behavior and the actions of each participant in a way that allowed constructing the timeline for each one.

After 10 minutes, an alert popped up informing the participants about the end of the search and redirecting them to the questionnaire. After completing the questionnaire, they were directed to a "thank you" page that displayed a generated code if the experiment was completed successfully. Multimedia Appendix 2 includes screenshots of these steps of the experiment.

Before the search and the questionnaire, participants were directed to a webpage (Multimedia Appendix 3) gateway that asked for the Amazon Mechanical Turk (MTurk) worker identification number before proceeding to the next step. MTurk is a crowdsourcing platform for requesting humans (MTurk workers) to work on executing specific tasks. In this way, we could track when they started the next phase and whether they passed through the experiment once or consecutively in one sitting. This ensured the validity of the flow of the experiment and allowed us to decide which participants followed the study instructions correctly.

Figure 1. Screenshot of an unmanipulated Google search page.



Search

In this phase, you are asked to search for health information concerning all topics and issues about ***Vaccination*** using the search box below. You can submit your search queries and change them as many as you want in the search box below.This search tool is powered by Google. Results will be Google retrieved results. Please make sure to use only the search box below for issuing your search. You can explore the retrieved search results by clicking on any of them and reading the different webpages. Make sure to vote by either **liking** or **disliking** only the pages you visited and read. This phase lasts for 10 minutes. The timer is visible under Google's logo. Do not close or refresh this page.Only insert your search keywords in the search box whenever you want until the end of the 10 minutes. Thank you for your cooperation and have a nice search!

vaccination

Vaccination – Wikipedia, the free encyclopedia
Vaccination is the administration of antigenic material (a **vaccine**) to stimulate an individual's immune system to develop adaptive immunity to a pathogen.
en.wikipedia.org/wiki/**Vaccination**

Vaccines: HOME page for **Vaccines** and Immunizations site
CDC's **Vaccines** & Immunizations web site HOME PAGE.
www.cdc.gov/**vaccines**/

Vaccination news, articles and information:
1/7/2013 – We **vaccinate** our infants thinking we are protecting them from deadly diseases. Yet, when we **vaccinate** them, we weaken their immune systems.
www.naturalnews.com/**vaccination**.html

Vaccination & **Vaccine** Safety | Flu.gov
Get information on the current flu **vaccine**, who should get it, how and where you can get it, and possible side effects at Flu.gov.
www.flu.gov/prevention-**vaccination**/**vaccination**/

VACCINATION
The **Vaccination** Racket exposed, especially Smallpox **vaccination**.
www.whale.to/**vaccines**.html

Vaccines news, photos and video – latimes.com
Items 1 – 12 of 2557 **...** Parents forgo **vaccines** for their teenage kids for a number of reasons, researchers said Monday in a paper reporting findings from the annual **...**
www.latimes.com/topic/health/**vaccines**-HEDAR00000154.topic

Vaccinations – Travelers' Health – CDC
Vaccinations. Simply looking for which **vaccinations** you need for your trip? Find the country or countries you plan to visit on the list of destinations.
wwwnc.cdc.gov/travel/page/**vaccinations**.htm

Facts About **Vaccines**: **Vaccines**.com
Sep 19, 2011 **...** Get the facts and the answers you seek about **vaccines** and immunizing yourself or your family against infectious diseases.
www.**vaccines**.com/

Immunizations – Health – The New York Times
Mar 23, 2012 **...** Immunization (**vaccination**) is a way to improve your immune system and prevent **...** A Multitude of **Vaccine** Benefits, Yet Controversy Persists **...**
health.nytimes.com/health/guides/specialtopic/...general.../overview.html

National **Vaccine** Information Center – **Vaccine** Watch Dog
National **Vaccine** Information Center is a non-profit organization dedicated to giving health **vaccine** information and preventing **vaccine** injuries and deaths.
www.nvic.org/

1 2 3 4 5 6 7 8 9 10 Next

Powered by Drupal

Recruitment and Participants

For experiment 2, a Web platform using Drupal framework was developed. We designed a human intelligence task (HIT) for each of the groups that explained the experiment in general terms asking would-be participants to go to the Web platform we prepared to start. We used the same template and wording of the HIT for all groups. We published a HIT, renewed each week, on the MTurk platform for approximately 6 weeks (October 9 to November 19, 2012) that included an interruption because of technical difficulties. Each HIT linked to the previously mentioned Web platform that allocated the participants to the different manipulations. JavaScript syntax was embedded in the webpage and controlled the workflow of loading, choosing the custom search engine, and allocating the participants by randomly generating a number from 1 to 6 when the page loaded. Participants who were registered as Amazon MTurk workers and who passed the qualification requirements described subsequently were able to preview the HIT and apply for it if they wanted. At the end of each week, we analyzed the submissions and accepted people who followed the experiment as explained. The qualification requirements for the HIT were 1-Number of HITs approved greater than or equal to 5000. The 2-HIT approval rate for all requesters' HITs was greater than 94%. The duration to complete the task was set to 30 minutes.

The sample size was N=197 (group 2: n=30, group 3: n=45, group 4: n=32, group 5: n=31, group 6: n=29, control group 1: n=30); 61.4% (121/197) were males and 38.6% (76/197) females. The nationalities of the participants were 0.5% (1/197) from England, 0.5% (1/197) from Hungary, 8.1% (16/197) from India, and 90.9% (179/197) from the United States. The participants' level of education was 17.8% (35/197) with high school level, 81.2% (160/197) with college or university level, and 1.0% (2/197) with vocational training. The mean age of the participants was 37.32 years (SD 11.39). The minimum age was 20 years and the maximum was 69 years. Additional information about the age and the self-reported profession of the participants for each group is represented in Multimedia Appendix 4.

Measures

In experiment 2, participants also had to fill a questionnaire after the search. It was similar to the posttest questionnaire in experiment 1 with the addition of:

1. Persuasion measure: A total of 9 items measuring participants' self-perceived persuasion by the sites visited on a Likert scale (versus the 5 categorical items in experiment 1).
2. Sociodemographic items that were presented in the pretest questionnaire of experiment 1.

Data Analysis

Knowledge scores were calculated for each participant and then ranked to apply the Kruskal-Wallis H test for testing for any significance among the experimental groups.

For the items measured on interval scales, we conducted factor analysis to find the latent variables [67]. Factor scores were calculated using Bartlett scores for each of the participants; Kruskal-Wallis H test, median test, and Jonckheere's trend test were used for the analysis of scores between the experimental groups. Jonckheere's test excluded the normal Google control group because the test is for a trend depending on decreasing and increasing shares of pro and con sites. The additional 2 items that indicated if the participants were persuaded by the information they read were analyzed using chi-square test. All analyses were conducted using SPSS 20 (IBM Corp, Armonk, NY, USA).

Results

Summary

All statistical analyses performed in this section were considered significant at $P<.05$. The Mann-Whitney U was used as a follow-up test for comparing different groups and a Bonferroni correction was applied; therefore, the significance level was changed to $P<.01$. Every time this occurs, it is clearly indicated. Moreover, we checked for the participants' characteristics among the groups in both experiments to see if there was any bias in any of the groups. The sociodemographic measure including the 2 items mentioned previously was the criteria. Results were insignificant in both experiments, implying that the participants in the experimental groups were similar with no significant differences among them. These results are reported in Multimedia Appendix 5.

Experiment 1

Knowledge

Knowledge questions from experiment 1 were calculated as scores and then ranked to apply the Wilcoxon signed rank test for checking any significance within each group in a before/after experiment. A significant increase in knowledge was observed in group 2 ($Z=-2.088$, $P=.03$), who were exposed to high-quality provaccination sites only. The other 2 groups did not show any significant increase in knowledge. An additional test was performed on the pretest knowledge scores of the 3 groups to check if there was any difference. The result was not significant ($H_2=4.02$, $P=.13$)

Beliefs and Attitudes

Two of the 15 attitude measures showed a significant change within group 2 (only high-quality provaccination sites), but none in groups 1 and 3. Importance of vaccination in adults against influenza ($Z=-2.326$, $P=.02$) and the effectiveness of vaccination against swine flu ($Z=-2.230$, $P=.03$), group 3 (only antivaccination sites) showed an increase in concern about the side effects of vaccination for adults ($Z=-2.582$, $P=.01$) and believing that vaccinations cause more harm than good ($Z=-2.200$, $P=.02$). In the remaining 11 measures, no significant change was observed.

Assessment of Information

Using 1-way ANOVA and Kruskal-Wallis test showed no difference between the 3 groups for trust in the information found ($F_{2,36}=1.83$, $P=.17$; $\chi^2_2=2.5$, $P=.28$), satisfaction with the information found ($F_{2,36}=1.84$, $P=.17$; $\chi^2_2=1.2$, $P=.54$), assessment of its persuasiveness ($F_{2,36}=0.99$, $P=.38$; $\chi^2_2=0.7$, $P=.68$), information relevance ($F_{2,36}=2.97$, $P=.06$; $\chi^2_2=5.4$, $P=.06$), and trust in Google ($F_{2,36}=3.07$, $P=.06$; $\chi^2_2=4.2$, $P=.12$). This means none of the measures employed produced any significant differences between the experimental groups.

Experiment 2

Knowledge

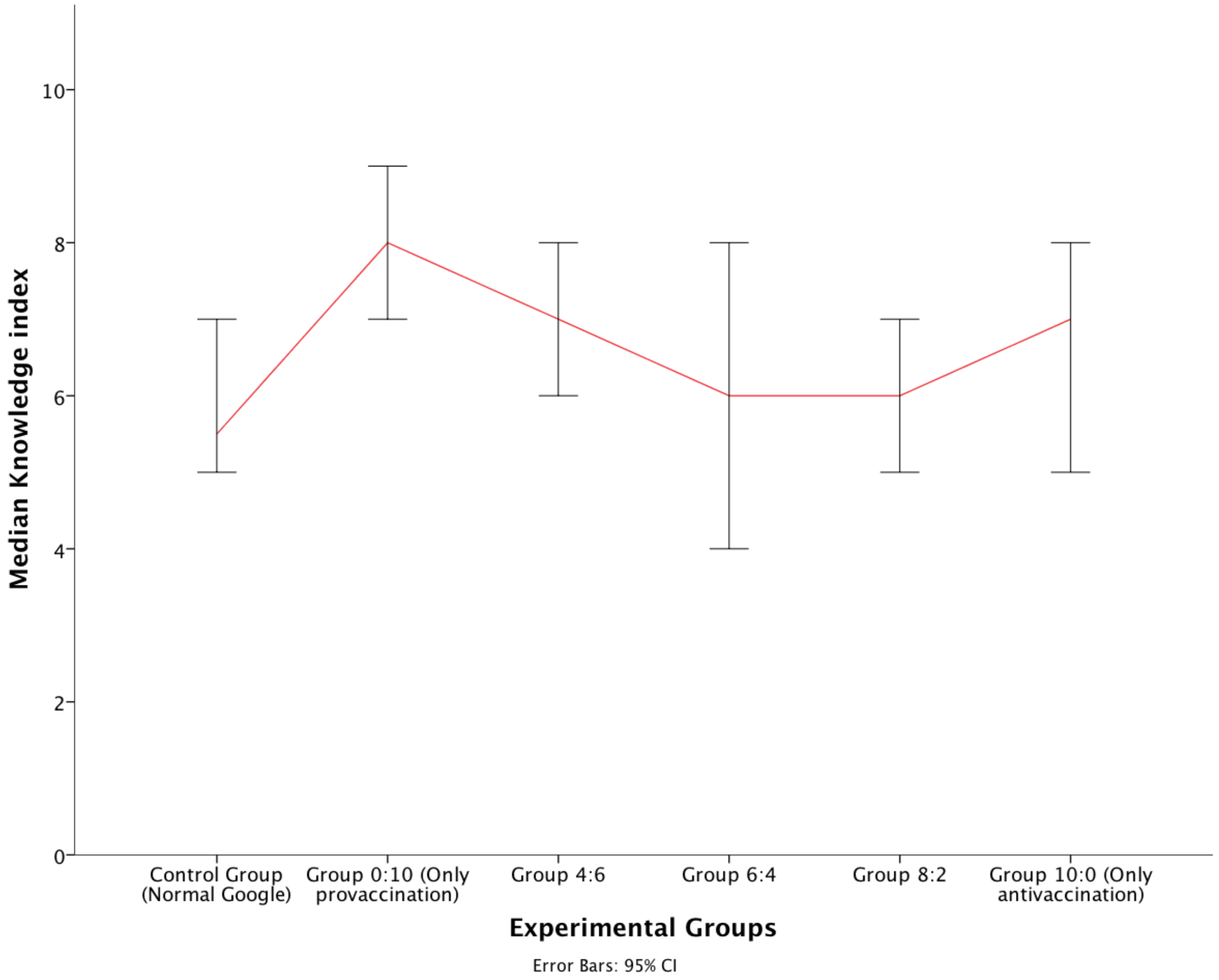
A knowledge index was computed for each participant based on the answers of 14 true-false questions in experiment 2. By using the Kolmogorov-Smirnov (*D*) and the Shapiro-Wilk tests, the distribution of the knowledge index (D_{197}=.10, P<.001) was found to be significantly nonnormal. As a result, we opted for nonparametric tests. The Kruskal-Wallis H test showed that the knowledge index was significantly affected by the different exposition to con vs pro vaccination websites (H_5=11.30, P=.04).

Mann-Whitney *U* tests were used to follow up this finding. A Bonferroni correction was applied and the effects were reported at .01 level of significance. The 4 comparisons between each of the groups 3 through 6 and normal Google (control group) did not show any significant result. Only group 2 showed a significantly higher level of knowledge versus the control group (*U*=241.5, *r*=–.4, *P*=.003). Figure 2 shows the medians for the experimental groups. It can be seen that as the share of search results belonging to antivaccination websites displayed increases, the median knowledge score decreases. In addition, the median of each group subjected to antivaccination websites is below the grand median.

Jonckheere’s trend test was applied to investigate whether the trend visible in Figure 2 was statistically significant. We hypothesized that the median of the knowledge index would decrease as the share of retrieved webpages from antivaccination websites in the search result pages increased. Jonckheere’s test revealed a significant trend in the data: the more participants were exposed to antivaccination websites, the lower the median of knowledge (*J*=4783.5, *z*=−2.142, *P*=.03).

In summary, the results from both experiments support our first hypothesis that users offered webpages from high-quality websites by their search engine will gain more knowledge than users offered webpages from low-quality websites.

Figure 2. Median vaccination knowledge scores and 95% CI (error bars) by experimental group (experiment 2).



Beliefs and Attitudes

In the factor analyses of attitudinal measures, an additional step was performed to check for multicollinearity by evaluating the variance inflation factor (VIF), making sure the items under analysis maintained low levels of VIF (<3). The factor analysis ran on 21 items with oblique rotation (promax) using the maximum likelihood extraction method. The Kaiser-Meyer-Olkin (KMO) measure verified the sampling adequacy for the analysis (KMO=.894), and by looking at the diagonal of the anti-image correlation matrix, all KMO values for individual items were >.70, well above the acceptable limits.

Bartlett's test of sphericity (χ^2_{190}=1784.1, P<.001) indicated that correlations between items were sufficiently large for factor analysis. An initial analysis was run to obtain eigenvalues for each component in the data. Four eigenvalues were greater than Kaiser's criterion of 1 and in combination explained 60% of the variance. We conducted a parallel analysis and looked into the scree plot in Figure 3, in which an inflection point was detected on the fourth eigenvalue. For this reason, we retained 3 factors for our final analysis. To assess the reliability of the subscales that emerged from factor analysis, we measured Cronbach alpha, which is the most common measure of scale reliability. Table 1 shows the pattern matrix of the factor analysis including each item's loadings on the obtained 3 factors. The eigenvalues were 7.13, 2.46, and 1.46 and the percentage of variance explained by each from the total variance was 35.64%, 12.34%, and 7.33%, respectively. In addition, the Cronbach alpha for each subscale that comprised items that had greater than or equal 0.4 loadings was .88, .86, and .75 respectively. The items that form factor 1 suggest that it represents skepticism/fear of vaccination side effects, acknowledgment of vaccination benefits for factor 2, and information qualities for factor 3. Factors 1 and 2 were negatively correlated (–.67) whereas both were almost zero correlated with factor 3.

To compare the groups, factor scores were calculated using the Bartlett method. Applying the Kruskal-Wallis H test revealed that the level of fear for vaccination side effects was significantly affected by the different exposure to con versus pro vaccination websites (H_5=16.88, P=.005).

Mann-Whitney *U* tests were used to follow up this finding. A Bonferroni correction was applied and the effects were considered significant at .01 alpha level. The 5 comparisons between each of the groups 1, 3, 4, 5, and 6 and group 2 (offered only high-quality provaccination sites) showed significant results, as reported in Table 2. Figure 4 shows the median of factor 1 score by the experimental groups. The most important comparison was with the real scenario that included normal Google (control group). Normal Google and group 2 differed, but the other groups did not differ from the control group. Thus, normal Google offered websites that created as much fear of vaccination as any of the customized search engines.

Table 1. Factor analysis of attitude and website assessment measures in experiment 2.

Pattern matrix[b]	Factor loadings		
	1	2	3
When recommending vaccination, doctors do not pay enough attention to side effects	.807[a]	.093	.055
In your opinion, how serious are the side effects of vaccination on adults?	.746[a]	.016	–.027
Many vaccinations today do more harm than good	.705[a]	–.127	.035
Many vaccinations recommended today are not really necessary because the disease is more or less extinct	.674[a]	–.106	.069
In your opinion, how serious are the side effects of vaccination on kids?	.668[a]	–.117	–.007
People who vaccinate run a risk of getting the disease from the vaccination	.630[a]	–.055	.067
When I read in the websites about the efficacy of vaccination I felt worried	.621[a]	–.011	–.054
Vaccination often does not really protect against a disease	.583[a]	–.207	.108
Health authorities should put the necessity of vaccination programs to the test more often	.564[a]	.348	.104
In your opinion, should adults get vaccinated for influenza?	–.528[a]	.040	.219
In your opinion, should babies get vaccinated for Hepatitis B?	–.402[a]	.039	.085
In your point of view, how effective is vaccination against swine flu?	–.399[a]	.120	.205
Vaccination is one of the great medical breakthroughs affecting our lives	.122	.931[a]	–.053
If it weren't for vaccination, many people today would have a shorter life span than they do	–.013	.794[a]	–.029
Sustaining and preserving current vaccination programs is a top priority of public health in our country	–.041	.686[a]	.118
People who opt against vaccination not only put themselves but also other people at risk	–.276	.568[a]	–.054
How much do you trust the information on vaccination you found in your search just now?	–.032	–.089	.779[a]
Was the information you found relevant?	.079	.094	.642[a]
I think that the information about vaccination I've read now from websites is comprehensible for me	–.111	–.092	.639[a]
How much do you trust Google to provide you with good information?	.034	.069	.582[a]

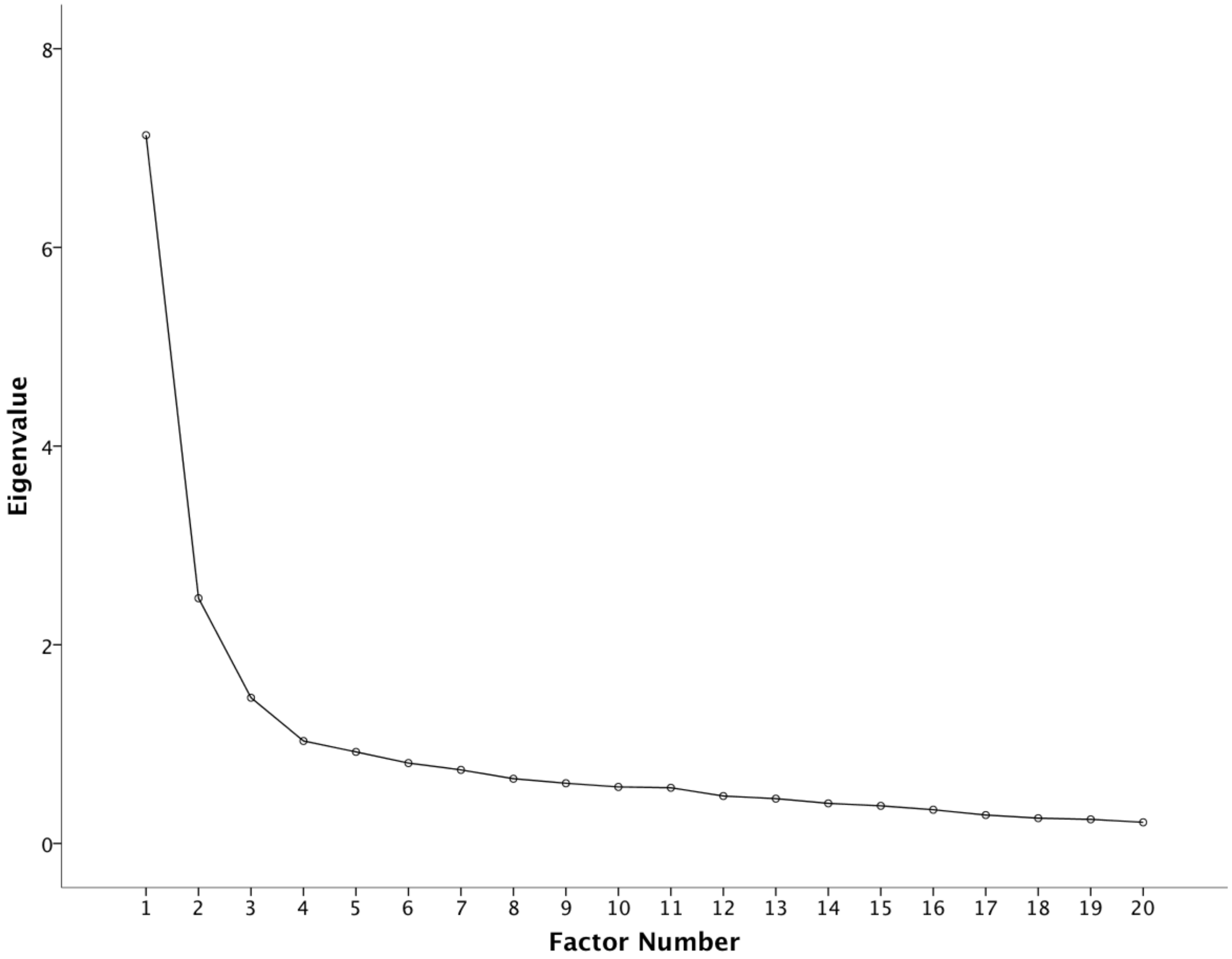
[a] Loading ≥0.4.

[b] Extraction method: maximum likelihood; rotation method: Promax with Kaiser normalization.

Table 2. Differences in user fears of vaccination between group 2 (only high-quality pro sites) and all others.

Group 2 vs:	Mann-Whitney *U*	Effect size *r*	*P*
Group 3	331	–.42	<.001
Group 4	252	–.40	.001
Group 5	280	–.34	.003
Group 6 (only con vaccination sites)	220	–.42	.001
Group 1 (normal Google)	282	–.32	.01

Figure 3. Scree plot for the factor analysis.



Regarding factor 2 representing the acknowledgment of vaccination benefits, Figure 4 displays the median factor score for different experimental groups. The Kruskal-Wallis H test showed a significant result (H_5=11.34, *P*=.04). Using Mann-Whitney *U* tests with Bonferroni correction applied, significant results were reported at .01 alpha level. Only the 2 comparisons between groups 2, 3, and 6 showed significant results (*U*=444, *r*=–.28, *P*=.01 and *U*=273, *r*=–.31, *P*=.01, respectively).

To investigate whether there was a significant trend in the median of the attitude factor scores among different groups, Jonckheere’s trend test was applied. The groups were ordered 0:10, 4:6, 6:4, 8:2, 10:0, with the share of con vaccination websites rising step-by-step. Significant results are reported in Table 3.

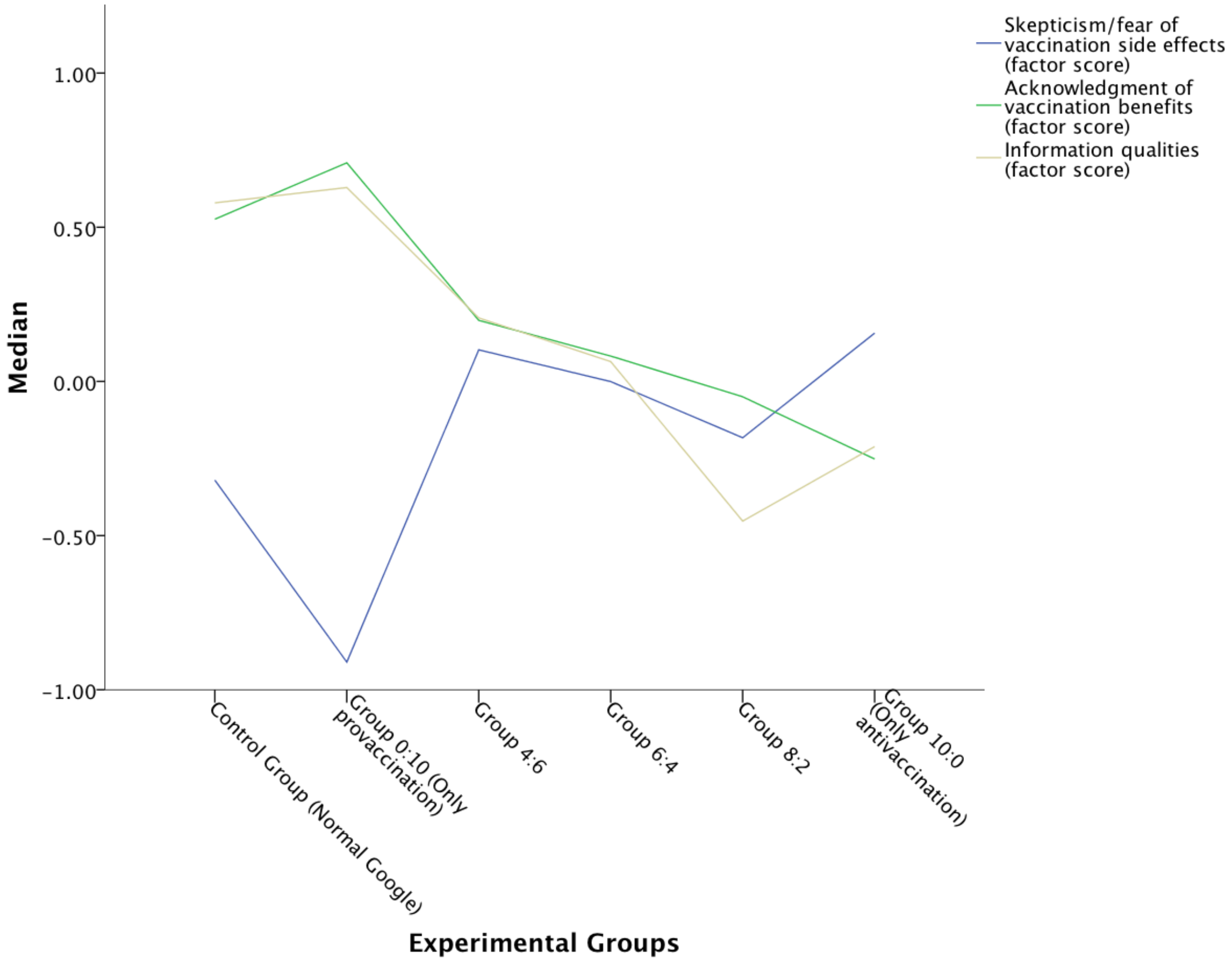
Table 3 suggests that as the share of con vaccination websites displayed increased, the fear of vaccination side effects also increased, whereas the acknowledgment of vaccination benefits decreased. In other words as the con/pro ratio changed from 4:6 up to 10:0 (only con vaccination websites), participants became more and more fearful of vaccination and acknowledged less and less that it had benefits.

The presented results from both experiments support hypothesis 2 that the ratio of con and pro vaccination webpages offered by the search engine affects information seekers’ views on vaccination. The more con sites offered by the search engine, the more skeptical they became.

Table 3. Jonckheere’s trend test results for attitude factor scores.

Factor score	Jonckheere trend test score		*P*
	Observed, *J*	Standardized, *z*	
Factor 1: skepticism/fear of vaccination side effects	6496	2.724	.006
Factor 2: acknowledgment of vaccination benefits	4805	–2.067	.03

Figure 4. Median of the skepticism/fear of side effects, acknowledgment of benefits, and information quality factor scores by experimental group (experiment 2).



Assessment of Information

Factor 3 offered a first measure of information assessment, which found a significant result (H_5=22.8, *P*<.001) in the Kruskal-Wallis H test. Comparisons that showed significant results after using Mann-Whitney *U* with Bonferroni correction at .01 alpha levels are reported in Table 4.

Table 4. Differences among groups in assessment of information quality.

Group comparison	Mann-Whitney *U*	Effect size *r*	*P*
Group 2 vs 5	216	–.45	<.001
Group 2 vs 6	231	–.40	.001
Group 1 vs 5	229	–.43	.001
Group 1 vs 6	242	–.38	.002

Additionally, 2 items that are indicators of perceived information quality were analyzed using chi-square test. The items measured whether the participant perceived any opinion change toward vaccination and, as an indicator of general assessment of the information retrieved, would recommend the information found.

The first item in Figure 5 resulted in χ^2_5=10.9, which corresponded to a value of *P*=.05 on the threshold of rejecting the null hypothesis. Participants in the group that did not receive lower-quality con sites, less often than the other groups attested to second thoughts after reading the information. Figure 6 shows the significant result (χ^2_5=14.11, *P*=.01) for the second item, indicating that when the share of the offered high-quality provaccination webpages decreased, the participants were less inclined to recommend to others the information they had retrieved. However, most in all groups recommended the information they found during the search phase.

The result from the first item implies that as the share of con vaccination websites increased, the doubt in vaccination efficacy increased. Moreover, the second item implied that health information seekers on the Web did recognize the quality of the sites they read to some extent, although most in all experimental groups recommended the information retrieved from the search engine.

We applied Jonckheere's trend test for the information quality score as we did for the attitude scores. The test showed that as the share of displayed webpages belonging to low-quality con vaccination websites increased, the assessment of information quality decreased (*J*=4154, *z*=−3.911, *P*<.001)

Additional information regarding participants' search terms is reported in Multimedia Appendix 6. The participants' ratings of con and pro vaccination webpages is evident from the clicks on the like/dislike buttons. Table 5 displays the results. The webpages belonging to high-quality provaccination sites were liked more often and disliked less often than pages belonging to lower-quality antivaccination sites. The latter, however, were on average also liked. In fact, 83% of the judgments of the high-quality webpages and 64% of the judgments of low-quality pages were positive. The difference of approximately 20% was similar across the groups. The mix of sites offered did not affect the evaluation of the sites.

Table 5 shows that participants who were offered high shares of high-quality information looked at more pages than the groups offered high shares of low-quality information. Various interpretations can be considered. Lower-quality information (eg, may be more difficult to process) could make respondents stay longer on the respective sites. It could also be that high-quality information was a bit boring, which could make respondents move on faster.

Table 5. Overview of webpage choice and evaluation by using the like/dislike button (experiment 2).

Webpages evaluation	Experimental group (ratio con to pro sites)					Total N=167
	2 (0:10) n=30	3 (4:6) n=45	4 (6:4) n=32	5 (8:2) n=31	6 (10:0) n=29	
Background						
Total unique webpages looked at, n	258	523	236	243	189	1449
Total like/dislike clicks, n	451	1086	569	429	336	2871
Webpages looked at per participant	8.6	11.6	7.4	7.8	6.5	8.7
Like/dislike clicks per participant	15.0	24.1	17.8	13.8	11.6	17.1
Like/dislike clicks per webpage looked at	1.7	2.1	2.4	1.8	1.8	2
Higher-quality provaccination webpages						
Webpages looked at,[a] n	258	252	68	42	—	620
Like/dislike clicks, n	451	533	189	63	—	1236
Like clicks, n (%)	341 (75.6)	465 (87.2)	165 (87.3)	49 (77.7)	—	1020 (82.5)
Dislike clicks, n (%)	110 (24.3)	68 (12.7)	24 (12.6)	14 (22.2)	—	216 (17.4)
Lower-quality antivaccination webpages						
Webpages looked at,[a] n	—	271	168	201	189	829
Like/dislike clicks, n	—	553	380	366	336	1635
Like clicks, n (%)	—	379 (68.5)	231 (60.7)	223 (60.9)	227 (67.5)	1060 (64.8)
Dislike clicks, n (%)	—	174 (31.4)	149 (39.2)	143 (39.0)	109 (32.4)	575 (35.1)
Proportion of lower-quality antivaccination webpages						
Proportion of lower-quality webpages in experimental manipulation (%)	0	40	60	80	100	
Proportion of lower-quality webpages looked at (%)	0	52	71	83	100	

[a] Most information not available for control group.

Groups offered a more evenly distributed mix of search result webpages from high- and low-quality sites tended to pass evaluation of the site more often. This could mean that the impression of 2 equally strong camps triggered site quality evaluations in users, whereas users offered a more uniform mix in either direction seemed less motivated to indicate how they thought of it. Across all groups who had a choice to select either type of sites, bias ran in favor of low-quality antivaccination sites, which were opened more frequently than their share in the manipulated search engine results would suggest. This could mean the con sites were, even just from looking at Google output, more interesting than the pro sites.

The results from experiment 2 support hypothesis 3 that webpages from high-quality sites were assessed more positively than webpages from lower-quality sites, which might be indicative of the ability to recognize high-quality medical websites. The hypothesis was not corroborated by experiment 1.

Figure 5. Bar graph of perceived persuasive effect of information retrieved by experimental group.

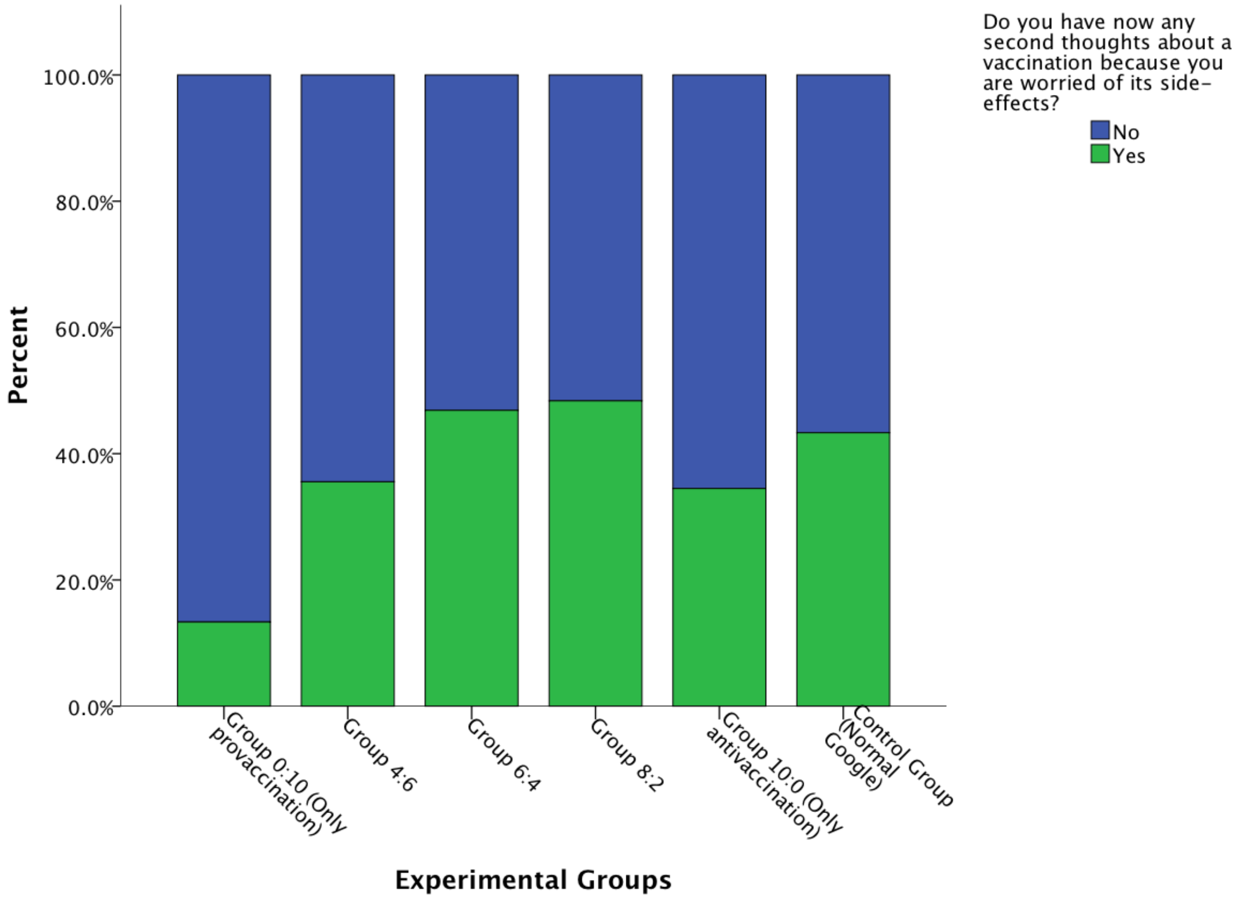
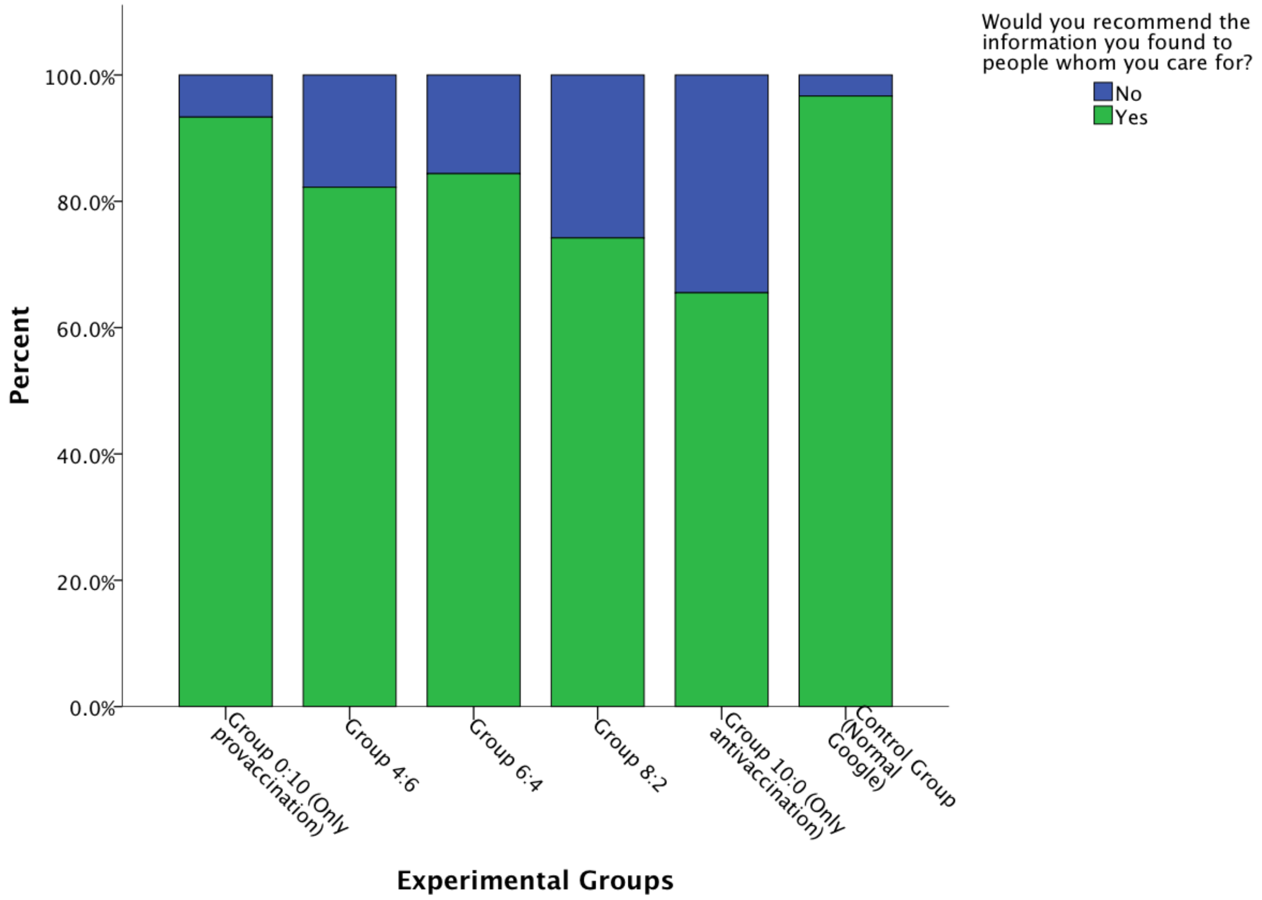


Figure 6. Bar graph of general assessment of information retrieved by experimental group.



Discussion

Principal Findings

The study presents a new type of experiment that manipulates the selection and sorting criteria of a general search engine to study their influence on users' knowledge and attitude toward health-related topics (vaccination in our case). The study retrieved results from 2 search spaces, one that was comprised of websites that were HON-certified or ran by credible trustworthy sources (eg, WHO, governmental health agencies) and supported the pro side of vaccination, and the other space was the websites skeptical of vaccination. The search engine algorithm for matching the queries with the website and all other parameters used in the algorithm to retrieve the relevant results and determine their rank/order were left intact.

The only group with an increase in knowledge in experiment 1 was the group offered only HON-certified and similar trustworthy webpages, and this group also differed from normal Google in experiment 2. This implies that knowledge gain by vaccination websites would be largest if websites critical of vaccination were filtered out. The result that participants scored higher in the knowledge index when they were less exposed to antivaccination websites supports assumptions that lay behind the initiatives promoting a health code of conduct [30], health-related seals of approval [16], and eHealth standards [31,32], and also the observation that websites with a high-quality rating using DISCERN or low scores on readability (grade levels) also contained health-related seals of approval [36]. As in Lorence and Greenberg [68], our study promotes the importance of quality proxies and encourages search engines to give higher rank to websites passed these quality criteria.

With regard to attitudes and beliefs about vaccination, the group offered only HON-certified and similar trustworthy credible webpages was the only group with positive change in experiment 1, and was the only group with lower levels of skepticism about vaccination in experiment 2, differing from both the normal Google group and all other experimental manipulations. In spite of a trend element showing more positive opinions as the proportion of provaccination Web pages grew, the results seem to indicate that attitude and belief change toward more support for vaccination programs depends on the absence of counter standpoints. Thus, the attitudes favored by websites offered by search engines translate into users' views on issues. Search engines' selection of websites according to bias, if it were to occur, can be expected to affect Internet users' views.

As to assessment of the information received, there was no difference between the groups in experiment 1 indicating participants were not able to tell good from bad sites. Actually, there was a trend element in experiment 2 showing the sites were assessed better the higher the share of HON-certified and other similar trustworthy sites was. But the lower-quality sites were not assessed as bad; rather, they were assessed as less good than the others. In other words, quality assessments are not completely independent of actual quality, but health information seekers seem to be blind to the fact that there are misleading and dangerous information sources on the Web. People were shown to hold positive views toward seeking health information even if they had been unsuccessful in such endeavors before [49].

This study finds 3 elements that together constitute bad or dangerous health literacy. The first element is the presence of bad information, caused by our experimental manipulation in this study, but certainly prevalent in reality. The variability in the quality of the content on the Internet has been demonstrated in many studies from the early stages of Internet adoption [16,20-24]. The second element is communication effects, in our case suboptimal knowledge gain and transfer of skeptical views into people's heads when they are presented with a large share of webpages from low-quality websites. Finally, a limited capacity for judging the informational value of the website was demonstrated and this is supposed to be the crucial condition that makes people misinform themselves on the Web. This capacity is not totally missing, but a weak sense of knowing what makes good sites does not exempt one from the influence of the bad ones.

Interestingly, results retrieved by normal Google were not as informative and beneficial to participants as the results retrieved by the manipulated search engine that retrieved only high-quality provaccination websites.The group offered provaccination websites only learned more than the normal Google group became less skeptical of vaccination; their support for vaccination was comparable as was their perceived information quality. Moreover, the intention to recommend the information found during the search was highest in the normal Google group. In fact, the factor that represented information quality was uncorrelated with the attitude scores. This suggests that effects are apparently independent of users' manifest of site credibility and evaluation judgments. Another supporting observation is that sites of low quality were rated comparatively positively and this could explain why the quality assessment appears to be unrelated to the effects.

Limitations

The study may have some limitations that are common to many online health information-seeking experiments. The search task is necessarily hypothetical and might turn out to be especially problematic if the participant is not interested or affected by the searching scenario. We chose, however, a medical topic that is considered more general and spans a wide range of people, especially all age groups. Vaccination concerns everybody and becomes more visible at the time of epidemics or the spread of infectious diseases, such as the H1N1 pandemic in 2009/2010. Therefore, we believe educating people and making sure they have good knowledge about vaccination and its benefits serves a good purpose in such scenarios. Moreover, Fox and Duggan [2] reported that half of online health information searches are conducted on the behalf of other individuals; this could justify the hypothetical scenario because searching for someone else might be similar to searching hypothetically.

The study exploited the design of an online system that took care of the experiment's workflow and, more importantly, the seamless tracking of the participants' actions during the entire search phase. The system recordings with the questionnaires' responses formed the measurements that we received from the participants. Many studies, specifically observational ones, use verbal protocols such as the think-aloud protocol to capture the online health information-seeking process. As a result, we believe such methods would result in richer details and could focus on aspects that might be missed by the online systems or the questionnaires. Therefore, the study could benefit by integrating and using such protocols as part of the measures, which completes the whole image and provides a deeper insight in the experiments.

Conclusions

By emulating a real-life scenario of a health-information search, this paper aimed to demonstrate suboptimal outcomes of such searches, not to attack search engines. Instead, we acknowledge the importance of the advances in the technology and the algorithms that are used by Google or any other widely used search engine. More importantly, our aim is to raise attention for the need of intelligent filters on top of these prominent engines that will help in redefining the search experience by providing different representation and evaluation of health information content on the Internet. Moreover, the study seeks to present new experiments that could be exploited further for future research especially in the area of search engine manipulation.

In conclusion, users are affected, be it beneficially or detrimentally, but the quality of the source of this effect largely escapes them. This suggests they are not consciously aware of indicators that steer them toward the promising sources or away from the dangerous ones. In this sense, the health information seeker on the Internet is flying blind.

Acknowledgments

The authors would like to thank the Swiss National Science Foundation (SNSF) for their ongoing support of this research.

Multimedia Appendix 1

Retrieved search results; each attached a like/dislike button and webpages 1,2 and 5 are rated.

[JPG File, 526KB - jmir_v16i4e100_app1.jpg]

Multimedia Appendix 2

Screenshots of the steps in the experiment after the search phase that includes pop up window indicating the end of the search phase, questionnaire phase and “Thank you” page displaying the automatic generated code.

[PDF File (Adobe PDF File), 503KB - jmir_v16i4e100_app2.pdf]

Multimedia Appendix 3

A gateway webpage that requests Mturk Worker ID before the search phase and the questionnaire phase.

[PDF File (Adobe PDF File), 67KB - jmir_v16i4e100_app3.pdf]

Multimedia Appendix 4

The age (in years) and self-reported profession of the participants in the six groups.

[PDF File (Adobe PDF File), 106KB - jmir_v16i4e100_app4.pdf]

Multimedia Appendix 5

Randomization check for Experiment 1 and 2. The experimental groups are compared to check if there is significant difference among them with respect the socio-demographic measure.

[PDF File (Adobe PDF File), 152KB - jmir_v16i4e100_app5.pdf]

Multimedia Appendix 6

Search terms used by participants during the search experiment. The size of the label is proportional to the frequency of occurrence of the keyword among all the groups.

[PDF File (Adobe PDF File), 69KB - jmir_v16i4e100_app6.pdf]

References

1. Internet World Stats. 2012. Internet Users in the World URL: http://www.internetworldstats.com/stats.htm [accessed 2013-03-10] [WebCite Cache ID 6F0YkwNot]
2. Fox S, Duggan M. Health Online 2013. Washington, DC: Pew Internet & American Life Project; 2013 Jan 15. URL: http://pewinternet.org/Reports/2013/Health-online.aspx [accessed 2013-07-23] [WebCite Cache ID 6IKHMUFmS]
3. Dutton WH, Blank G. Oxford Internet Survey 2011 Report. Oxford: Oxford Internet Institute, University of Oxford; 2011. Next Generation Users: The Internet in Britain URL: http://microsites.oii.ox.ac.uk/oxis/ [accessed 2013-07-23] [WebCite Cache ID 6IKGy0tu1]
4. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Arch Intern Med 2005;165(22):2618-2624. [doi: 10.1001/archinte.165.22.2618] [Medline: 16344419]
5. Hesse BW, Moser RP, Rutten LJ. Surveys of physicians and electronic health information. N Engl J Med 2010 Mar 4;362(9):859-860. [doi: 10.1056/NEJMc0909595] [Medline: 20200398]
6. Higgins O, Sixsmith J, Barry MM, Domegan C. ECDC Technical Report. Stockholm: ECDC; 2011. A literature review on health information-seeking behaviour on the web: a health consumer and health professional perspective URL: http://ecdc.europa.eu/en/publications/Publications/Literature%20review%20on%20health%20information-seeking%20behaviour%20on%20the%20web.pdf [accessed 2013-07-23] [WebCite Cache ID 6IKHshPyy]
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. Bader JL, Theofanos MF. Searching for cancer information on the internet: analyzing natural language search queries. J Med Internet Res 2003 Dec 11;5(4):e31 [FREE Full text] [doi: 10.2196/jmir.5.4.e31] [Medline: 14713659]
9. O'Keeffe J, Willinsky J, Maggio L. Public access and use of health research: an exploratory study of the National Institutes of Health (NIH) Public Access Policy using interviews and surveys of health personnel. J Med Internet Res 2011;13(4):e97 [FREE Full text] [doi: 10.2196/jmir.1827] [Medline: 22106169]
10. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002 Mar 9;324(7337):573-577 [FREE Full text] [Medline: 11884321]

11. Graham L, Tse T, Keselman A. Exploring user navigation during online health information seeking. AMIA Annu Symp Proc 2006:299-303 [FREE Full text] [Medline: 17238351]
12. Google Zeitgeist 2012. All 2012 search trends in the world URL: http://www.google.com/zeitgeist/2012/ [accessed 2013-03-10] [WebCite Cache ID 6F0Y2Lu0j]
13. Wang L, Wang J, Wang M, Li Y, Liang Y, Xu D. Using Internet search engines to obtain medical information: a comparative study. J Med Internet Res 2012;14(3):e74 [FREE Full text] [doi: 10.2196/jmir.1943] [Medline: 22672889]
14. Maurer H. Institute for Information Systems and Computer Media, Graz University of Technology. Austria: Graz University of Technology; 2007 Sep 30. Report on dangers and opportunities posed by large search engines, particularly Google URL: http://www.iicm.tugraz.at/iicm_papers/dangers_google.pdf [accessed 2013-03-14] [WebCite Cache ID 6F7JDMqVR]
15. Schulz PJ, Nakamoto K. "Bad" literacy, the Internet, and the limits of patient empowerment. In: AAAI Spring Symposium Series: Artificial Intelligence & Health Communication. 2011 Mar Presented at: AAAI Spring Symposium Series: Artificial Intelligence & Health Communication, 21st-23rd March; March 21-23, 2011; Palo Alto, CA URL: http://aaai.org/ocs/index.php/SSS/SSS11/paper/view/2430
16. Murray E, Lo B, Pollack L, Donelan K, Catania J, Lee K, et al. The impact of health information on the Internet on health care and the physician-patient relationship: national U.S. survey among 1.050 U.S. physicians. J Med Internet Res 2003;5(3):e17 [FREE Full text] [doi: 10.2196/jmir.5.3.e17] [Medline: 14517108]
17. Stvilia B, Mon L, Yi YJ. A model for online consumer health information quality. J Am Soc Inf Sci 2009 Sep;60(9):1781-1791. [doi: 10.1002/asi.21115]
18. Gustafson DH, Wyatt JC. Evaluation of ehealth systems and services. BMJ 2004 May 15;328(7449):1150 [FREE Full text] [doi: 10.1136/bmj.328.7449.1150] [Medline: 15142895]
19. Berland GK, Elliott MN, Morales LS, Algazy JI, Kravitz RL, Broder MS, et al. Health information on the Internet: accessibility, quality, and readability in English and Spanish. JAMA 2001;285(20):2612-2621. [Medline: 11368735]
20. Eysenbach G, Diepgen TL. Towards quality management of medical information on the internet: evaluation, labelling, and filtering of information. BMJ 1998 Nov 28;317(7171):1496-1500 [FREE Full text] [Medline: 9831581]
21. Kunst H, Groot D, Latthe PM, Latthe M, Khan KS. Accuracy of information on apparently credible websites: survey of five common health topics. BMJ 2002 Mar 9;324(7337):581-582 [FREE Full text] [Medline: 11884323]
22. Roberts JM, Copeland KL. Clinical websites are currently dangerous to health. Int J Med Inform 2001 Jul;62(2-3):181-187. [Medline: 11470621]
23. Cline RJ, Haynes KM. Consumer health information seeking on the Internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692 [FREE Full text] [Medline: 11780707]
24. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002 May;287(20):2691-2700. [Medline: 12020305]
25. Bernstam EV, Walji MF, Sagaram S, Sagaram D, Johnson CW, Meric-Bernstam F. Commonly cited website quality criteria are not effective at identifying inaccurate online information about breast cancer. Cancer 2008 Mar 15;112(6):1206-1213 [FREE Full text] [doi: 10.1002/cncr.23308] [Medline: 18266210]
26. Hardey M. 'E-health': the internet and the transformation of patients into consumers and producers of health knowledge. Info., Comm. & Soc 2001 Oct;4(3):388-405. [doi: 10.1080/713768551]
27. Wilson P. How to find the good and avoid the bad or ugly: a short guide to tools for rating quality of health information on the internet. BMJ 2002 Mar 9;324(7337):598-602 [FREE Full text] [Medline: 11884329]
28. Eysenbach G. Consumer health informatics. BMJ 2000;320(7251):1713-1716. [doi: 10.1136/bmj.320.7251.1713]
29. Purcell GP, Wilson P, Delamothe T. The quality of health information on the internet. BMJ 2002;324(7337):557-558. [doi: 10.1136/bmj.324.7337.557]
30. Boyer C, Selby M, Scherrer JR, Appel RD. The Health On the Net Code of Conduct for medical and health Websites. Comput Biol Med 1998 Sep;28(5):603-610. [Medline: 9861515]
31. World Health Organization. Geneva: World Health Organization; 1999. Medical products and the Internet: a guide to finding reliable information URL: http://apps.who.int/medicinedocs/en/d/Js2277e/ [accessed 2014-03-29] [WebCite Cache ID 6OR1rn8Bo]
32. e-Health Ethics Initiative. e-Health Code of Ethics (May 24). J Med Internet Res 2000;2(2):E9 [FREE Full text] [doi: 10.2196/jmir.2.2.e9] [Medline: 11720928]
33. Zulman DM, Kirch M, Zheng K, An LC. Trust in the internet as a health resource among older adults: analysis of data from a nationally representative survey. J Med Internet Res 2011;13(1):e19 [FREE Full text] [doi: 10.2196/jmir.1552] [Medline: 21324832]
34. Fogg BJ, Marshall J, Laraki O, Osipovich A, Varma C, Fang N, et al. What makes Web sites credible?: A report on a large quantitative study. In: CHI '01 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2001 Presented at: SIGCHI Conference on Human Factors in Computing Systems; March 31-April 4, 2001; Seattle, WA p. 61-68. [doi: 10.1145/365024.365037]
35. Fogg BJ, Marshall J, Kameda T, Solomon J, Rangnekar A, Boyd J, et al. Web credibility research: a method for online experiments and early study results. In: CHI '01 Proceedings of the SIGCHI Conference on Human Factors in Computing

Systems. 2001 Presented at: SIGCHI Conference on Human Factors in Computing Systems; March 31-April 4. 2001; Seattle, WA p. 295-296. [doi: 10.1145/634067.634242]

36. Kaicker J, Debono VB, Dang W, Buckley N, Thabane L. Assessment of the quality and variability of health information on chronic pain websites using the DISCERN instrument. BMC Med 2010;8:59 [FREE Full text] [doi: 10.1186/1741-7015-8-59] [Medline: 20939875]
37. Griffiths KM, Christensen H. Website quality indicators for consumers. J Med Internet Res 2005;7(5):e55 [FREE Full text] [doi: 10.2196/jmir.7.5.e55] [Medline: 16403719]
38. Ilic D, Bessell TL, Silagy CA, Green S. Specialized medical search-engines are no better than general search-engines in sourcing consumer information about androgen deficiency. Hum Reprod 2003 Mar;18(3):557-561 [FREE Full text] [Medline: 12615824]
39. Maloney S, Ilic D, Green S. Accessibility, nature and quality of health information on the Internet: a survey on osteoarthritis. Rheumatology (Oxford) 2005 Mar;44(3):382-385 [FREE Full text] [doi: 10.1093/rheumatology/keh498] [Medline: 15572390]
40. Ilic D, Risbridger G, Green S. Searching the Internet for information on prostate cancer screening: an assessment of quality. Urology 2004 Jul;64(1):112-116. [doi: 10.1016/j.urology.2004.02.027] [Medline: 15245946]
41. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111 [FREE Full text] [Medline: 10396471]
42. Khazaal Y, Chatton A, Cochand S, Coquard O, Fernandez S, Khan R, et al. Brief DISCERN, six questions for the evaluation of evidence-based content of health-related websites. Patient Educ Couns 2009 Oct;77(1):33-37. [doi: 10.1016/j.pec.2009.02.016] [Medline: 19372023]
43. Brin S, Page L. Anatomy of a large-scale hypertextual web search engine. In: WWW7 Proceedings from the 7th International World Wide Web Conference. 1998 Presented at: WWW7: The 7th International World Wide Web Conference; April 14-18, 1998; Brisbane, Australia URL: http://ilpubs.stanford.edu:8090/361/
44. Birru MS, Monaco VM, Charles L, Drew H, Njie V, Bierria T, et al. Internet usage by low-literacy adults seeking health information: an observational analysis. J Med Internet Res 2004 Sep 3;6(3):e25 [FREE Full text] [doi: 10.2196/jmir.6.3.e25] [Medline: 15471751]
45. Knapp C, Madden V, Wang H, Sloyer P, Shenkman E. Internet use and eHealth literacy of low-income parents whose children have special health care needs. J Med Internet Res 2011;13(3):e75 [FREE Full text] [doi: 10.2196/jmir.1697] [Medline: 21960017]
46. Lorence D, Hatton TL. Assessing domains of uncertainty in critical Web-based healthcare communities. J Med Syst 2006 Apr;30(2):117-122. [Medline: 16705996]
47. Keselman A, Browne AC, Kaufman DR. Consumer health information seeking as hypothesis testing. J Am Med Inform Assoc 2008;15(4):484-495 [FREE Full text] [doi: 10.1197/jamia.M2449] [Medline: 18436912]
48. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? Soc Sci Med 2007 May;64(9):1853-1862. [doi: 10.1016/j.socscimed.2007.01.012] [Medline: 17328998]
49. Zeng QT, Kogan S, Plovnick RM, Crowell J, Lacroix EM, Greenes RA. Positive attitudes and failed queries: an exploration of the conundrums of consumer health information retrieval. Int J Med Inform 2004 Feb;73(1):45-55. [doi: 10.1016/j.ijmedinf.2003.12.015] [Medline: 15036078]
50. Ivanitskaya L, O'Boyle I, Casey AM. Health information literacy and competencies of information age students: results from the interactive online Research Readiness Self-Assessment (RRSA). J Med Internet Res 2006;8(2):e6 [FREE Full text] [doi: 10.2196/jmir.8.2.e6] [Medline: 16867969]
51. Ericsson KA, Simon HA. Protocol Analysis: Verbal Reports as Data. Cambridge, MA: MIT Press; 1993.
52. Ziebland S, Chapple A, Dumelow C, Evans J, Prinjha S, Rozmovits L. How the internet affects patients' experience of cancer: a qualitative study. BMJ 2004 Mar 6;328(7439):564 [FREE Full text] [doi: 10.1136/bmj.328.7439.564] [Medline: 15001506]
53. Hanik B, Stellefson M. Int Electronic J of Health Educ. 2011. EHealth literacy competencies among undergraduate health education students URL: http://www.aahperd.org/aahe/publications/iejhe/upload/11_B-Hanik.pdf [accessed 2011-07-22] [WebCite Cache ID 60NJfktbm]
54. Centers for Disease Control and Prevention. 2004. Smallpox fact sheet URL: http://www.bt.cdc.gov/agent/smallpox/overview/disease-facts.asp [accessed 2013-03-10] [WebCite Cache ID 6F0ZjHPqx]
55. Fenner F, Henderson DA, Arita I, Jezekv Z, Ladnyi ID. Smallpox and its Eradication. Geneva: World Health Organization; 1988. URL: https://extranet.who.int/iris/restricted/bitstream/10665/39485/1/9241561106.pdf [accessed 2013-07-23] [WebCite Cache ID 6IKeK4RKN]
56. Stern AM, Markel H. The history of vaccines and immunization: familiar patterns, new challenges. Health Aff (Millwood) 2005;24(3):611-621 [FREE Full text] [doi: 10.1377/hlthaff.24.3.611] [Medline: 15886151]
57. Omer SB, Salmon DA, Orenstein WA, deHart MP, Halsey N. Vaccine refusal, mandatory immunization, and the risks of vaccine-preventable diseases. N Engl J Med 2009 May 7;360(19):1981-1988. [doi: 10.1056/NEJMsa0806477] [Medline: 19420367]

58. Crutzen R, Cyr D, de Vries NK. The role of user control in adherence to and knowledge gained from a website: randomized comparison between a tunneled version and a freedom-of-choice version. J Med Internet Res 2012;14(2):e45 [FREE Full text] [doi: 10.2196/jmir.1922] [Medline: 22532074]
59. 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]
60. Powell J, Inglis N, Ronnie J, Large S. The characteristics and motivations of online health information seekers: cross-sectional survey and qualitative interview study. J Med Internet Res 2011;13(1):e20 [FREE Full text] [doi: 10.2196/jmir.1600] [Medline: 21345783]
61. 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]
62. Hovland CI, Weiss W. The influence of source credibility on communication effectiveness. Public Opin Q 1951;4:635-650. [doi: 10.1086/266350]
63. Hovland CI, Janis IL, Kelly HH. Communication and Persuasion: Psychological Studies of Opinion Change. New Haven, CT: Yale University Press; 1953.
64. Nikolaus J. Credibility effects. In: Donsbach W, editor. Blackwell International Encyclopedia of Communication. Oxford: Wiley; 2008:1044-1047.
65. Google Developers. Google Custom Search API URL: https://developers.google.com/custom-search/ [accessed 2013-03-10] [WebCite Cache ID 6F0Z3bdiF]
66. Google Developers. Google custom search rendering URL: https://developers.google.com/custom-search/docs/js/rendering [accessed 2013-03-14] [WebCite Cache ID 6F7Mya9Sh]
67. Tinsley HE, Tinsley T. Uses of factor analysis in counseling psychology research. Journal of Counseling Psychology 1987;34(4):414-424. [doi: 10.1037/0022-0167.34.4.414]
68. Lorence DP, Greenberg L. The zeitgeist of online health search. Implications for a consumer-centric health system. J Gen Intern Med 2006 Feb;21(2):134-139 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.0295.x] [Medline: 16336621]

Abbreviations

AJAX: asynchronous JavaScript
API: application programming interface
HIT: human intelligence task
HON: Health on the Net
KMO: Kaiser-Meyer-Olkin
MTurk: Mechanical Turk
RRSA: Research Readiness Self-Assessment
VIF: variance inflation factor
XML: extensible markup language

Edited by G Eysenbach; submitted 29.03.13; peer-reviewed by S Rubinelli, N Allen; comments to author 11.06.13; revised version received 23.07.13; accepted 18.01.14; published 02.04.14.

Please cite as:
Allam A, Schulz PJ, Nakamoto K
The Impact of Search Engine Selection and Sorting Criteria on Vaccination Beliefs and Attitudes: Two Experiments Manipulating Google Output
J Med Internet Res 2014;16(4):e100
URL: http://www.jmir.org/2014/4/e100/
doi:10.2196/jmir.2642
PMID:24694866

Viewpoint

Is Biblioleaks Inevitable?

Adam G Dunn[1], PhD; Enrico Coiera[1], MBBS, PhD; Kenneth D Mandl[2,3], MD, MPH

[1]Centre for Health Informatics, Australian Institute of Health Innovation, The University of New South Wales, Sydney, Australia

[2]Center for Biomedical Informatics, Harvard Medical School, Boston, MA, United States

[3]Children's Hospital Informatics Program, Harvard-MIT Health Sciences and Technology, Boston Children's Hospital, Boston, MA, United States

Corresponding Author:
Adam G Dunn, PhD
Centre for Health Informatics
Australian Institute of Health Innovation
The University of New South Wales
Sydney, 2052
Australia
Phone: 61 9385 8699
Fax: 61 9385 8692
Email: a.dunn@unsw.edu.au

Abstract

In 2014, the vast majority of published biomedical research is still hidden behind paywalls rather than open access. For more than a decade, similar restrictions over other digitally available content have engendered illegal activity. Music file sharing became rampant in the late 1990s as communities formed around new ways to share. The frequency and scale of cyber-attacks against commercial and government interests has increased dramatically. Massive troves of classified government documents have become public through the actions of a few. Yet we have not seen significant growth in the illegal sharing of peer-reviewed academic articles. Should we truly expect that biomedical publishing is somehow at less risk than other content-generating industries? What of the larger threat—a "Biblioleaks" event—a database breach and public leak of the substantial archives of biomedical literature? As the expectation that all research should be available to everyone becomes the norm for a younger generation of researchers and the broader community, the motivations for such a leak are likely to grow. We explore the feasibility and consequences of a Biblioleaks event for researchers, journals, publishers, and the broader communities of doctors and the patients they serve.

(J Med Internet Res 2014;16(4):e112) doi:10.2196/jmir.3331

KEYWORDS

bibliographic databases; compromising of data; open access; public access to information; peer-to-peer architectures

The Hypothetical Biblioleaks Scenario

Through a concerted effort, hackers gain access to the databases of six publishers that together control access to the majority of subscription-based biomedical journal articles. This group makes copies of every article from every journal and releases them into the public domain. Subsets of articles are mirrored in anonymous peer-to-peer networks, creating a decentralized and multiply-redundant repository that is accessible to any human or computer algorithm. The repository grows when its users begin to add new and missing articles, creating a self-sustaining system of frictionless, free, and universal access to published research. While there would be recourse against offenders and while the wider academic community may be unlikely to embrace illicit activity, a robust international article-sharing underground is created. Academics in wealthy countries generally enjoy the privilege of institutional subscriptions to many journals, but articles that require payment to read or download (paywalled) are largely beyond the reach of everyone else and there is a substantial motivation to access this new resource.

The potential for this form of guerrilla open access is rarely discussed [1], despite the massive scale of recent cyber-attacks against commercial and government interests. Large-scale events, like the rise of illegal music file sharing on Napster and the massive releases of government documents including The War Logs and global surveillance disclosures, can force these issues of access and transparency into the mainstream public debate.

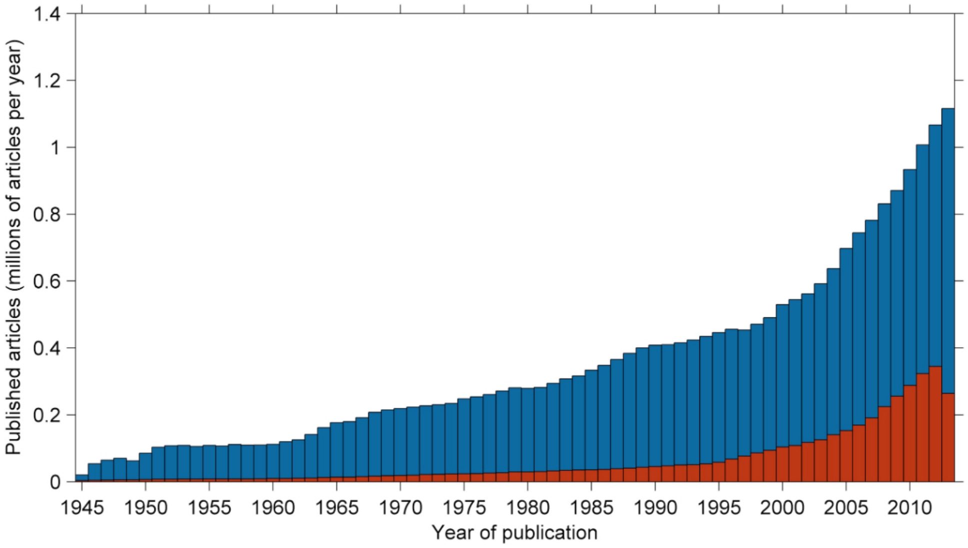
The likely consequences of such an event for publishers, journals, researchers, and the wider community are largely unknown. Speculating on the consequences of a leak and the plausibility of a decentralized article-sharing underground, we consider the current behaviors of the producers and consumers of biomedical research, the sizes and forms of recent data breaches, and the technologies underpinning anonymous sharing.

Current Public Access

Of the 23.6 million articles currently indexed by PubMed (a search engine that accesses the MEDLINE database of life science and biomedical literature), the full text versions of just over 3 million are available for free via PubMed Central (Figure 1). This means that today around 13% of peer-reviewed biomedical articles currently indexed by PubMed are directly available for free via PubMed Central. Although open access publishing is growing rapidly (a 16-fold growth between 2000 and 2011 [2]), the overall volume of publishing in biomedical research appears to be outpacing the volume of growth in open access, creating a persistent archive of potentially inaccessible biomedical research.

PubMed indexes the bulk of all biomedical research that meets a minimum standard of quality but PubMed Central does not capture all of the articles that can be accessed for free. Other access options include library or personal journal subscriptions, emailing authors, a series of balkanized repositories like research-based social networks and institutional webpages [3], and paying publishers for access to individual articles. This process for gaining access to the full text of paywalled articles is inefficient even for experts actively engaged in research, but its most severe effects are likely felt by the groups that have fewer options for access—clinicians and the broader public.

Figure 1. Volumes of articles currently indexed by PubMed (blue) and volumes freely available via PubMed Central (orange), arranged by year of publication, for articles published between 1945 and 2013 (data accessed 17 March 2014).



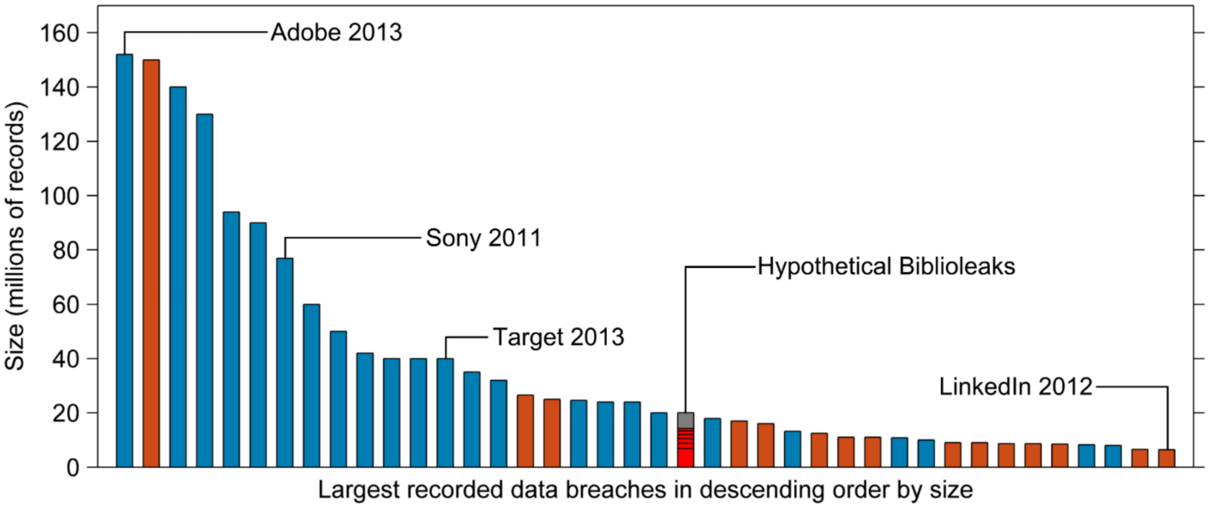
Feasibility

The motivation behind the ethos that “all information should be free” has been explicitly built into the fabric of academia for at least 50 years [4]. Where the costs of accessing published research are unaffordable, the drivers to subvert access restriction seem no different to those driving recent large scale data breaches. In a time where once secret or restricted information is increasingly leaked in the public interest, and at least one advocate has publicly promoted guerrilla open access in peer-reviewed literature [1], publishers should address the threat of a massive data breach.

Scale is no barrier to a cyber-attack. From a database of over 7600 recorded data breaches [5], 21 involved over 20 million records each, indicating clear and recent precedents for a data breach of this scale (Figure 2). Among the 21 largest, hackers were responsible for 18, and most of these were in the last five years. From these records, it is clear that large businesses are not immune to data breaches and that large data breaches are increasing unabated.

Once released into the public domain, articles may be difficult or impossible to recover because there are no technical barriers to leaking published research once it has been acquired. The software used for cleaning documents and anonymously disseminating them online are available [6,7]. The peer-to-peer network structures that could be used to store, track, and provide access to the leaked articles became mainstream with Napster in the late 1990s [8,9]. While publishers are currently involved in issuing take-down notices to authors and institutions that release their own articles in contravention of licenses, this strategy for enforcing copyright ownership could not be used if articles were leaked anonymously online. New forms of peer-to-peer networks also resist this form of censorship through the privacy and security of darknet structures [10,11], and by using distributed storage, where files are split into encrypted chunks so that all users have access but no individual user stores an entire file [11].

Figure 2. Largest recorded data breaches by number of records (accessed 7 January 2014). Hacks are in blue, all other breach types in orange (eg, stolen/lost disks)–compared to a hypothetical breach equivalent to the numbers of articles indexed by PubMed for which full-text versions require a subscription or payment to access. The proportions associated with the 6 largest publishers (sampled from outgoing PubMed links on 7 January 2014) make up 72% of these inaccessible articles (in red).



Publishers

The publishing industry has experimented with a variety of open access models [12,13]. Traditional publishers have embraced gold, green, and hybrid open access models for both new and existing journals. In gold open access, authors typically pay to publish at the time of publication and articles are then free to access. This does not necessarily mean that the articles are released without restrictions on their use, however. In green open access, authors are permitted to upload some version of the article to a repository, sometimes after a delay. Traditional publishers continue to be exceedingly profitable even when the majority of their articles are released under green open access licenses [14], and new predatory publishers have also flourished in an ever-expanding market [15,16].

The commercial risk of a massive data leak would be skewed toward traditional publishers with business models involving charging for access to published research, and journals that rely on traffic through their websites for advertising revenue. Such a scenario may also affect the community structure in some disciplines by redirecting readership away from journal websites, reducing interaction within societies [17]. In a Biblioleaks scenario, open access publishers that receive the bulk of their income at the point of publication and do not rely on subscriptions or advertising revenue would therefore be at an advantage relative to other publishers.

Authors

The license given by subscription journals to authors typically allows them to share their published work by uploading versions of their manuscripts to government and academic institutional repositories, or personal webpages. Known as green open access, around 81% of published articles fall into to this category (62% immediately after publication, the rest after varying delays) [2]. Despite the growing requirement from public research funding bodies that research be open access, only around 12% of green open access articles can be found by searching repositories or personal webpages because authors have not availed themselves of this option [2]. What this means is that despite the ability and obligation to do so, the rate of self-archiving by researchers is low.

In contrast to the extraordinary expansion of the Napster music sharing network in the late 1990s, relatively few researchers are involved in guerrilla open access—sharing articles in contravention of the conditions of a publisher. Two sharing practices have recently emerged on Twitter. The first, #icanhazpdf [18], started in 2011 and is a “pull” practice, where those who need articles request them and wait for someone with access to provide a copy. The second, #pdftribute [19], started in 2013 as a “push” practice, where authors advertised an online location for full versions of articles they wrote or held. Both practices stabilized at a low rate of requests and uploads. From this evidence, and given the low levels of observed self-archiving and civil disobedience from within the academic community, we speculate that a disruptive change is more likely to come from a Biblioleaks scenario—a small number of massive breaches, potentially from outside academia, rather than *en masse* civil disobedience from within academic communities.

A massive leak would appear to have few or no direct negative consequences for individual researchers. The major impact would be increased access to their published works. Overall, the capacity to better link and synthesize research could also lead to improvements in automated literature discovery [20,21], expanding opportunities for new forms of research. With passive roles in the Biblioleaks scenario, the interests of authors of peer-reviewed research would be served, with no directly associated risks.

Clinicians and Patients

While access to published literature is problematic for researchers inside universities, the broader community faces formidable barriers. There is some evidence to suggest that the public want access to published research and are hampered by paywalls [22]. Evidence showing that open access articles are viewed and downloaded more often without necessarily leading to higher citation rates is a further hint that the wider community is engaging with peer-reviewed research [23].

In the medical context, we know that health care workers are less likely to read peer-reviewed literature than they are to ask colleagues, use reference books, or visit websites via Google or Wikipedia [24]. Among patients in the United States, 58% looked online for health information and one in four encountered a paywall [25]. The evidence suggests that clinicians and the public try to use the Internet to find literature but are often unable to reach what they need. With no paid institutional access, and without the personal networks to help circumvent access restrictions, the average member of the community is more likely to ignore inaccessible articles and rely on sources that are not peer-reviewed. As such, a massive leak has every chance of creating a more informed clinical and patient community once they become comfortable with accessing such a repository.

At the population level, the gap between research consensus and public understanding has major consequences for global health, where for issues including vaccination, homoeopathy, and climate change, there is a clear dissonance between what peer-reviewed evidence shows and what large sections of the public believe [26-28]. While prior beliefs feature heavily in decision making, the first document accessed in a search plays an important role in the potential to switch beliefs [29]. By removing the barriers that restrict access to most peer-reviewed literature, a massive leak could help to reduce problematic public opinions by providing greater transparency and shifting the weight of available information away from grey literature and toward peer-reviewed research.

Futures

Even as open access increases, the motivations for a massive leak will persist because the archive of inaccessible research continues to grow. The threats are clear. Chelsea Manning and Edward Snowden demonstrated that individuals can bring about tectonic shifts in the ability of government to maintain secrecy and the public attitudes toward clandestine programs, although at great personal cost to the leakers themselves. The entertainment industry navigated similar terrain over a decade ago, when consumers moved from recording onto cassettes from the radio to sharing on local and then global computer networks. That disruption left a legacy of file-sharing networks that have become increasingly secure and resistant to censorship. Today, instantaneous access to music, television, and movies is taken for granted by many, while the entertainment industry continues to flourish by finding legitimate and low-cost ways to reach audiences that would otherwise turn to illegal file-sharing services.

From this view, biomedical publishing faces threats, but also opportunities. The current forms of illegitimate sharing in academia rely largely on personal networks or easily censored websites. Relatively few academics have started to explore broader forms of civil disobedience. Since open access has become a mainstream issue, academics and the public are beginning to expect free and immediate access to new research as the norm and not the exception. As forms of illegitimate sharing become more sophisticated and widespread, publishers face a situation reminiscent of the one faced by the entertainment industry more than a decade ago.

We think that low-level civil disobedience (or authors unaware of which versions of their articles they are allowed to upload to repositories) is by itself unlikely to lead to a critical mass of illegal article sharing. Large-scale leaks are a bigger threat because they could immediately influence the way published research is accessed. For this reason, publishers might see value in strengthening the systems already in place to detect and prevent unusually large volume downloads, or atypical systematic or ordered access to full texts.

Prescient publishers may also consider alternatives that would minimize the motivation behind any illegal access and avoid the costs of a technical and legal arms race that may only delay the inevitable. Publishers may choose to deliberately release articles on their own terms, an approach that improved the reputation of and trust in GlaxoSmithKline when they responded to growing demand for access to comprehensive clinical trial data [30]. They might also consider alternative forms of low-cost access that could greatly expand the market for peer-reviewed research into the broader community. Examples of new forms of low-cost access, such as time-limited rentals, are already available [31].

From the limited evidence available in this area, it seems clear that a Biblioleaks event is technically feasible. There is some evidence that new forms of illegal file sharing are emerging among researchers and the broader community, suggesting that the current environment is similar to the nascent period of illegal file sharing. In that time, online users increasingly encountered the tools that provided free access to music, and fragmented communities began to coalesce into a global sharing network. If precipitated by targeted data breaches, a similar growth in underground article sharing could see negative effects for some publishers, disruptive changes to the way biomedical research is accessed by the public, the rapid development of new low-cost access options, and improved public engagement with medical research.

Acknowledgments

No specific funding was received in relation to this research.

Conflicts of Interest

None declared.

References

1. Swartz A. Guerilla Open Access Manifesto. 2008. URL: https://ia700808.us.archive.org/17/items/GuerillaOpenAccessManifesto/Goamjuly2008.pdf [accessed 2014-02-17] [WebCite Cache ID 6P3J9evVH]
2. Björk B, Laakso M, Welling P, Paetau P. Anatomy of green open access. J Assn Inf Sci Tec 2013 Nov 06;65(2):237-250. [doi: 10.1002/asi.22963]
3. Laakso M, Björk BC. Anatomy of open access publishing: a study of longitudinal development and internal structure. BMC Med 2012;10:124 [FREE Full text] [doi: 10.1186/1741-7015-10-124] [Medline: 23088823]
4. Levy S. Hackers: heroes of the computer revolution. Garden City, N.Y: Anchor Press/Doubleday; 1984.
5. Open Security Foundation. Data Loss DB – Largest Incidents. 2013. URL: http://datalossdb.org/index/largest [accessed 2014-02-17] [WebCite Cache ID 6NQcmoASL]
6. Dingledine R, Mathewson N, Syverson P. Tor: the second-generation onion router. 2004 Presented at: Proceedings of the 13th Conference on USENIX Security Symposium; 2004; San Diego, CA p. 21.
7. Ramakrishnan C, Patnia A, Hovy E, Burns GA. Layout-aware text extraction from full-text PDF of scientific articles. Source Code Biol Med 2012;7(1):7 [FREE Full text] [doi: 10.1186/1751-0473-7-7] [Medline: 22640904]
8. DeVoss DN, Porter JE. Why Napster matters to writing: filesharing as a new ethic of digital delivery. Computers and Composition 2006 Jan;23(2):178-210. [doi: 10.1016/j.compcom.2006.02.001]
9. Saroiu S, Gummadi KP, Gribble SD. Measuring and analyzing the characteristics of Napster and Gnutella hosts. Multimedia Systems 2003 Aug 1;9(2):170-184. [doi: 10.1007/s00530-003-0088-1]
10. Biddle P, England P, Peinado M, Willman B. The darknet and the future of content protection. Digital Rights Management, Lecture Notes in Computer Science 2003;2696:155-176. [doi: 10.1007/10941270_23]
11. Clarke I, Miller S, Hong T, Sandberg O, Wiley B. Protecting free expression online with Freenet. IEEE Internet Comput 2002;6(1):40-49. [doi: 10.1109/4236.978368]
12. Guédon J. The "green" and "gold" roads to open access: the case for mixing and matching. Serials Review 2004 Jan;30(4):315-328. [doi: 10.1016/j.serrev.2004.09.005]
13. Kennedy D. Government policies and the cost of doing research. Science 1985 Feb 1;227(4686):480-484. [Medline: 3966157]
14. The Economist. Scientific publishing: the price of information URL: http://www.economist.com/node/21545974 [accessed 2014-02-17] [WebCite Cache ID 6NQdsgYX6]
15. Bohannon J. Who's afraid of peer review? Science 2013 Oct 4;342(6154):60-65. [doi: 10.1126/science.342.6154.60] [Medline: 24092725]
16. Haug C. The downside of open-access publishing. N Engl J Med 2013 Feb 28;368(9):791-793. [doi: 10.1056/NEJMp1214750] [Medline: 23445091]
17. Davis PM. Public accessibility of biomedical articles from PubMed Central reduces journal readership--retrospective cohort analysis. FASEB J 2013 Jul;27(7):2536-2541 [FREE Full text] [doi: 10.1096/fj.13-229922] [Medline: 23554455]
18. Kuszewski A. @AndreaKuszewski. 2011 Jan 21. OMG, that should be the new “I'm requesting a paper” hashtag! #icanhazPDF RT @BoraZ: I can haz PDF? http://bit.ly/eCsfKo URL: https://twitter.com/AndreaKuszewski/status/28257118322688000 [accessed 2014-02-17] [WebCite Cache ID 6NQfKvez6]
19. Vivalt E. @EvaVivalt. 2013 Jan 13. @csik In tribute, I'll put all pdfs collected for research online. I encourage everyone to do the same. http://aideconomics.com/index.php/2013/01/12/in-memoriam/ #pdftribute URL: https://twitter.com/evavivalt/status/290287659681726464 [accessed 2014-02-17] [WebCite Cache ID 6NQfTXM01]
20. Rzhetsky A, Seringhaus M, Gerstein M. Seeking a new biology through text mining. Cell 2008 Jul 11;134(1):9-13 [FREE Full text] [doi: 10.1016/j.cell.2008.06.029] [Medline: 18614002]
21. Wren JD, Bekeredjian R, Stewart JA, Shohet RV, Garner HR. Knowledge discovery by automated identification and ranking of implicit relationships. Bioinformatics 2004 Feb 12;20(3):389-398 [FREE Full text] [doi: 10.1093/bioinformatics/btg421] [Medline: 14960466]
22. Davis PM, Walters WH. The impact of free access to the scientific literature: a review of recent research. J Med Libr Assoc 2011 Jul;99(3):208-217 [FREE Full text] [doi: 10.3163/1536-5050.99.3.008] [Medline: 21753913]
23. Davis PM. Open access, readership, citations: a randomized controlled trial of scientific journal publishing. FASEB J 2011 Jul;25(7):2129-2134 [FREE Full text] [doi: 10.1096/fj.11-183988] [Medline: 21450907]
24. O'Keeffe J, Willinsky J, Maggio L. Public access and use of health research: an exploratory study of the National Institutes of Health (NIH) Public Access Policy using interviews and surveys of health personnel. J Med Internet Res 2011;13(4):e97 [FREE Full text] [doi: 10.2196/jmir.1827] [Medline: 22106169]
25. Fox S, Duggan M. Health Online.: Pew Internet & American Life Project; 2013. URL: http://www.pewinternet.org/files/old-media//Files/Reports/PIP_HealthOnline.pdf [accessed 2014-02-17] [WebCite Cache ID 6NQh134O7]

26. Kahan DM, Peters E, Wittlin M, Slovic P, Ouellette LL, Braman D, et al. The polarizing impact of science literacy and numeracy on perceived climate change risks. Nature Climate change 2012 May 27;2(10):732-735. [doi: 10.1038/nclimate1547]
27. Maurer J, Uscher-Pines L, Harris KM. Perceived seriousness of seasonal and A(H1N1) influenzas, attitudes toward vaccination, and vaccine uptake among U.S. adults: does the source of information matter? Prev Med 2010 Aug;51(2):185-187. [doi: 10.1016/j.ypmed.2010.05.008] [Medline: 20510270]
28. Shaw DM. Homeopathy is where the harm is: five unethical effects of funding unscientific 'remedies'. J Med Ethics 2010 Mar;36(3):130-131. [doi: 10.1136/jme.2009.034959] [Medline: 20211989]
29. Lau AY, Coiera EW. A Bayesian model that predicts the impact of Web searching on decision making. J. Am. Soc. Inf. Sci 2006 May;57(7):873-880. [doi: 10.1002/asi.20377]
30. Coombes R. GlaxoSmithKline grants researchers access to clinical trial data. BMJ 2012;345:e6909. [Medline: 23065357]
31. Baynes G, Hanson M. Trialing mobile and article rental access options for journal content. The Serials Librarian 2012 Jan;62(1-4):217-221. [doi: 10.1080/0361526X.2012.652924]

Edited by G Eysenbach; submitted 16.02.14; peer-reviewed by I Oransky, P Kudlow; comments to author 14.03.14; revised version received 03.04.14; accepted 13.04.14; published 22.04.14.

Please cite as:
Dunn AG, Coiera E, Mandl KD
Is Biblioleaks Inevitable?
J Med Internet Res 2014;16(4):e112
URL: http://www.jmir.org/2014/4/e112/
doi:10.2196/jmir.3331
PMID:24755534

Original Paper

Executive Functioning in Alcoholics Following an mHealth Cognitive Stimulation Program: Randomized Controlled Trial

Pedro Gamito[1,2], PhD, DSc; Jorge Oliveira[1,2], PhD; Paulo Lopes[1,2], PhD; Rodrigo Brito[1,2], PhD; Diogo Morais[1,2], MSc; Diana Silva[1], MSc; Ana Silva[1], MSc; Sara Rebelo[1], MSc; Marta Bastos[1], MSc; Alberto Deus[3], PhD

[1]Lusophone University of Humanities and Technologies, Lisbon, Portugal
[2]COPELABS-Cognition and People-Centric Computing Labs, Lisbon, Portugal
[3]São João de Deus Institute, Sintra, Portugal

Corresponding Author:
Pedro Gamito, PhD, DSc
Lusophone University of Humanities and Technologies
Campo Grande, 376
Lisbon, 1749 - 024
Portugal
Phone: 351 217 515 500
Fax: 351 21 757 7006
Email: pedro.gamito@ulusofona.pt

Abstract

Background: The consequences of alcohol dependence are severe and may range from physical disease to neuropsychological deficits in several cognitive domains. Alcohol abuse has also been related to brain dysfunction specifically in the prefrontal cortex. Conventional neuropsychological interventions (paper-and-pencil cognitive stimulation training) have a positive effect but are time-consuming, costly, and not motivating for patients.

Objective: Our goal was to test the cognitive effects of a novel approach to neuropsychological intervention, using mobile technology and serious games, on patients with alcohol dependence.

Methods: The trial design consisted of a two-arm study assessing the cognitive outcomes of neuropsychological intervention with mobile serious games (mHealth) versus control (treatment-as-usual with no neuropsychological intervention) in patients undergoing treatment for alcohol dependence syndrome. Sixty-eight patients were recruited from an alcohol-rehab clinic and randomly assigned to the mHealth (n=33) or control condition (n=35). The intervention on the experimental group consisted of a therapist-assisted cognitive stimulation therapy for 4 weeks on a 2-3 days/week basis.

Results: Fourteen patients dropped out of the study. The results of the neuropsychological assessments with the remaining 54 patients showed an overall increase ($P<.05$) of general cognitive abilities, mental flexibility, psychomotor processing speed, and attentional ability in both experimental (n=26) and control groups (n=28). However, there was a more pronounced improvement ($P=.01$) specifically in frontal lobe functions from baseline (mean 13.89, SE 0.58) to follow-up (mean 15.50, SE 0.46) in the experimental group but not in the control group.

Conclusions: The overall increase in general cognitive function for both experimental and control groups supports the beneficial role of existing alcohol treatment protocols aimed at minimizing withdrawal symptoms, but the differential improvements observed in frontal lobe functioning supports the use of mobile serious games for neuropsychological stimulation to overcome executive dysfunction in patients with alcohol dependence. This trial was negative on two neuropsychological/cognitive tests, and positive on one.

Trial Registration: ClinicalTrials.gov NCT01942954; http://www.clinicaltrials.gov/ct2/show/NCT01942954 (Archived by WebCite at http://www.webcitation.org/6OYDqHLwB).

(J Med Internet Res 2014;16(4):e102) doi:10.2196/jmir.2923

KEYWORDS

alcohol; addiction; cognitive stimulation; executive function; mobile health; serious games

Introduction

Alcohol abuse is a global health problem and cause of excess mortality and morbidity, as well as a source of personal disruption for both abusers and their families. A representative survey-based study of the US population found that excessive consumption of alcohol increased risk of mortality over a 14-year period [1]. A study in Northern Germany concluded that alcohol dependency (as assessed by the Diagnostic and Statistical Manual of Mental Disorders, 4th edition [DSM-IV]) increased the mortality rate almost two-fold for men and more than four-fold for women [2].

Apart from excess mortality, the negative effects of alcohol on the brain can also be severe. Alcohol dependence syndrome has been related to brain dysfunction specifically in the prefrontal cortex, which is associated with cognitive functioning [3]. In fact, the available evidence suggests that alcohol abuse produces a decrease in specific cognitive abilities, particularly those associated with executive functions [4]. Many studies have found a positive correlation between the integrity of white matter and cognitive performance in teenagers and a negative effect of prolonged alcohol dependence on the integrity of white matter in adults (eg, [5]). This suggests that alcohol use is related to the reduction of the integrity of grey matter, as the integrity of grey matter depends on functional white matter, in particular, the superior longitudinal fasciculus. Also, several studies with general population samples indicate that the prefrontal cortex is particularly vulnerable to the neurotoxic effects of alcohol [6]. A recent review of 62 papers assessing cognitive dysfunction in alcoholics concluded that (1) several cognitive functions become significantly compromised, and remain so, even during abstinence, with stable effects for 1 year of abstinence, (2) without intervention, these effects start reversing after 1 year of abstinence, (3) they appear to support the diffuse cerebral hypothesis, but (4) it is still unclear which cognitive functions are more susceptible to be compromised, as well as what the necessary timeframe is for cognitive recovery [7]. What is known is that there is a relation between the excessive consumption of alcohol and the compromising of cognitive functions in a variety of domains, namely, attention, working memory, processing speed, visual-spatial capacities, impulsivity, learning, memory, verbal fluency, decision making, and executive functions [8-14]. One study showed that, even when sober, binge drinking (BDs) university students performed significantly worse on planning tasks (they took significantly longer to plan an action) and on attention/working memory tasks (they got fewer correct answers on the Paced Auditory Serial Addition Task [PASAT]) than university students who did not binge drink (non-BDs) [15]. Another study found that BDs showed a lower capacity to retain and manipulate information in the verbal working memory (Digit Span Backward, Wechsler Memory Scale [WMS-III]), as well as a greater number of persevering responses (although only on the speed of processing test [SOPT], and not on the Wisconsin Card Sorting Test [WCST]) than non-BDs [16], indicating a lower capacity on executive tasks, which depend on the functional integrity of the prefrontal cortex. Another study also found that BD university students had worse results on perseverative and intrusion errors and more false positives (Complutense Verbal Learning Test [TAVEC]), as well as greater interference (Stroop tests) and worse performance in the Digit Span (WMS-III), Corsi Block Test, and the Series Recall Test [17].

In sum, patients with alcohol dependence are clearly an important target population for neuropsychological rehabilitation programs in general and cognitive rehabilitation programs in particular. However, the literature on the effectiveness of neuropsychological rehabilitation in alcohol addicts is still scarce. A few years ago, a review of cognitive rehabilitation programs with alcoholics noted that although patients' responses to cognitive rehabilitation techniques were generally satisfactory, those techniques had not been used in traditional treatment programs [18]. Subsequent studies have shown that training and stimulating cognitive functions that have been compromised by the effects of drugs have some positive effects. For example, a study of 120 drug addicts in treatment found that patients benefiting from cognitive rehabilitation remained in treatment for significantly longer periods, and 38% of this group ended treatment successfully, compared to only 18% of other groups [19]. Another study with 60 drug addicts in in-residence treatment found that patients undergoing cognitive rehabilitation treatment became significantly more committed to the treatment, assimilated therapeutic content better, remained in abstinence for longer periods after ending the treatment, and had better scores on social and family behavioral normalization, as well as greater reduction of legal problems, than those who did not undergo treatment [20]. In addition, a study with 40 alcoholics found that a group subjected to cognitive rehabilitation significantly improved their level of information processing, visual-constructive abilities, and decision-making process compared to a group that did not undergo cognitive rehabilitation [21]. A recent review of all available studies on cognitive rehabilitation (CR) related to addiction concluded that there is a clear tendency for improvement of success rates in the treatment of patients that are targeted by specifically cognitive rehabilitation programs [22].

Despite the apparent added value of CR, the large majority of programs are still based on pencil-and-paper tasks with questionable ecological validity for functionality in daily life. Serious games (SG), that is, games designed for other purposes than gaming, seem to be a sound way to overcome this flaw by simulating real life activities or by simply challenging patients' cognitive functions through an interactive and appealing interface. Several have already been designed and applied to stroke and traumatic brain injury rehabilitation (eg, [23,24]). SGs are usually platforms that encompass training environments where repetition and visual and auditory feedback can be systematically manipulated according to individual specificities. And because these are games, patients are usually motivated to execute the proposed exercises. Thus, SGs combine three important aspects that may contribute to the effectiveness of this approach: repetition, feedback, and motivation. Repetition refers to an intrinsic characteristic of games, which is the possibility of repeating over and over the same action in a pleasing way. Feedback refers to the fact that, while carrying out exercises, the patient's senses are provided with feedback on the accomplishments achieved during each task. Finally, motivation is probably enhanced in SG because they are usually presented on a multimodal platform with different immersive cues, such as images and sounds, where patients may be more willing to engage and pursue an exercise.

The increasing development of both hardware and software has allowed the use of mobile devices in cognitive rehabilitation programs [23]. It is now possible to design, develop, and apply training and treatment programs over mobile health (mHealth) applications. Mobile device-based interventions are already used, for example, to enhance emotional awareness [25] or to treat emotional disorders such as depression [26] or anxiety [27], as self-help programs to reduce cocaine consumption [28], or to support patients with mild acquired cognitive impairments [29]. There are also some studies on alcohol abusers that reflect this shift of attention. For example, Hester et al developed a Web-based application so that alcoholics could overcome their drinking problems [30]. They found that patients using their program significantly decreased their average number of drinks per drinking day and alcohol-related problems, and also increased their average number of abstinence days. The application, however, did not produce better results than the traditional approach.

Another study compared the outcomes of two types of Web-based interventions on 170 problem drinkers. In one of the interventions, Check Your Drinking (CYD) [31], screened participants were invited to answer questions about the quantity and frequency of their drinking and with the severity of their drinking problems. In the other, the Alcohol Help Center (AHC) [32], participants had to undergo several exercises based on cognitive and behavioral principles, designed to increase motivation levels and prevent relapse. A significant additional reduction in drinking for participants in the AHC group was found when compared with the CYD group [33].

However, for addiction-related cognitive impairments, research assessing the effect of cognitive stimulation using mHealth SG programs for the improvement of cognitive functioning is still in its infancy. The objective of the current study was to assess the outcome of cognitive stimulation exercises over a Web platform developed to tackle cognitive impairments of individuals with Alcohol Dependence Syndrome (ADS) [34]. The potential impact of using such an approach relies on the possibility of democratizing the access and the usage of cognitive stimulation programs throughout the health care system (ie, both through caregivers and directly to patients), thus also reducing the societal burden associated with the costs of cognitive impairments and treatments.

Methods

Trial Design

The study design consisted of a two-arm randomized controlled trial (RCT) developed to assess the neuropsychological effects of mHealth applications in alcoholics. (The trial was registered retrospectively because our funding agency, the Portuguese National Science Foundation [FCT] does not require registration for psychological tests nor does our home country have a registration system.) Our sample size was estimated based on a priori power analysis according to commonly accepted standards in the field (see Statistical Procedures) and agreed to in advance with the clinical institution where the trial took place.

The method of patient assignment was based on simple randomization with a random number generator. The patients were randomly assigned to experimental and control groups in a specialized institution for treatment of alcohol dependence (see Multimedia Appendix 1 for the CONSORT eHealth Checklist). The trial was approved by the Ethics Committee of the research center where the authors were affiliated at the time (Centre for the Study of Cognitive and Learning Psychology [CEPCA]) and adhered to the principles of the World Medical Association’s Declaration of Helsinki.

Participants

Sixty-eight patients diagnosed with ADS according to DSM-IV criteria were recruited from a specialized institution for treatment of alcohol dependence, the Novo Rumo Clinic, São João de Deus Institute in the Lisbon region, Portugal, and were asked to participate in a study on the effects of their treatment on cognitive abilities. In the experimental condition, they were told that their treatment would include cognitive exercises. Due to a dropout rate of 20.6% (14/68 patients), the final sample consisted of 54 patients.

Inclusion Criteria

Only patients that scored higher than the cut-off values for their age (see Outcomes) on the Mini Mental State Examination (MMSE) [35] and with no clinical scores on the Symptoms Checklist Revised (SCL-90-R) [36] were included in the study.

Patients continued during the entire study their regular medication regimen consisting of anxiolytics, mostly Diazepam and Tiapride, which help minimize withdrawal symptoms, and vitamins. Each patient’s assistant psychiatrist guaranteed the stability of the medication regimens throughout the program.

Exclusion Criteria

Patients with dependency on substances other than alcohol or with a history of previous neurological disorders were excluded from the study. Patients were also screened for minimal computer literacy, but no patient was excluded due to this criterion (ie, all patients demonstrated the minimum literacy required).

Study Procedure

After initial recruitment and screening, participants were randomly assigned to either the experimental group with mHealth SG-based cognitive stimulation plus treatment-as-usual (n=33) or to the control group (n=35), which received solely treatment-as-usual. The treatment-as-usual consisted of an alcohol-abstinence program adapted from the Minnesota Model (see [37] for a detailed description of this program) and lasted for an average period of 1 month.

Figure 1 illustrates the flow of participants in the experimental group throughout the protocol. Participants failing to complete training sessions within the assigned time-frame were considered to have dropped out and their data were not analyzed.

The trial took place over a period of 6 months, and no changes to the stimulus program were made during this period. Both the treatment and the assessments took place on location at the clinic where participants were recruited.

Each participant underwent two complete neuropsychological assessments, once after they had undergone screening and given their written informed consent to treatment and assessment, and again after at least 30 days. In the case of the treatment group, this was done after completing the intervention.

The intervention consisted of ten 60-minute sessions of cognitive stimulation with mobile technology using SG (2-3 sessions per week over the usual 4-6 week period of treatment). No institutional affiliations were presented in the eHealth media. The executive training exercises performed by participants in the experimental group were selected in order to develop cognitive abilities related to executive functioning. Each session started with a brief training period, when participants were able to (re)acquire interaction skills with the touchscreen devices. Participants accessed the exercises freely over the Internet, and their responses were registered using the input from the device's touchscreen.

Therapists from the research and intervention team were involved in both the recruitment and the cognitive stimulation. These therapists were introduced to patients by in-house therapists and asked patients to participate in the study, explaining its benefits, duration, and demands on patients' time and commitment. During the assessments, other therapists provided, explained, and collected the assessment forms. In the cognitive simulation sessions, the first group of therapists provided the mobile devices, launched the exercises, and explained how they worked to participants.

The mobile cognitive stimulation program consisted of several mobile applications developed to run on Android OS, which were adapted from the traditional paper-and-pencil exercise conceived for cognitive stimulation on patients with cognitive impairments, independently of the cause, but selected for their relevance to address the most common cognitive impairments in addicts. Cognitive stimulation in each session comprised attention, working memory, and logical reasoning exercises (see Textbox 1 for a more detailed description; these exercises (Figure 2) are available online [34]). The level of difficulty of each task increased progressively throughout the cognitive stimulation rationale. In the last session, the same neuropsychological tests used in the first assessment were again applied.

The hardware used to perform the exercises consisted of Samsung Galaxy 10.1 tablets. The applications were developed using Unity 2.5, and their alpha and beta versions had been previously tested by a group of students.

Textbox 1. Cognitive stimulation program—sessions and mHealth applications. a: Perception; b: Processing speed; c: Reasoning; d: Attention; e: Memory; f: Decision making; g: Planning; h: Spatial vision.

Session 1. Slot[d]/Memory[e]/Parking Zone[g]/Under pressure[h]/Snowflakes[b]

Session 2. Slot[d]/Memory[e]/Under pressure[h]/Snowflakes[b]/Right order[c]

Session 3. Slot[d]/Memory[e]/Hanoi Tower[f]/Snowflakes[b]/Right order[c]

Session 4. Slot[d]/Memory[e/]Odd-even[c]/Parking Zone[g]/Snowflakes[b]

Session 5. Basket[c]/Odd-even[c]/Hand tricks[h]/Brick[a]/Memory[e]

Session 6. Hanoi Tower[f]/Parking Zone[g]/Under pressure[h]/Memory[e]/Snowflakes[b]

Session 7. Parking Zone[g]/Under pressure[h]/Selective transfer[d]/Memory[e]/Snowflakes[b]

Session 8. Selective transfer[d]/Brick[a]/Hand tricks[h]/Memory[e]

Session 9. Parking Zone[g]/Brick[a]/Hand tricks[h]/Memory[e]

Session 10. Slot[d]/Memory[e]/Parking Zone[g]/Under pressure[h]/Snowflakes[b]

Figure 1. Flow chart describing the flow of the participants in the treatment group throughout the protocol.

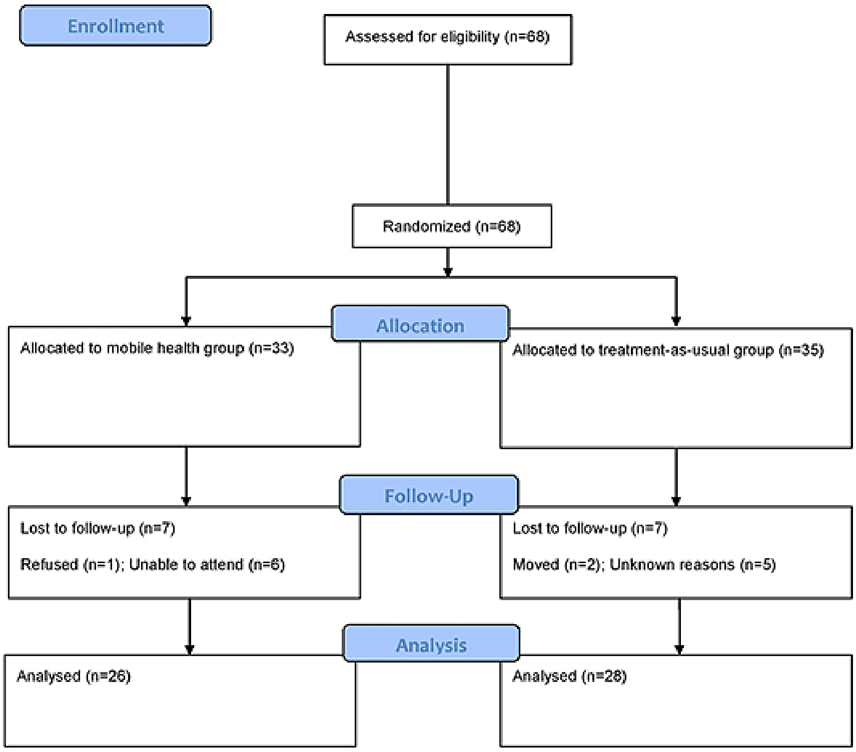
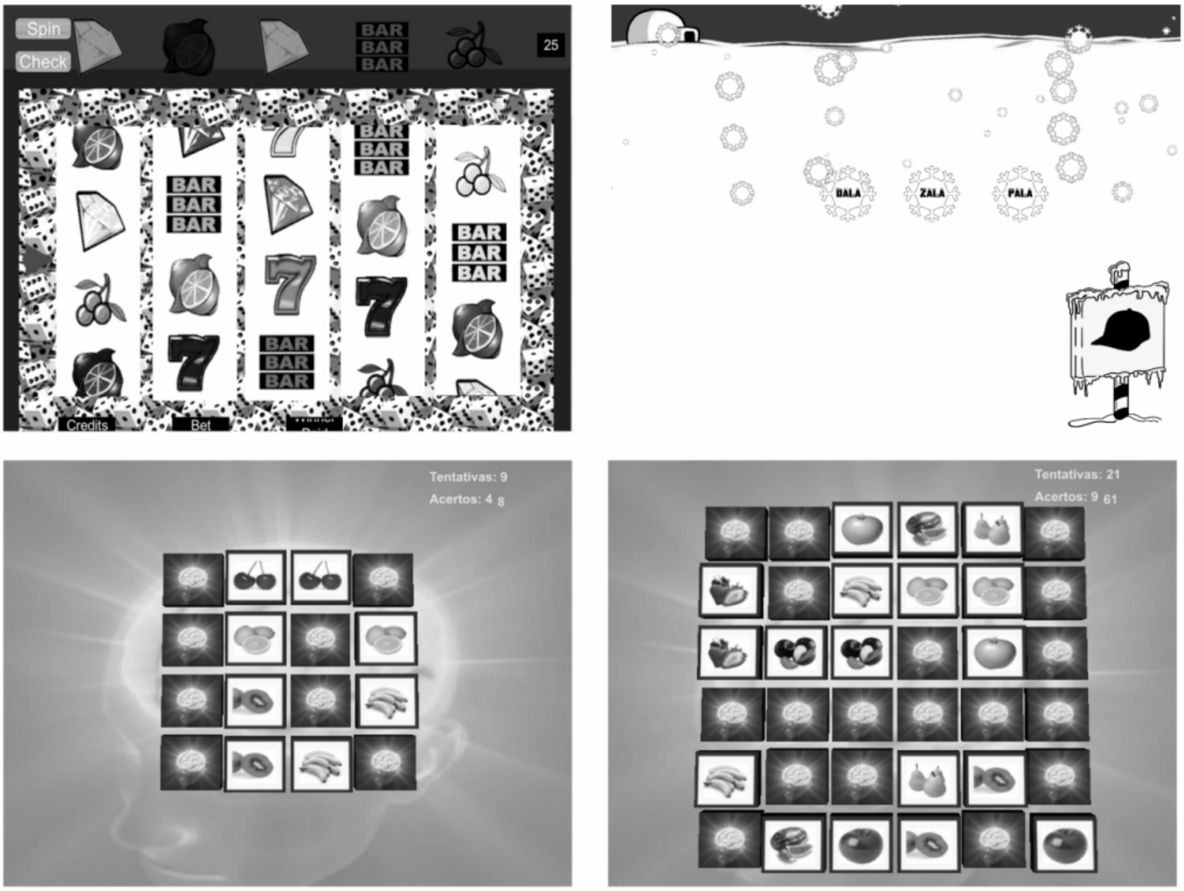


Figure 2. Slot machine (top-left) for attention; visual memory task (bottom-left and bottom-right with increased difficulty) for working memory; and word-object correspondence (top-right) for logical reasoning.



Outcomes

All neuropsychological assessments were carried out with the pencil-and-paper forms of well-established cognitive tests. Global cognitive abilities were assessed with the MMSE [35], a brief screening test that has been validated for the Portuguese population [38]. The MMSE assesses diverse aspects of cognitive mental function, allowing an overall assessment of cognitive performance based on 30 items grouped into 6 categories: Orientation, Retention, Attention and Calculation, Language, and Visual-spatial abilities. The maximum score is 30 points and the current cut-off values for the Portuguese population were estimated according to age and education, specifically for people aged over 40 years: (1) 0-2 years of schooling—22, (2) 3-6 years of schooling—24, and (3) more than 7 years of schooling—27. MMSE was dictated on the basis of education.

Frontal brain functioning was evaluated with the Frontal Assessment Battery (FAB) [39]. The FAB assesses conceptualization, mental flexibility, motor programming, sensitivity to interference, inhibitory control, and environmental autonomy. The total score of the FAB (maximum 18 points) is estimated through the sum of each of the subtest scores. The FAB assesses 6 domains of frontal functioning through the ability to generate similarities (conceptualization), verbal fluency (mental flexibility), Luria's motor series (motor programming), conflicting instructions (sensitivity to interference), go-no go paradigm (inhibitory control), and prehension behavior assessments (environmental autonomy).

Cognitive flexibility was measured with the WCST [40], which evaluates cognitive functions in several executive domains, namely the ability to develop and maintain appropriate strategies for problem solving and planning, and the ability to use environmental feedback to modify a cognitive response. The WCST assesses cognitive abilities such as abstraction, mental flexibility, and sustained attention. We focused on the overall index of performance through correct responses, categories, and perseverative errors. In the current study, we used the short version of the test (WCST-64).

Psychomotor processing speed and attentional abilities were estimated through the Color Trail Test (CTT) [41]. The CTT assesses focused and divided attention, sequencing, mental flexibility, visual search, and motor functions. The CTT consists of an A4 sheet containing circles with numbers printed in yellow and pink. Participants are asked to link the numbered circles in the right order, as fast as possible, and without lifting the pen. The CTT consists of two different forms (CTT1 and CTT2), which differ in difficulty. Two measures of performance were analyzed: errors and execution time.

Statistical Analysis

The variables evaluated during the neuropsychological assessments were submitted to parametric statistical analyses. The dependent variables were based on the neuropsychological outcomes from the MMSE, FAB, WCST, and the CTT. Scores reported refer to before intervention (baseline) and after intervention (follow-up).

Sample and baseline characteristics were compared between the experimental and control groups with Student's *t* test for independent samples for interval dependent variables and chi-square test for categorical dependent variables.

The statistical analyses of evaluation outcomes were based on generalized linear model procedures using type III Sum of Squares. A full factorial model (within-between interaction) was tested using repeated measures analysis of variance (ANOVA). The factorial design consisted of one within-subjects factor with 2 levels (baseline vs follow-up) and one between-subjects factor with 2 levels (experimental vs control group). Missing data imputation was completed using linear interpolation method. A significance level of .05 was adopted for all statistical procedures. Post-hoc effect sizes are eta-square in the ANOVA procedure, and within-between interactions were tested using simple effects.

Inferential statistics were carried out using IBM SPSS v.20. A priori power analysis was estimated with G*Power v.3.1 with Cohen's *f* effect size for *F* tests [42].

Sample Size

The expected effect size for calculating the required sample size was 0.35 (medium effect) based on a power of 0.80 for a significance level of .05. According to these criteria, a total sample size of 68 patients would be required for this trial. This was the initial size of the sample recruited for this study.

Results

Sample Characteristics

Due to a dropout rate of 20.6% (14/68 patients), the final sample consisted of 54 patients diagnosed with alcohol dependence syndrome: 45 males and 9 females, mean age 45.37 years (SD 10.12) with an average of 10 years of formal education (SD 4.62). Of these, 26 patients had been assigned to the mobile health cognitive stimulation condition: 19 males and 7 females, mean age 45.50 years (SD 10.18) with an average of 10 years of education (SD 4.39). Twenty-eight patients were assigned to the control condition consisting of treatment-as-usual: 26 males and 2 females, mean age 45.25 years (SD 10.26) with an average of 10 years of formal education (SD 4.95). No statistically significant differences at the conventional $P<.05$ level were observed between groups for the demographic characteristics of gender (Table 1), age, and education (Table 2).

Table 1. Baseline demographics for categorical data.

Characteristics	mHealth, n (%)	Treatment-as-usual, n (%)	χ^2	*P* value
Gender			3.798	.051
Male	19 (35.2)	26 (48.1)		
Female	7 (13.0)	2 (3.7)		

Table 2. Baseline demographics for interval data.

Characteristics	mHealth, mean (SD)	Treatment-as-usual, mean (SD)	*t*	*P* value
Age, yrs	45.50 (10.18)	45.25 (10.26)	0.090	.929
Education, yrs	10.18 (4.93)	10.25 (4.95)	–0.052	.958

Baseline Characteristics

The statistical analyses for the neuropsychological data at baseline were focused on comparisons between groups for the overall score of the MMSE and the FAB, the correct responses, categories and perseverative errors in the WCST, and the mean errors and execution times in the CTT. The results revealed no statistically significant differences between the experimental and control group for neuropsychological outcomes at baseline (all *P*s>.41) (Table 3).

Table 3. Baseline characteristics for both mHealth and control.

Characteristics	mHealth, mean (SD)	Treatment-as-usual, mean (SD)	*t*	*P* value
MMSE				
Total score	26.46 (2.97)	26.38 (2.67)	0.098	.922
FAB				
Total score	13.89 (2.86)	14.14 (3.05)	–0.320	.750
WCST				
Correct responses	52.85 (20.05)	52.48 (17.01)	0.068	.946
Categories	54.86 (35.60)	50.00 (35.27)	0.485	.630
Perseverative errors	24.78 (15.30)	25.57 (13.56)	–0.193	.848
CTT1				
Errors	0.23 (0.58)	0.26 (0.08)	–0.146	.409
Execution time (seconds)	86.65 (42.45)	114.02 (162.60)	–0.832	.885
CTT2				
Errors	0.81 (1.09)	0.85 (1.29)	–0.134	.894
Execution time (seconds)	173.42 (82.06)	169.71 (93.44)	0.154	.878

Evaluation Outcomes

Global cognitive function was estimated at baseline and follow-up assessment with the MMSE. The ANOVA on the overall score of the MMSE revealed a main effect of assessment ($F_{1,52}$=20.68, eta-square=0.41, $P<.001$), indicating an improvement of cognitive ability from baseline (mean 26.42, SE 0.38) to follow-up (mean 27.70, SE 0.26). No statistically significant main effect of group or interaction effect between factors (*F*s<1) was found (Figure 3).

Frontal lobe functions were also evaluated under the same factorial design. The data from the FAB showed a significant interaction effect between factors ($F_{1,52}$=8.00, eta-square=0.16, P=.01). Simple effects analysis was performed to test differences between assessments on each group (experimental and control). These data indicated a significant increase in the FAB score from baseline (mean 13.89, SE 0.58) to follow-up (mean 15.50, SE 0.46) only in the experimental group, but not in the control group (mean 14.14, SE 0.56) and (mean 14.14, SE 0.45), respectively for baseline and follow-up (Figure 4).

Cognitive flexibility was measured on three different dimensions of the WCST: the correct responses, the number of categories, and the number of perseverative errors during the task. The ANOVA showed a main effect of assessment on the number of correct responses ($F_{1,52}$=15.10, eta-square=0.29, $P<.001$) and on the number of completed categories ($F_{1,52}$=4.94, eta-square=0.09, *P*=.03), indicating an improvement in performance of the WCST from baseline to follow-up assessment, but there was no interaction effect with the treatment factor (*F*<1). No significant effects (*F*<1) were reported for perseverative errors (Figure 5).

Speed of processing and attentional functioning were assessed respectively with execution time and error rate of the two forms of the CTT. We found a main effect of assessment for both error rate and execution time in the CTT1 and in the CTT2. A significant decrease in error rate ($F_{1,52}$=5.20, eta-square=0.10, *P*=.03) and execution time ($F_{1,52}$=4.26, eta-square=0.08, *P*=.04) was found in the CTT1 (Figure 6). The same pattern, but with a stronger effect, was found for the results of the CTT2, with a decrease in error rate ($F_{1,52}$=13.23, eta-square=0.25; *P*<.001) as well as execution time ($F_{1,52}$=14.41, eta-square=0.28, *P*<.001) (Figure 7). In both the CTT1 and the CTT2, there was no significant interaction (*F*<1) with the treatment factor.

Figure 3. MMSE total scores for the experimental and control conditions at baseline and follow-up.

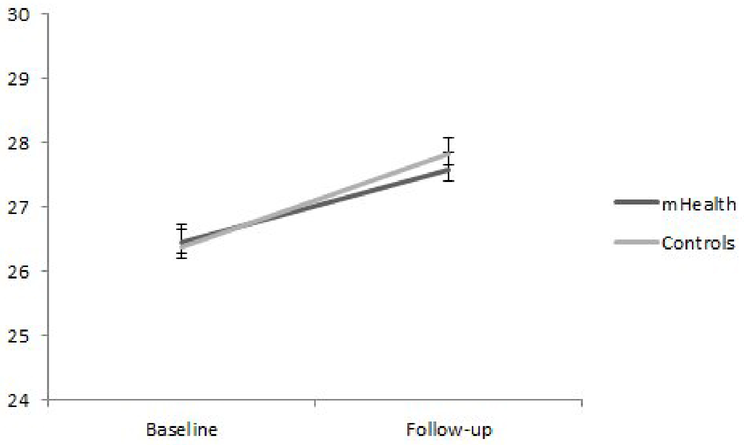


Figure 4. FAB total scores for the experimental and control conditions at baseline and follow-up.

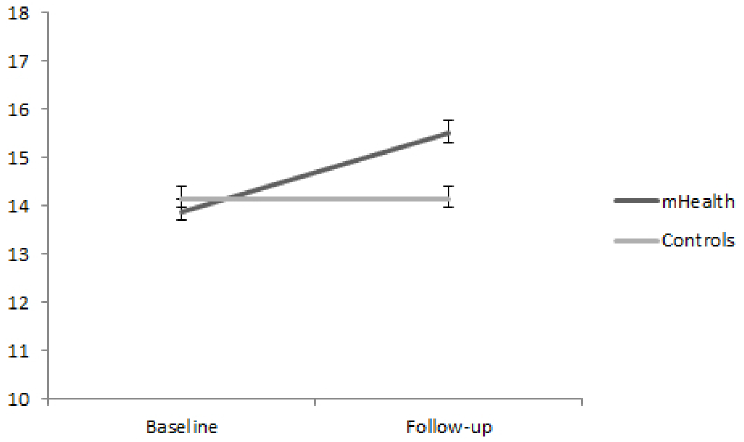


Figure 5. WCST correct responses (left) perseverative errors (middle) and completed categories (right) for the experimental and control conditions at baseline and follow-up.

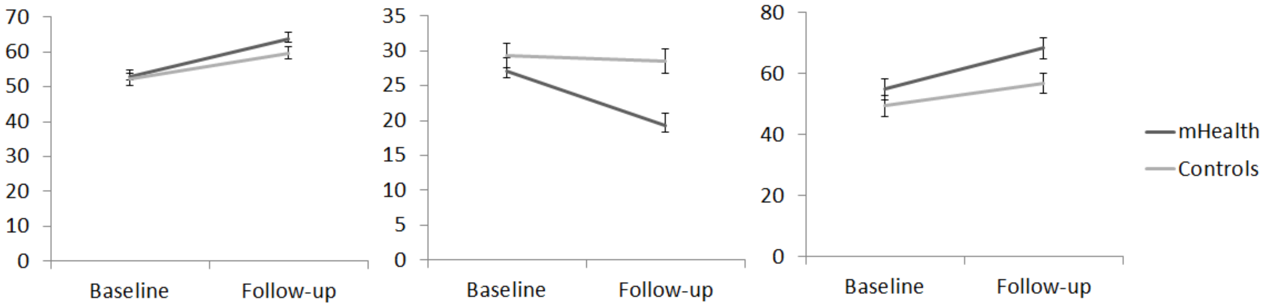


Figure 6. CTT1 execution time in seconds (left) and mean errors (right) for the experimental and control conditions at baseline and follow-up.

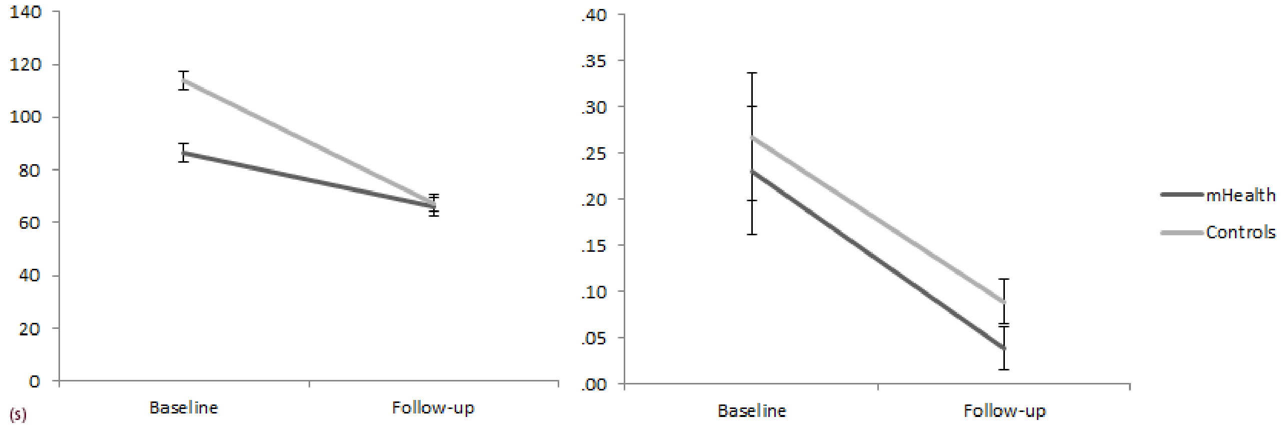
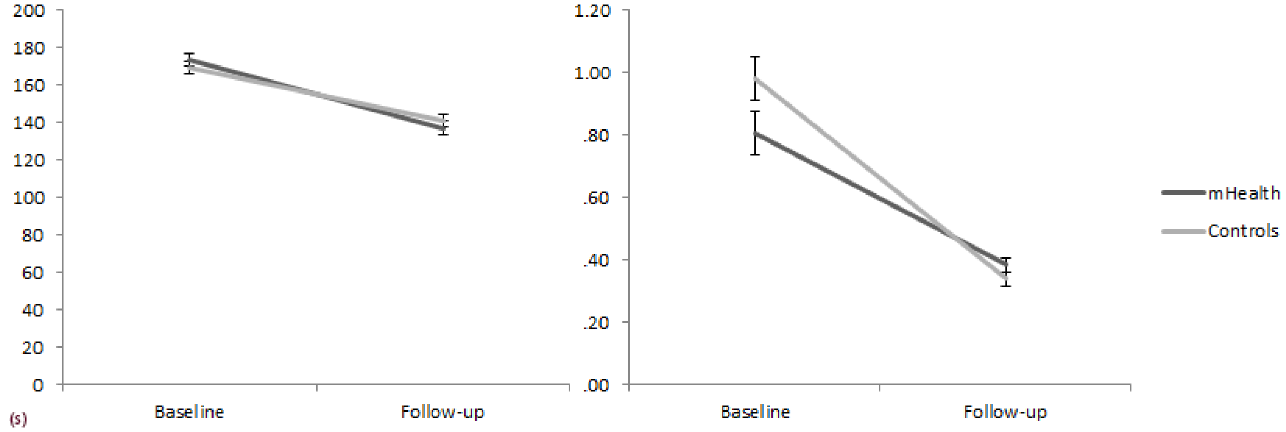


Figure 7. CTT2 execution time in seconds (left) and mean errors (right) for the experimental and control conditions at baseline and follow-up.



Discussion

Principal Findings

The current experiment was designed to study the effectiveness of a cognitive stimulation program in alcohol abusers' executive functioning with mobile SG applications. According to previous studies, alcohol dependence may impair information processing in the prefrontal areas, such as the dorsolateral prefrontal cortex and the anterior cingulate cortex, which are thought to be related to executive functioning in cognitive domains ranging from attention and working memory to higher-order functions of reasoning and decision making. In this study, our aim was to evaluate the effect of cognitive stimulation with mobile technology and serious games on the general cognitive ability of patients diagnosed with ADS.

To do this, we carried out an RCT on patients undergoing an alcohol rehabilitation treatment at a specialized institution for treatment of alcohol dependence. The overall MMSE scores at baseline suggest deficits in general cognitive ability of ADS patients, which concurs with previous studies highlighting the relationship between the alcohol abuse and cognitive impairments [4]. There were no differences at baseline between control and treatment groups, which indicates a successful random distribution of patients.

Our results indicate a general improvement of cognitive abilities in both patient groups, which reflects an effect of the withdrawal of the direct impact of alcohol on the brain. In support of the specific effectiveness of the cognitive stimulation, this effect was qualified with an interaction with the treatment factor in the FAB test, indicating a more pronounced increase of frontal cognitive abilities in patients subjected to mHealth SG approach, with improvements in frontal lobe function and executive functioning. However, there were no significant effects of the treatment on the MMSE and WCST, which measure more specific, not general, domains of frontal lobe functioning. These results are in agreement with those of other cognitive rehabilitation trials with addict populations, suggesting that the positive effects of neuropsychological rehabilitation in addicted populations are restricted to cognitive abilities related to frontal lobes [21].

The explanation for this is that both traditional exercises and SG applications focused mainly on cognitive abilities related to frontal lobe functioning, such as attention, working memory, decision making, and planning. The systematic and repeated stimulation of these functions may have had a more pronounced effect on those domains, but not on the others. Our results suggest that the neuropsychological benefits of our mHealth SG cognitive stimulation program in alcoholics may be limited to frontal lobe general functioning and that may arise from the response demands of the game tasks. That is, the processes required to meet the games' tasks are most closely related to the processes that are captured by the FAB.

In addition, the overall feedback from the participants was positive. Qualitative comments were mainly related to the technological and innovative features of this approach and intrinsic aspects such as a positive motivation to pursue a goal in the tasks.

Limitations

This study was not designed to test the relative effectiveness of mobile intervention with online stimulation and serious games against traditional pencil-and-paper cognitive stimulation exercises. It was simply to test if this mHealth solution could add effectiveness in cognitive rehabilitation to treatment-as-usual procedures. Further studies testing the relative effectiveness of SG vs traditional approaches are needed. Due to the characteristics of SG and of the clinical context, it would not have been possible to conduct a double blind study. However, the therapists conducting the assessment were not aware of which group (experimental vs control) the patients were enrolled in.

Conclusions

The increase in general cognitive function for both experimental and controls supports the beneficial role of existing alcohol treatment protocols, which are based on medication to help minimize withdrawal symptoms along with behavioral therapy mainly to help patients manage their stress levels during abstinence, and indicates that the neuropsychological effects of alcohol abuse on brain structure are reversible through rehabilitation. The effect of the trial is encouraging, suggesting an improvement specifically in frontal cognitive general functioning in alcoholics following an mHealth approach with SG in line with previous studies using cognitive stimulation with addicts in rehab.

Acknowledgments

Hardware and software acquisition was funded by Centre for the Studies on Cognitive and Learning Psychology (where the first five authors were researchers until its merger into a larger center, COPELABS of Universidade Lusófona de Humanidades e Tecnologias, where they are currently researchers) via the Foundation for Science and Technology (FCT) of Portugal (PEst-OE/PSI/UI0700/2011).

Authors' Contributions

All authors contributed to the design and analysis of the study and have approved the final manuscript, but specific authors had a major role in the following aspects of the study: Pedro Gamito: concept, intervention design, and original draft; Jorge Oliveira: statistical analyses; Paulo Lopes: evaluation protocol and supervision of the intervention; Rodrigo Brito: additional literature review and critical revision; Diogo Morais: original literature review; Diana Silva and Ana Silva: neuropsychological assessment; Alberto Deus: supervision of the assessment; Sara Rebelo and Marta Bastos: intervention.

Conflicts of Interest

The authors have developed and own the majority of the applications, but these are freely available online and no commercial profit is intended from them.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V1.6.2 [43].

[PDF File (Adobe PDF File), 980KB - jmir_v16i4e102_app1.pdf]

References

1. John U, Rumpf HJ, Bischof G, Hapke U, Hanke M, Meyer C. Excess mortality of alcohol-dependent individuals after 14 years and mortality predictors based on treatment participation and severity of alcohol dependence. Alcohol Clin Exp Res 2013 Jan;37(1):156-163. [doi: 10.1111/j.1530-0277.2012.01863.x] [Medline: 23072405]
2. Breslow RA, Graubard BI. Prospective study of alcohol consumption in the United States: quantity, frequency, and cause-specific mortality. Alcohol Clin Exp Res 2008 Mar;32(3):513-521. [doi: 10.1111/j.1530-0277.2007.00595.x] [Medline: 18215212]
3. Chan RC, Shum D, Toulopoulou T, Chen EY. Assessment of executive functions: review of instruments and identification of critical issues. Arch Clin Neuropsychol 2008 Mar;23(2):201-216. [doi: 10.1016/j.acn.2007.08.010] [Medline: 18096360]
4. Moselhy HF, Georgiou G, Kahn A. Frontal lobe changes in alcoholism: a review of the literature. Alcohol Alcohol 2001;36(5):357-368 [FREE Full text] [Medline: 11524299]
5. Elofson J, Gongvatana W, Carey KB. Alcohol use and cerebral white matter compromise in adolescence. Addict Behav 2013 Jul;38(7):2295-2305. [doi: 10.1016/j.addbeh.2013.03.001] [Medline: 23583835]
6. Oscar-Berman M, Marinković K. Alcohol: effects on neurobehavioral functions and the brain. Neuropsychol Rev 2007 Sep;17(3):239-257. [doi: 10.1007/s11065-007-9038-6] [Medline: 17874302]
7. Stavro K, Pelletier J, Potvin S. Widespread and sustained cognitive deficits in alcoholism: a meta-analysis. Addict Biol 2013 Mar;18(2):203-213. [doi: 10.1111/j.1369-1600.2011.00418.x] [Medline: 22264351]

8. Beatty WW, Tivis R, Stott HD, Nixon SJ, Parsons OA. Neuropsychological deficits in sober alcoholics: influences of chronicity and recent alcohol consumption. Alcohol Clin Exp Res 2000 Feb;24(2):149-154. [Medline: 10698365]
9. Cantrell H, Finn PR, Rickert ME, Lucas J. Decision making in alcohol dependence: insensitivity to future consequences and comorbid disinhibitory psychopathology. Alcohol Clin Exp Res 2008 Aug;32(8):1398-1407 [FREE Full text] [doi: 10.1111/j.1530-0277.2008.00714.x] [Medline: 18565158]
10. Davies SJ, Pandit SA, Feeney A, Stevenson BJ, Kerwin RW, Nutt DJ, et al. Is there cognitive impairment in clinically 'healthy' abstinent alcohol dependence? Alcohol Alcohol 2005;40(6):498-503 [FREE Full text] [doi: 10.1093/alcalc/agh203] [Medline: 16186142]
11. Goudriaan AE, Grekin ER, Sher KJ. Decision making and response inhibition as predictors of heavy alcohol use: a prospective study. Alcohol Clin Exp Res 2011 Jun;35(6):1050-1057 [FREE Full text] [doi: 10.1111/j.1530-0277.2011.01437.x] [Medline: 21332527]
12. Le Berre AP, Vabret F, Cauvin C, Pinon K, Allain P, Pitel AL, et al. Cognitive barriers to readiness to change in alcohol-dependent patients. Alcohol Clin Exp Res 2012 Sep;36(9):1542-1549. [doi: 10.1111/j.1530-0277.2012.01760.x] [Medline: 22458335]
13. Noël X, Bechara A, Dan B, Hanak C, Verbanck P. Response inhibition deficit is involved in poor decision making under risk in nonamnesic individuals with alcoholism. Neuropsychol 2007;21:778-786. [doi: 10.1037/0894-4105.21.6.778]
14. Pitel AL, Rivier J, Beaunieux H, Vabret F, Desgranges B, Eustache F. Changes in the episodic memory and executive functions of abstinent and relapsed alcoholics over a 6-month period. Alcohol Clin Exp Res 2009 Mar;33(3):490-498. [doi: 10.1111/j.1530-0277.2008.00859.x] [Medline: 19120052]
15. Hartley DE, Elsabagh S, File SE. Binge drinking and sex: effects on mood and cognitive function in healthy young volunteers. Pharmacol Biochem Behav 2004 Jul;78(3):611-619. [doi: 10.1016/j.pbb.2004.04.027] [Medline: 15251270]
16. Parada M, Corral M, Mota N, Crego A, Rodríguez Holguín S, Cadaveira F. Executive functioning and alcohol binge drinking in university students. Addict Behav 2012 Feb;37(2):167-172. [doi: 10.1016/j.addbeh.2011.09.015] [Medline: 21996093]
17. García-Moreno L, Expósito J, Sanhueza C, Angulo M. Actividad prefrontal y alcoholismo de fin de semana en jóvenes. Prefrontal activity and weekend alcoholism in young people Adicciones 2008;20:271-279.
18. Allen DN, Goldstein G, Seaton BE. Cognitive rehabilitation of chronic alcohol abusers. Neuropsychol Rev 1997 Mar;7(1):21-39. [Medline: 9243529]
19. Grohman K, Fals-Stewart W. Computer-assisted cognitive rehabilitation with substance-abusing patients: Effects on treatment response. J Cogn Rehab 2003;21(4):10-17.
20. Fals-Stewart W, Lam WK. Computer-assisted cognitive rehabilitation for the treatment of patients with substance use disorders: a randomized clinical trial. Exp Clin Psychopharmacol 2010 Feb;18(1):87-98. [doi: 10.1037/a0018058] [Medline: 20158298]
21. Goldstein G, Haas GL, Shemansky WJ, Barnett B, Salmon-Cox S. Rehabilitation during alcohol detoxication in comorbid neuropsychiatric patients. J Rehabil Res Dev 2005;42(2):225-234 [FREE Full text] [Medline: 15944887]
22. Ruiz-Sánchez de León JM, Pedrero-Pérez EJ, Rojo-Mota G, Llanero-Luque M, Puerta-García C. [A proposal for a protocol of neuropsychological assessment for use in addictions]. Rev Neurol 2011 Oct 16;53(8):483-493 [FREE Full text] [Medline: 21960389]
23. Gamito P, Oliveira J, Lopes P, Morais D, Brito R, Saraiva T, et al. Assessment of frontal brain functions in alcoholics following a health mobile cognitive stimulation approach. Stud Health Technol Inform 2013;191:110-114. [Medline: 23792854]
24. Gamito P, Oliveira J, Pacheco J, Morais D, Saraiva T, Lacerda R, et al. Traumatic Brain Injury memory training: a Virtual Reality online solution. International Journal on Disability and Human Development 2011;10(4):309-312. [doi: 10.1515/IJDHD.2011.049]
25. Morris ME, Kathawala Q, Leen TK, Gorenstein EE, Guilak F, Labhard M, et al. Mobile therapy: case study evaluations of a cell phone application for emotional self-awareness. J Med Internet Res 2010;12(2):e10 [FREE Full text] [doi: 10.2196/jmir.1371] [Medline: 20439251]
26. van Bastelaar KM, Pouwer F, Cuijpers P, Riper H, Twisk JW, Snoek FJ. Is a severe clinical profile an effect modifier in a Web-based depression treatment for adults with type 1 or type 2 diabetes? Secondary analyses from a randomized controlled trial. J Med Internet Res 2012;14(1):e2 [FREE Full text] [doi: 10.2196/jmir.1657] [Medline: 22262728]
27. Klein B, Meyer D, Austin DW, Kyrios M. Anxiety online: a virtual clinic: preliminary outcomes following completion of five fully automated treatment programs for anxiety disorders and symptoms. J Med Internet Res 2011;13(4):e89 [FREE Full text] [doi: 10.2196/jmir.1918] [Medline: 22057287]
28. Schaub M, Sullivan R, Haug S, Stark L. Web-based cognitive behavioral self-help intervention to reduce cocaine consumption in problematic cocaine users: randomized controlled trial. J Med Internet Res 2012;14(6):e166 [FREE Full text] [doi: 10.2196/jmir.2244] [Medline: 23192752]
29. Eghdam A, Scholl J, Bartfai A, Koch S. Information and communication technology to support self-management of patients with mild acquired cognitive impairments: systematic review. J Med Internet Res 2012;14(6):e159 [FREE Full text] [doi: 10.2196/jmir.2275] [Medline: 23165152]

30. Hester RK, Lenberg KL, Campbell W, Delaney HD. Overcoming Addictions, a Web-based application, and SMART Recovery, an online and in-person mutual help group for problem drinkers, part 1: three-month outcomes of a randomized controlled trial. J Med Internet Res 2013;15(7):e134 [FREE Full text] [doi: 10.2196/jmir.2565] [Medline: 23846588]
31. Check Your Drinking. 2013. URL: http://www.checkyourdrinking.net/CYD/CYDScreenerP1_0.aspx [accessed 2013-11-03] [WebCite Cache ID 6Kq0ygRTr]
32. Alcohol Help Center. 2013. URL: http://www.alcoholhelpcenter.net/ [accessed 2013-11-03] [WebCite Cache ID 6Kq0iIMNR]
33. Cunningham JA. Comparison of two Internet-based interventions for problem drinkers: randomized controlled trial. J Med Internet Res 2012;14(4):e107 [FREE Full text] [doi: 10.2196/jmir.2090] [Medline: 22954459]
34. Gamito. 2013. URL: http://labpsicom.ulusofona.pt/ [accessed 2013-08-29] [WebCite Cache ID 6JEYtW6JH]
35. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [Medline: 1202204]
36. Derogatis LR. SCL-90-R: administration, scoring and procedures manual II. Towsen, MD: Clinical Psychometric Research; 1992.
37. Morgenstern J, Bux D, Labouvie E, Blanchard KA, Morgan TI. Examining mechanisms of action in 12-step treatment: the role of 12-step cognitions. J Stud Alcohol 2002 Nov;63(6):665-672. [Medline: 12529066]
38. Guerreiro M, Silva AP, Botelho MA, Leitão A, Castro-Caldas A Garcia C. Adaptação à população portuguesa da tradução do Mini-Mental State Examination (MMSE). Revista Portuguesa de Neurologia 1994;3(1):9-10.
39. Dubois B, Slachevsky A, Litvan I, Pillon B. The FAB: a Frontal Assessment Battery at bedside. Neurology 2000 Dec 12;55(11):1621-1626. [Medline: 11113214]
40. Heaton RK, Chelune GJ, Talley JL, Kay GG, Curtiss G. Wisconsin card sorting test (WCST) — manual revised and expanded. Odessa, FL: Psychological Assessment Resources; 1993.
41. D'Elia L, Satz P, Uchiyama C, White T. Color trails test professional manual. Odessa, FL: Psychological Assessment Resources; 1999.
42. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: L. Erlbaum Associates; 1988.
43. 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

ADS: Alcohol Dependence Syndrome
AHC: AlcoholHelpCenter.net
ANOVA: analysis of variance
BD: binge drinking
CR: cognitive rehabilitation
CTT: Color Trail Test
CYD: Check Your Drinking
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, 4th edition
FAB: Frontal Assessment Battery
mHealth: mobile health
MMSE: Mini Mental State Examination
PASAT: Paced Auditory Serial Addition Task
RCT: randomized controlled trial
SCL-90-R: Symptoms Checklist Revised
SG: serious games
SOPT: Speed of Processing Test
SPSS: Statistical Package for Social Sciences
TAVEC: Verbal Learning Test of the Complutense University
WCST: Wisconsin Card Sorting Test
WMS: Wechsler Memory Scale

Edited by A Mashat; submitted 31.08.13; peer-reviewed by S Pérez; comments to author 22.10.13; revised version received 05.11.13; accepted 23.11.13; published 17.04.14.

Please cite as:
Gamito P, Oliveira J, Lopes P, Brito R, Morais D, Silva D, Silva A, Rebelo S, Bastos M, Deus A
Executive Functioning in Alcoholics Following an mHealth Cognitive Stimulation Program: Randomized Controlled Trial
J Med Internet Res 2014;16(4):e102
URL: http://www.jmir.org/2014/4/e102/
doi:10.2196/jmir.2923
PMID:24742381

Original Paper

Web-Based Virtual Patients in Nursing Education: Development and Validation of Theory-Anchored Design and Activity Models

Carina Georg[1,2*], RN, CCRN, MSN; Nabil Zary[1*], PhD

[1]Center for Learning and Knowledge (CLK), Department of Learning, Informatics, Management and Ethics (LIME), Karolinska Institutet, Stockholm, Sweden

[2]Department of Neurobiology, Care Science and Society, Division of Nursing, Karolinska Institutet, Huddinge, Sweden

[*]all authors contributed equally

Corresponding Author:
Carina Georg, RN, CCRN, MSN
Center for Learning and Knowledge (CLK)
Department of Learning, Informatics, Management and Ethics (LIME)
Karolinska Institutet
Widerströmska building, 4th Floor
Tomtebodavägen 18a
Stockholm, 171 77
Sweden
Phone: 46 73 3783608
Fax: 46 8345128
Email: carina.georg@ki.se

Abstract

Background: Research has shown that nursing students find it difficult to translate and apply their theoretical knowledge in a clinical context. Virtual patients (VPs) have been proposed as a learning activity that can support nursing students in their learning of scientific knowledge and help them integrate theory and practice. Although VPs are increasingly used in health care education, they still lack a systematic consistency that would allow their reuse outside of their original context. There is therefore a need to develop a model for the development and implementation of VPs in nursing education.

Objective: The aim of this study was to develop and evaluate a virtual patient model optimized to the learning and assessment needs in nursing education.

Methods: The process of modeling started by reviewing theoretical frameworks reported in the literature and used by practitioners when designing learning and assessment activities. The Outcome-Present State Test (OPT) model was chosen as the theoretical framework. The model was then, in an iterative manner, developed and optimized to the affordances of virtual patients. Content validation was performed with faculty both in terms of the relevance of the chosen theories but also its applicability in nursing education. The virtual patient nursing model was then instantiated in two VPs. The students' perceived usefulness of the VPs was investigated using a questionnaire. The result was analyzed using descriptive statistics.

Results: A virtual patient Nursing Design Model (vpNDM) composed of three layers was developed. Layer 1 contains the patient story and ways of interacting with the data, Layer 2 includes aspects of the iterative process of clinical reasoning, and finally Layer 3 includes measurable outcomes. A virtual patient Nursing Activity Model (vpNAM) was also developed as a guide when creating VP-centric learning activities. The students perceived the global linear VPs as a relevant learning activity for the integration of theory and practice.

Conclusions: Virtual patients that are adapted to the nursing paradigm can support nursing students' development of clinical reasoning skills. The proposed virtual patient nursing design and activity models will allow the systematic development of different types of virtual patients from a common model and thereby create opportunities for sharing pedagogical designs across technical solutions.

(J Med Internet Res 2014;16(4):e105) doi:10.2196/jmir.2556

KEYWORDS

virtual patient; patient simulation; nursing education; clinical reasoning; e-learning, simulation technology

Introduction

Background

One of the main goals in health care education is to supply society with knowledgeable, up-to-date, and skilled professionals [1]. The societal and health care expectations in providing adequate learning experiences can be overwhelming for both learners and educators [2], and the health care environment is dynamic and continuously changing with fewer patients in hospitals staying shorter periods and with specialized health care needs. There are also technological advances, financial challenges, and numerous regulations [3]. The teaching and learning of professional skills must be conducted without waiving patient safety. Therefore, this continuously changing health care environment requires new models for training health care professionals [4]. One challenge is how to teach students to apply their knowledge when they are dealing with clinical problems [5]. Technology-enhanced simulation is one possible solution [4]. Different types of simulation modalities as a teaching method in health care education have developed rapidly during the last decade [6,7]. Simulation offers a safe and realistic environment in which to learn, practice, and make mistakes without direct contact with and risk of harm to patients [8]. A type of computer-based simulation called virtual patients (VPs) has been proposed to support nursing students in their acquisition of scientific knowledge as a way to integrate theory and practice and promote clinical reasoning.

Challenges in Nursing Education

Nursing as a profession is at a turning point. Nurses' work has evolved as a result of a profound change in health care environments, science and technology, and the settings and nature of nursing practice. These changes have important implications for nursing education [3].

Nursing education varies in different parts of the world [9]. The variation lies in the organization, content, and quality of education [10]. In Sweden, the nursing study program covers a total of 180 credits and is conducted at a university level leading to a Bachelor of Science degree in Nursing. After graduation, nursing students become registered nurses (RNs) and obtain a license to work. The goal for nursing education is to support students in acquiring the knowledge and competencies required to provide a good quality of care. The newly graduated nurse should be prepared to practice safely, accurately, and compassionately in various settings. To be able to conduct safe and effective care, the newly graduated nurse must be proficient both in practical knowledge and science and also in a wide range of skills such as clinical reasoning [3]. The development of clinical performance and competencies among nursing students continues to challenge nurse educators, specifically when trying to link nursing theory and evidence-based clinical practice for the student [2,3].

A first step for students toward achieving competencies is to acquire theoretical scientific knowledge and then use this theoretical knowledge in patient-related practice. However, research has shown that nursing students find it difficult to translate and apply their theoretical knowledge in a clinical context [3]. Therefore, a challenge in nursing education is to find ways of preparing nursing students to manage, interpret data, evaluate nursing activities and interventions, and to translate theoretical knowledge to the clinical context. Nurses must think and reason across diverse contexts, which requires the development of strong clinical reasoning skills [11]. Nursing care also integrates knowledge from other disciplines such as medicine and behavioral science [3]. The students usually develop and integrate various aspects of knowledge during their encounters with patients while they observe, understand, and argue for their choices. Theory and reflection form the basis for nursing knowledge that is the prerequisite for good management of the patient [11]. Additionally, critical thinking is an essential skill that requires both cognitive and metacognitive capabilities [12].

Many courses in nursing have learning outcomes that include clinical decision making, clinical reasoning, and critical thinking [13]. However, the literature reports limited skills in clinical reasoning among nursing students. Therefore, nursing students must learn how to apply theoretical knowledge acquired in classrooms to the clinical context. Nursing educators have to develop teaching and learning approaches that enable students to acquire knowledge and skills and to then apply theoretical knowledge in the clinical context.

Clinical Reasoning in Nursing

Clinical reasoning (CR) is not unique to nursing—all professionals use CR to reach decisions [14]. CR is a complex task [15], and in this study, we have adopted Higgs' definition:

> *Clinical reasoning (or practice decision making) is a context-dependent way of thinking and decision making in professional practice to guide practice actions. It involves the construction of narratives to make sense of the multiple factors and interests pertaining to the current reasoning task. It occurs within a set of problem spaces informed by the practitioner´s unique frames of reference, workplace context and practice model, as well as by the patient's or clients' contexts. It utilizes core dimensions of practice knowledge, reasoning and metacognition and draws on these capacities in others. Decision making within clinical reasoning occurs at micro, macro and meta-levels and may be individually or collaboratively conducted. It involves metaskills of critical conversations, knowledge generation, practice model authenticity and reflexivity.* [16]

CR is a logical process where nurses and other health professionals collect cues, process information, and come to an understanding about patients' situations or problems, plan and implement interventions, and evaluate outcomes [17]. CR represents the essence of nursing practice [12]. There are several different types of clinical reasoning, including problem-effect, which means identifying the problem, seeing the factors influencing the problem, and finding solutions to the problem [15]. This approach helps nurses identify nursing problems and prescribe and implement nursing actions [18].

In terms of CR, a nurse's focus is the process of care and the well-being of the patient. A nurse evaluates both subjective factors (like descriptions of symptoms) and objective factors (like lab data). The nurse should try to see possible patterns and decide on, plan, and perform patient care, as well as continuously assess care interventions and status of the patient. A competent nurse that provides safe patient care possesses skills in caring sciences and knowledge in both medical and behavioral sciences [19].

CR is widely acknowledged as a fundamental part of health professionals' education, but there is still a need for the development of innovative teaching and learning methods that enhance CR skills among nursing students [3,18,20]. Virtual patients (VPs) have been proposed as a learning activity that can support students to integrate theory and practice in their development of clinical reasoning skills [21,22].

Virtual Patients

VPs can take many different forms [23] and can be realized using a wide range of presentations, styles, and configurations [24]. There are also several definitions for the concept of the virtual patient [21], and the term VP is often used in an ambiguous manner [25]. In this study, we adopted Ellaway's definition: "An interactive computer simulation of real-life clinical scenarios for the purpose of health care and medical training, education or assessment" [26].

VPs can be designed in different ways and are often dependent on the technical affordances of the system used to author them [25]. Most VPs have common features including medical history taking, physical examinations, lab/imaging tests, as well as features for suggesting an appropriate diagnosis and treatment [21]. An essential characteristic of VPs is the interactive interface that enables the user to query the patient and receive a patient response supplied by the computer [21,22,27].

A particular strength of VPs is that they seem to support learning on clinical reasoning and decision making. Students may be exposed to a large number of VP cases in a safe and controlled environment [21,24]. Clinical learning experience is difficult to standardize and schedule in a reproducible manner. However, VPs can provide exactly the same experience repeatedly and also allow students to revisit their actions during the interaction with the virtual patient, and then compare them with the best practice protocol. VPs also facilitate a venue for safe and repetitive practice and stand as a model where progressive clinical variation and difficulty can be presented [28]. Studies also show that VPs are a cost-effective way to teach and assess clinical skills and clinical reasoning among medical students [29].

Other studies indicate that participation in education using VPs may bridge the gap between experimental learning (learning by doing) and information representation (lectures) and consequently achieve a variety of learning outcomes [22,27]. Studies have also shown that VPs can enhance training of essential nursing skills and knowledge [30]. Although the research surrounding best practice of VPs is in its infancy [21], there are some studies that suggest that new learning technologies such as VPs and virtual reality may be useful to support nursing education. A challenge with most of the reported studies is their focus on the medical curricula [31,32]. There is a therefore an urgent need to investigate how to model VPs for nursing education.

Unfortunately, research on these learning technologies has not kept pace with the rapid technological development [27] and often lacks a theoretical foundation. The lack of empirical base in research lies at three levels: (1) the VPs reported in the literature have been created without an explicit and consistent underlying theoretical model, making it difficult for researchers and practitioners to build on the findings, (2) while a learning activity encompasses more than the VPs, we could not find that aspect systematically presented, and (3) there are different types of VPs and therefore a need to clearly understand the strengths and weaknesses of each type.

Objective of the Study

The aim of this study was to (1) develop a theory-anchored model for developing virtual patients in nursing education, (2) investigate how VPs could be instantiated as a learning activity, and finally (3) explore the students' perceived usefulness of virtual patients created from the developed model. To reach this aim, three research questions were asked:

1. What aspects of clinical reasoning should be present in a virtual patient nursing model?
2. How should the virtual patient nursing model be represented as a learning activity?
3. How do the nursing students perceive the usefulness of virtual patients as an artefact of the developed models?

Methods

Modeling of the Aspects in a Virtual Patient

The process of modeling started by a review of existing theoretical frameworks reported in the literature and in actual use when designing learning and assessment activities in nursing education. The 3rd generation nursing process model, the Outcome-Present State Test (OPT) model [11,33], was chosen because it is widely used internationally and was relevant for the focus of the study. This theoretical framework is outcome driven and builds on earlier versions of the nursing process. The OPT model provides structure for the iterative clinical reasoning necessary for contemporary nursing practice. The OPT model emphasizes reflection, outcome specification, and tests for judgment within the context of the individual patient story. It provides a way for nursing students to frame and attribute meaning to patient stories while concurrently considering relationships among diagnoses intervention and outcomes, with attention to evidence-based nursing and the decision-making process [34]. Essential processes contained within the OPT model are reflection, framing, cue logic, testing, decision making, and judgment [33]. An essential part of the OPT model is the emphasis on framing a situation based on the story of a client. The OPT model uses the facts associated with the client's story to frame the context, content, and major issues for clinical reasoning. The patient stories are a key element of clinical reasoning [11,33].

The virtual patient nursing model was then, in an iterative manner, developed and optimized. Validation of the model was performed with faculty both in terms of relevance of chosen theories but also applicability of aspects represented in the model for the nursing education context.

Representation of the Virtual Patient Model as a Learning Activity

Curriculum integration plays a key role in the implementation of new learning technologies [21]. In order to translate the virtual patient Nursing Design Model (vpNDM) to a learning activity, we chose the context of a second year undergraduate nursing course at a university in Sweden. The course content includes topics from nursing and medicine, and theory with practice. A large part of the course is based in a clinical setting.

Subject matter experts and educational researchers reviewed which intended learning outcomes in the course plan were suitable to be addressed using virtual patients. The faculty group then defined learning goals for the learning activity based on both the chosen intended learning outcomes and the developed vpNDM. The goal for the student while working with the VP cases was based on the nursing process, training nursing documentation, and nursing diagnosis. Students must increase knowledge of how to frame a patient story, identify and analyze it, and then document and evaluate it. In other words, they are trained to develop a nursing care plan based on the individual patient's specific needs. An additional aim was to describe nursing interventions related to the nursing needs of the patient and to practice using measurement instruments for the assessment.

Based on the defined goal and requirements, a multidisciplinary team of subject matter experts, nurse educators, educational researchers, and clinical active nurses created two VP cases. Both cases were a loose amalgamation of actual cases experienced by the clinical team members.

Authoring of the Virtual Patients

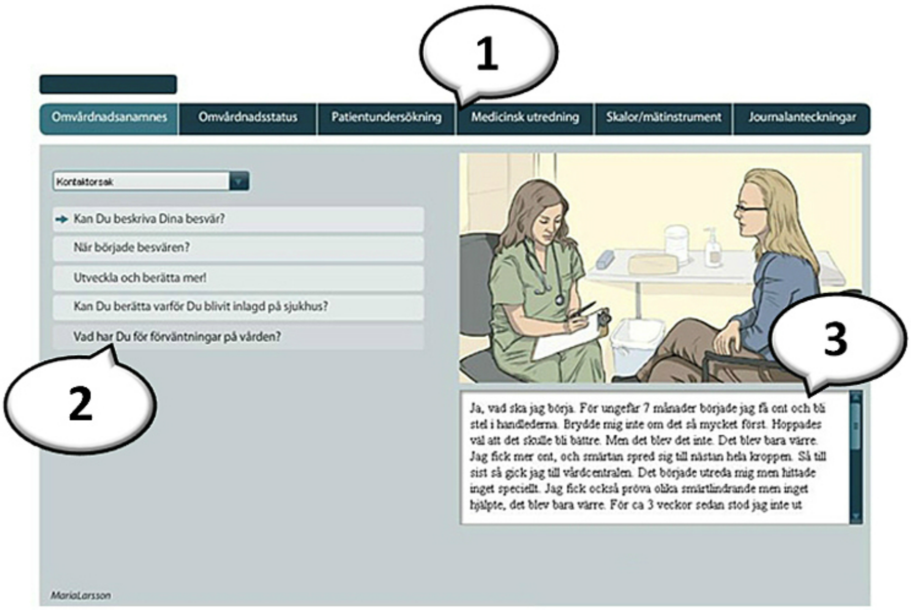
Method

We used the Virtual Interactive Case system (VIC) as a prototyping tool to create the VPs. VIC is a Web-based application using Adobe Flash. The VPs were accessible to the students through the university's online course management system. The navigation was global linear, which means that there was a linear process but the user could freely navigate between the different sections of the VP [25]. The answers given by the VP were a combination of text and multimedia (images, sounds) [35].

A common model of interaction in VPs is the presence of long menus of questions, formulated by the author of the patient case. Students choose from a wide range of questions about different issues, and part of the task is to be able to select relevant questions. Figure 1 shows how students can navigate and decide on different options (1) containing the patient's past history, the patient's perception of the current situation, physical examination, medical investigation, scales and measuring instruments, and patient journal records. (2) Within each option, the student can then select among a number of questions or inquiries. (3) The answer is then presented as text, image, movie, or sound file.

In this study, we created two VPs. The first case was a woman in her 40s who was admitted to a ward with a 6-month history of pain related to joint inflammation (rheumatoid arthritis). She had several different nursing diagnoses related to chronic pain and readiness for enhanced comfort, impaired physical mobility, insomnia, and risk for impaired skin integrity. The second VP was an older man with type 2 diabetes mellitus. He was admitted to the ward with heart failure, hyperglycemia, and an acute diabetic foot. The nursing diagnosis in this VP case was related to excess fluid volume, impaired gas exchange, risk for unstable blood glucose level, ineffective self-health management, deficient knowledge, and impaired skin integrity.

Figure 1. Screenshot of the system (VIC) used to author virtual patients.



Ja, vad ska jag börja. För ungefär 7 månader började jag få ont och bli stel i handlederna. Brydde mig inte om det så mycket först. Hoppades väl att det skulle bli bättre. Men det blev det inte. Det blev bara värre. Jag fick mer ont, och smärtan spred sig till nästan hela kroppen. Så till sist så gick jag till vårdcentralen. Det började utreda mig men hittade inget speciellt. Jag fick också pröva olika smärtlindrande men inget hjälpte, det blev bara värre. För ca 3 veckor sedan stod jag inte ut

Participants

The participants (N=102) were undergraduate nursing students whose age ranged between 20-53 years (mean age 23 years). The gender distribution was 12.5% male and 87.5% female. The VPs were integrated in the regular course as mandatory self-study learning.

The participants were free to take part in the study, and their anonymity and confidentiality were respected. The local ethical board approved the study.

Questionnaire

An online self-administered questionnaire developed by the eViP (electronic virtual patients) project team [36] was adjusted to fit nursing education. The questionnaire evaluated the students' experiences of learning while working with virtual patients, focusing on the development of clinical reasoning skills. The questionnaire contains 17 items clustered into five subsets. The subset covered areas such as authenticity of patient encounter and the consultation, professional approach in the consultation, coaching during consultation, learning effect of consultation, and overall judgment of case workup. Responses were on a 5-point Likert scale (strongly disagree to strongly agree) as well as additional free text open-ended questions [36,37]. The participants answered the questionnaire after completing the two VPs. Completion of the survey was a voluntary part of the study.

Analysis

Descriptive statistics were performed and percentages reported. Cronbach alpha was calculated to provide an overall reliability coefficient for the set of questions included in the study. Item-total statistic was examined to check if any item was inconsistent with the others and could thus be discarded. Item-item correlation matrix was observed to examine the correlation between the items. "Strongly agree" and "agree" were combined to create a positive response.

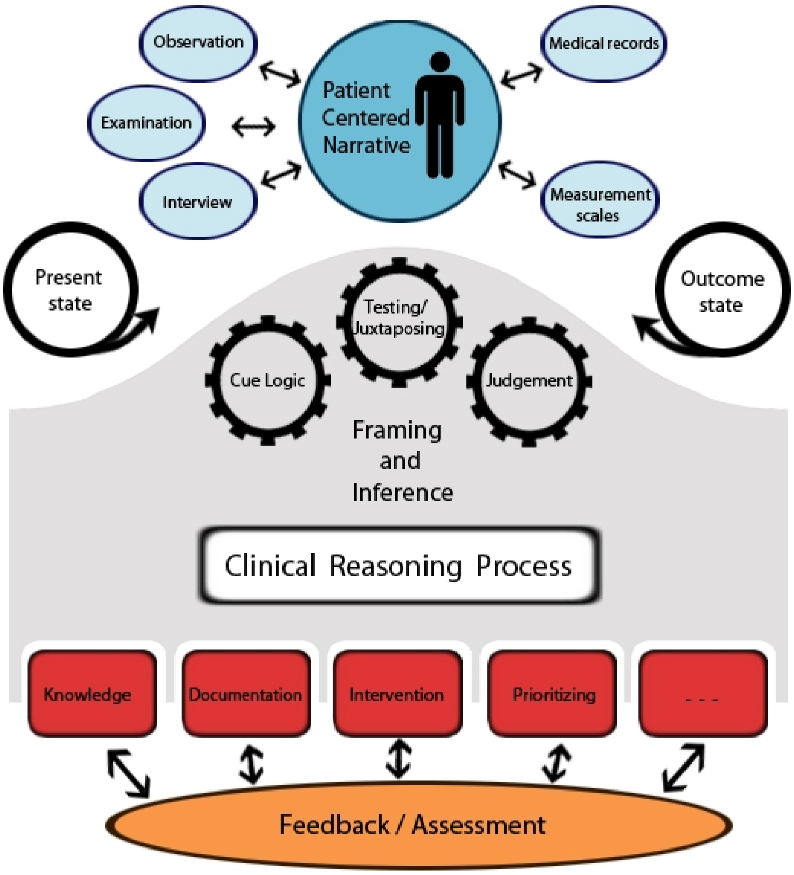
Results

The Virtual Patient Nursing Design Model

Overview

The model we produced (Figure 2) is based on and grounded in the 3rd generation nursing process, using the OPT model [11,33]. This model describes the elements that should be part of a VP meant for nursing education. The vpNDM model is composed of three main layers.

Figure 2. The three layers in the virtual patient nursing design model.



Layer 1: Patient-Centric Data Collection

The first layer includes different methods for collecting data on the patient. By capturing the patient story, the nurse obtains the necessary data to support the identification of nursing care needs. An essential part of nursing involves listening and understanding the meaning in a patient encounter. To capture the patient-centered narrative, the learner is given the opportunity to collect both subjective and objective data from the VP.

In order to capture the patient story, nurses seek information from various sources, for example by conversations or interviews with the patient or their relatives, physical examinations, other health care professionals' evaluations, and patient records. By interviewing the patient, the nurse obtains a history of the patient's past and present health. The past health history describes the patient's previous health status and living conditions while the nursing status describes the current status and factors affecting the current nursing care.

In order to increase authenticity while students interact with the VP, the patient's answers are provided in informal "spoken" language—breaks and emotional expressions are included as part of the answer. As in real life, the patient does not always directly answer the question asked. The question might be misunderstood or the patient chooses not to answer the question at all.

Physical examination provides important information for nurses to understand the patient's situation. Physical examination is an assessment of the physical and mental status of the patient. In the box "Observation", the student can gather data (assessment) by observation and inspection of the patient. It is also possible for learners to have access to clinical investigations conducted by other professionals, such as laboratory data and radiology reports. In these VP cases, the result was presented through text and images. The model does allow the use of other types of media such as video clips, but this option was not used in our study.

The box "Measurement scales" includes different assessment instruments used in nursing to systematically collect information to support judgments and decisions. Commonly used measuring scales are the Downton Fall Risk index and the Minimal Nutritional Assessment (MNA) scales.

Collecting data from the patient is not solely the nurse's responsibility. Different health professionals contribute with information to the patient's medical records. Different care providers then use information collected and reported by different health professionals. For example, both the nurse and the physician perform physical examination and patient history taking, but they use different formats and analyze the data differently because of each profession's patient care focus. Therefore the medical records from different health professions include valuable information that can help the nurse to understand the patient's story and determine the strengths of the patient or assess the responses the patient exhibits or could potentially exhibit, as a result of a health problem.

Layer 2: Iterative Clinical Reasoning

After obtaining the patient-centered narrative, the learner starts to frame or derive the meaning and significant issues and identify the patient's nursing care needs by identifying the present state and the desired outcome state. The interplay between the description of the present state and how to attain the outcome state will drive the reasoning. One approach is juxtaposing where specific outcome state criteria are compared with present state data. A test is then created to evaluate the gap between the two states. Through a process of cue logic, juxtaposing, and judgment, the students choose the appropriate interventions that will bring present state and the desired outcome state closer to each other.

At this stage, the students' strategies are important. In VPs, this process is supported by didactic questions that simulate reflection and make thinking skills visible.

Layer 3: Measurable Outcomes

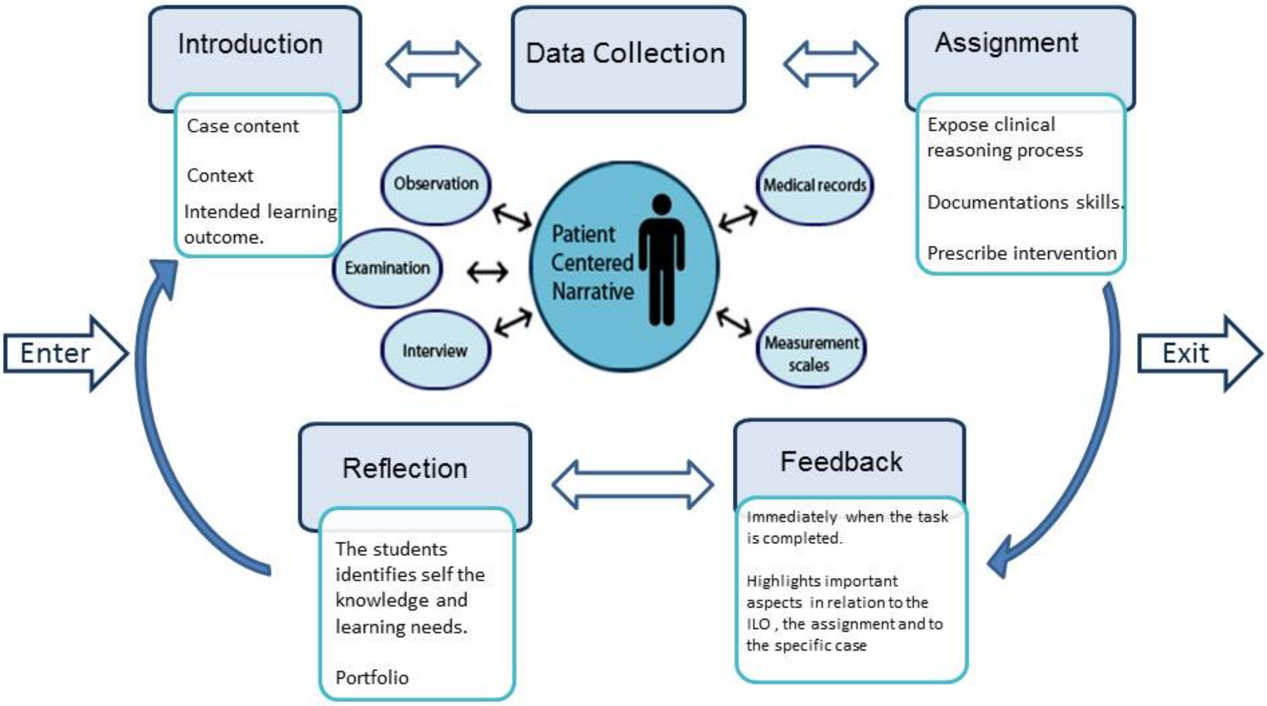
The VP model allows assessment of how the student collects and analyzes data in order to identify and frame nursing care needs and the student's ability to document their reasoning process. The patient's care management can also be exposed in terms of the student's ability to plan for accurate intervention and accurate prioritizing of nursing actions. The model also allows assessment of knowledge levels and the student's ability to reflect on some specific aspects.

The Virtual Patient Nursing Activity Model

In order to be useful, the model needs to be contextualized in an educational setting with specific intended learning and/or assessment outcomes. The virtual patient Nursing Activity Model (vpNAM) contains both aspects that are necessary for a learning activity (see the light blue boxes in Figure 3) and is based on the intended learning outcomes specific instructional strategies. To illustrate the vpNAM, we applied it to the third semester of the undergraduate nursing education program at Karolinska Institutet. The goals of the learning activity in that semester were to define the patients' nursing problems, create a nursing plan, train on documentation, and practice how to prescribe evidence-based nursing intervention.

When the vpNAM is applied to the VP, it means that the student is first introduced to the patient and the aims of the learning activity. The student is then expected to collect relevant data to frame an understanding of the patient's needs and also capture the patient's story. After completing the encounter with the VP, the student is given formative feedback that the case author formulated in advance. The student uses the feedback to reflect and to identify further knowledge and learning needs.

Figure 3. The different aspects in the virtual patient nursing activity model.



Students' Perceptions of the Virtual Patient Learning Activity

We investigated how the VPs, based on the developed models, were perceived by nursing students. After completing two VPs, the participants answered an adapted version of the eViP questionnaire [36] adjusted to fit nursing education. All the students in the course (N=102) were invited to participate in the study. We then excluded those that either did not complete the cases within a given timeframe (first case in 3 weeks and second case in 2 weeks) or chose not to participate in the study. Figure 4 summarizes the process of recruitment of participants.

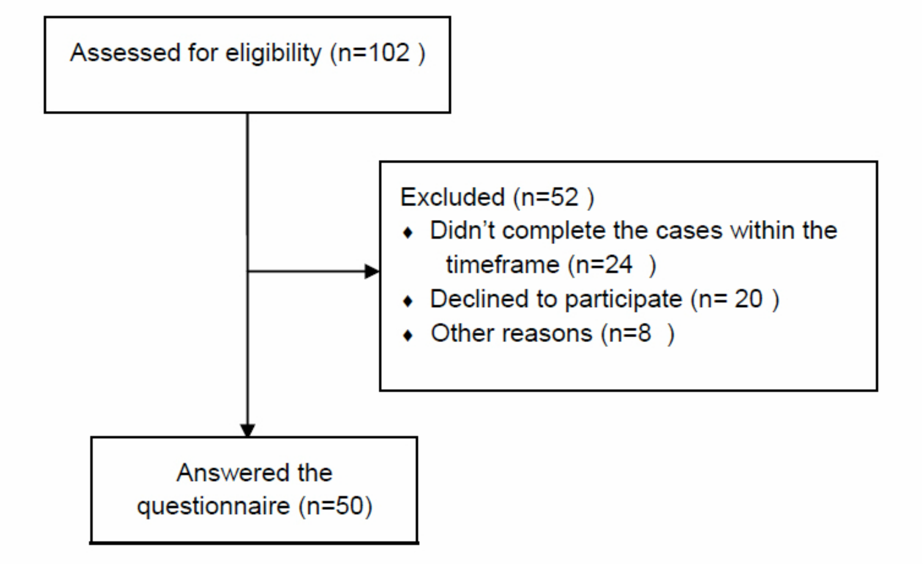
The questionnaire is a 14-item self-reported evaluation that explores five domains: authenticity, professionalism, learning, coaching, and overall judgment. The Cronbach alpha is .88 (greater than .7), which indicates a high level of internal consistency of scale with this specific sample. The construct has a good reliability. Table 1 shows the proportion of students that answered "agree" and "strongly agree" to the statements. Perceived usefulness is presented in a relative manner in order to highlight strengths and weaknesses of the type of VP created.

Table 1. Descriptive statistics and analysis of students’ perceived usefulness (full sample N=50).

Questions	Agree		
	n	%	Perceived usefulness[a]
Authenticity of patient encounter and the consultation			
Q1.While working on this case, I felt I had to make the same decisions a nurse would make in real life.	22	43	+
Q2. While working on this case, I felt I were the nursing student caring for this patient.	24	47	+
Professional approach in the consultation			
Q3. While working through this case, I was actively engaged in gathering the information (eg, history questions, physical exams, lab tests) I needed, to characterize the patient’s nursing problem.	41	80	+++
Q4. While working through this case, I was actively engaged in revising my initial image of the patient’s problem as new information became available.	32	63	++
Q5. While working through this case, I was actively engaged in creating a short summary of the patient’s problem using nursing terms.	26	51	+
Q6. While working through this case, I was actively engaged in thinking about which findings supported or refuted each nursing diagnosis.	31	61	++
Coaching during consultation			
Q7. I felt that the case was at the appropriate level of difficulty for my level of training.	31	61	++
Q8. The questions I was asked while working through this case were helpful in enhancing my diagnostic reasoning in this case	24	47	+
Q9. The feedback I received was helpful in enhancing my diagnostic reasoning in this case.	27	53	+
Q10. The feedback I received was helpful to improve my ability to identify nursing problems and nursing diagnoses	25	49	+
Q11. The feedback I received helped me improve my ability to prescribe nursing intervention.	23	45	+
Learning effect of consultation			
Q12. After completing this case, I feel better prepared to confirm a nursing diagnosis and exclude differential diagnoses in a real life patient with this complaint.	29	60	++
Q13. After completing this case, I feel better prepared to care for a real life patient with this complaint.	23	45	+
Overall judgment of the case workup			
Q14. Overall, working through this case was a worthwhile learning experience.	38	74	+++

[a] The level of perceived usefulness is indicated as medium (+), high (++), and very high (+++).

Figure 4. Flowchart showing the recruitment process of the participants.



Discussion

Principal Findings

A Design Blueprint for Virtual Patients in Nursing Education

The aim of this study was (1) to develop a theory-anchored model for creating virtual patients in nursing education, (2) to investigate how VPs could be instantiated as a learning activity, and finally (3) to explore undergraduate nursing students' perceived usefulness of two virtual patients.

We chose a theory-anchored design blueprint (Figure 2) that provides a structure for the authoring of virtual patients. The virtual patient Design Nursing Model (vpDNM) builds on the 3rdgeneration nursing process and is composed of three main layers. Layer 1 focuses on patient-centric data collection, Layer 2 is about iterative clinical reasoning, and Layer 3 covers measurable outcomes. The nursing process has provided structure and thinking in nursing education since the 1950s, and nurses recognize themselves in this way of thinking and working. However, the nursing process has evolved over time. While the first generation focused on problems and processes, the second generation emphasized diagnosis and reasoning. Finally, the third generation highlighted reflection, outcome specification, and testing given a patient's story [11]. The developed design model reflects the nursing process (observation, assessment, nursing diagnosis, nursing prescription, planning, implementation, and evaluation) because it is recognized as a guideline for nurses in their clinical practice. However, the OPT model differs from the traditional nursing process by highlighting clinical reasoning. It is built on a foundation of reflective judgment. It honors the holistic nature of nursing and approaches patients' situations in terms of outcomes. It identifies the thinking skills and thinking strategies involved in clinical decisions and judgment.

A further advantage is that this model can be used with several taxonomies used in nursing that can provide the content for clinical reasoning [11]. Choosing a different theoretical basis would probably have had an impact on the vpDNM by highlighting other aspects. We acknowledge that a model could hide the complexity of people's health care needs that require the collective competence, skills, knowledge, and actions of different professions. Health care professionals share competencies, knowledge, and skills, but they also make unique contributions to the patient's care [38]. As different professions contribute with different knowledge and skills to the health care environments, the different educational program has to focus on different learning goals. This is why this model is needed for nursing education, but also all contextualized models should be revisited when addressing other health care professions.

In the proposed design model, the nursing process has most influence on Layers 1 and 2. In Layer 1, it influences the way patient data are collected and what type of clinical data was made available. The purpose of data collection is to capture the patient-centered narrative. The learner is given the opportunity to collect both subjective and objective data from the VP. The third generation nursing process also had a substantial effect on Layer 2. In Layer 2, the emphasis is on reflection, outcome specification, and testing given a patient's narrative.

The vpDNM is intended to work as guideline for teachers when they create VPs. The model will also allow the systematic development of different types of virtual patients from a common model and thereby create opportunities for sharing pedagogical designs across technical solutions. Finally, the model might help provide a basis for research that builds on shared empirical findings and best practices.

Virtual Patients in a Learning Activity

An activity model was built using the vpDNM combined with a nursing educational context, specifically, the third semester in an undergraduate nursing program. The intended learning outcomes informed the modeling process (Figure 3). The vpNAM we developed reinforces the increased understanding that VPs are note merely an object/tool but rather a learning activity. The proposed learning activity takes into account the core drivers of learning in medical simulation, which are scenario-based learning, task deliverables, trigger events, data collection guidance/instruction, feedback, and reflection [39]. The findings indicate that a learning activity using VPs is similar to other types of simulation-based learning opportunities that offer the potential to investigate constructive alignments between different types of simulation modalities for both learning and assessment. It also creates an incentive to further investigate the value of VP affordances in relation to the learning objectives and competencies.

Computer-based education and simulation have been proposed as an instructional method to develop nursing students' ability to translate and apply theoretical knowledge in a clinical context [9]. The vpNAM contributes to systematizing the curricular integration that plays a key role in the implementation of new technologies and thereby ensures that the VPs are used effectively and consistently [21].

Students' Perceptions of Global Linear Virtual Patients

We investigated the nursing students' experiences of learning while working with virtual patients, created using the vpNDM and vpNAM, focusing on the development of clinical reasoning skills. The aspects covered authenticity of patient encounter and the consultation, professional approach in the consultation, coaching during consultation, learning effect of consultation, and overall judgment of case workup. The global linear VPs were perceived by students as useful in terms being actively engaged in gathering patient information in order to characterize the patients' nursing problem. Others useful aspects, but to a lesser extent, were the ability to actively engage in revising the patients' problems as new information became available and the fact that VPs triggered diagnostic reasoning.

Conclusions

Virtual patients adapted to the nursing paradigm could support nursing students' development of clinical reasoning skills. Building on theory-anchored models for designing and implementing virtual patients can aid constructive alignment with intended learning and assessment objectives. Furthermore, this approach could strengthen the research in this field. Finally, nursing students perceive the strengths of global linear VPs mainly in the ability to support clinical data gathering and as a worthwhile learning experience, and to a lesser extent, in VPs that actively drive the reasoning process.

Acknowledgments

This project was supported with intramural funds from Karolinska Institutet. The authors thank Sherry Fahimi for revising the figures, Dr Gordon Tait for giving us access to the VIC tool, and the participants of the study for their positive engagement in this research project.

Conflicts of Interest

None declared.

References

1. Swanwick T, Buckley G. In: Swanwick T, editor. Understanding Medical Education: Evidence, Theory and Practice. London: Wiley-Blackwell; 2010.
2. Jenson CE, Forsyth DM. Virtual reality simulation: using three-dimensional technology to teach nursing students. Comput Inform Nurs 2012 Jun;30(6):312-8; quiz 319. [doi: 10.1097/NXN.0b013e31824af6ae] [Medline: 22411409]
3. Benner P, Sutphen M, Leonard V, Day L, Shulman LS. Educating nurses: a call for radical transformation. San Francisco: Jossey-Bass; 2010.
4. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. JAMA 2011 Sep 7;306(9):978-988. [doi: 10.1001/jama.2011.1234] [Medline: 21900138]
5. Huwendiek S, Reichert F, Bosse HM, de Leng BA, van der Vleuten CP, Haag M, et al. Design principles for virtual patients: a focus group study among students. Med Educ 2009 Jun;43(6):580-588. [doi: 10.1111/j.1365-2923.2009.03369.x] [Medline: 19493183]
6. Gaba DM. The future vision of simulation in health care. Qual Saf Health Care 2004 Oct;13 Suppl 1:i2-10 [FREE Full text] [doi: 10.1136/qhc.13.suppl_1.i2] [Medline: 15465951]
7. Gobbi M, Monger E, Weal MJ, McDonald JW, Michaelides D, De Roure D. The challenges of developing and evaluating complex care scenarios using simulation in nursing education. Journal of Research in Nursing 2012 Jul 10;17(4):329-345. [doi: 10.1177/1744987112449969]
8. Salas E, Bowers CA, Rhodenizer L. It is not how much you have but how you use it: toward a rational use of simulation to support aviation training. Int J Aviat Psychol 1998;8(3):197-208. [doi: 10.1207/s15327108ijap0803_2] [Medline: 11541532]
9. World Health Organization. Global standards for the initial education of professional nurses and midwives URL: http://www.who.int/hrh/nursing_midwifery/hrh_global_standards_education.pdf [accessed 2013-03-03] [WebCite Cache ID 6EqM3jSrb]
10. Warne T, Johansson UB, Papastavrou E, Tichelaar E, Tomietto M, Van den Bossche K, et al. An exploration of the clinical learning experience of nursing students in nine European countries. Nurse Educ Today 2010 Nov;30(8):809-815. [doi: 10.1016/j.nedt.2010.03.003] [Medline: 20409620]
11. Pesut D, Herman JA, Herman J. Clinical reasoning: the art and science of critical and creative thinking. Albany: Delmar; 1999.
12. Higgs J. Clinical reasoning in the health professions. Edinburgh: Elsevier, Churchill Livingstone; 2008.
13. Gaberson KB, Oermann MH. Clinical teaching strategies in nursing. New York: Springer Pub; 2010.
14. Thompson C, Yang H. Nurses' decisions, irreducible uncertainty and maximizing nurses' contribution to patient safety. Healthc Q 2009;12 Spec No Patient:e178-e185 [FREE Full text] [Medline: 19667766]
15. Burns N, Grove SK. The practice of nursing research: conduct, critique, and utilization. St. Louis, Mo: Elsevier/Saunders; 2005.
16. Higgs J, Jones MA, Titchen A. Knowledge reasoning and evidence for practice. In: Higgs J, Jones MA, Loftus S, Christensen N, editors. Clinical reasoning in the health professions. Edinburgh: Elsevier, Churchill Livingstone; 2008:151-159.
17. Levett-Jones T, Hoffman K, Dempsey J, Jeong SY, Noble D, Norton CA, et al. The 'five rights' of clinical reasoning: an educational model to enhance nursing students' ability to identify and manage clinically 'at risk' patients. Nurse Educ Today 2010 Aug;30(6):515-520. [doi: 10.1016/j.nedt.2009.10.020] [Medline: 19948370]
18. Banning M. Clinical reasoning and its application to nursing: concepts and research studies. Nurse Educ Pract 2008 May;8(3):177-183. [doi: 10.1016/j.nepr.2007.06.004] [Medline: 17869587]
19. Higuchi KA, Donald JG. Thinking processes used by nurses in clinical decision making. J Nurs Educ 2002 Apr;41(4):145-153. [Medline: 11954966]

20. Narayan S, Corcoran-Perry S. Teaching clinical reasoning in nursing education. In: Higgs J, Jones MA, Loftus S, Christensen N, editors. Clinical reasoning in the health professions. Edinburgh: Elsevier; 2008:405-413.
21. Cook DA, Triola MM. Virtual patients: a critical literature review and proposed next steps. Med Educ 2009 Apr;43(4):303-311. [doi: 10.1111/j.1365-2923.2008.03286.x] [Medline: 19335571]
22. Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: a systematic review and meta-analysis. Acad Med 2010 Oct;85(10):1589-1602. [doi: 10.1097/ACM.0b013e3181edfe13] [Medline: 20703150]
23. Hurst HM, Marks-Maran D. Using a virtual patient activity to teach nurse prescribing. Nurse Educ Pract 2011 May;11(3):192-198. [doi: 10.1016/j.nepr.2010.08.008] [Medline: 20869319]
24. Ellaway RH, Poulton T, Smothers V, Greene P. Virtual patients come of age. Med Teach 2009 Aug;31(8):683-684. [Medline: 19811203]
25. Huwendiek S, De leng BA, Zary N, Fischer MR, Ruiz JG, Ellaway R. Towards a typology of virtual patients. Med Teach 2009 Aug;31(8):743-748. [Medline: 19811212]
26. Ellaway R, Poulton T, Fors U, McGee JB, Albright S. Building a virtual patient commons. Med Teach 2008;30(2):170-174. [doi: 10.1080/01421590701874074] [Medline: 18464142]
27. LeFlore JL, Anderson M, Zielke MA, Nelson KA, Thomas PE, Hardee G, et al. Can a virtual patient trainer teach student nurses how to save lives--teaching nursing students about pediatric respiratory diseases. Simul Healthc 2012 Feb;7(1):10-17. [doi: 10.1097/SIH.0b013e31823652de] [Medline: 22228285]
28. Cendan J, Lok B. The use of virtual patients in medical school curricula. Adv Physiol Educ 2012 Mar;36(1):48-53 [FREE Full text] [doi: 10.1152/advan.00054.2011] [Medline: 22383412]
29. Gesundheit N, Brutlag P, Youngblood P, Gunning WT, Zary N, Fors U. The use of virtual patients to assess the clinical skills and reasoning of medical students: initial insights on student acceptance. Med Teach 2009 Aug;31(8):739-742. [Medline: 19811211]
30. Guise V, Chambers M, Välimäki M. What can virtual patient simulation offer mental health nursing education? J Psychiatr Ment Health Nurs 2012 Jun;19(5):410-418. [doi: 10.1111/j.1365-2850.2011.01797.x] [Medline: 22070549]
31. Kononowicz AA, Narracott AJ, Manini S, Bayley MJ, Lawford PV, McCormack K, et al. A framework for different levels of integration of computational models into web-based virtual patients. J Med Internet Res 2014;16(1):e23 [FREE Full text] [doi: 10.2196/jmir.2593] [Medline: 24463466]
32. Salminen H, Zary N, Björklund K, Toth-Pal E, Leanderson C. Virtual patients in primary care: developing a reusable model that fosters reflective practice and clinical reasoning. J Med Internet Res 2014;16(1):e3 [FREE Full text] [doi: 10.2196/jmir.2616] [Medline: 24394603]
33. Pesut DJ, Herman J. OPT: transformation of nursing process for contemporary practice. Nurs Outlook 1998;46(1):29-36. [Medline: 9513228]
34. Kautz DD, Kuiper R, Pesut DJ, Knight-Brown P, Daneker D. Promoting clinical reasoning in undergraduate nursing students: application and evaluation of the Outcome Present State Test (OPT) model of clinical reasoning. Int J Nurs Educ Scholarsh 2005;2:Article 1. [doi: 10.2202/1548-923x.1052] [Medline: 16646903]
35. Toronto General Hospital Department of Anesthesia. Perioperative Interactive Education Virtual Interactive Case system. 2013. URL: http://pie.med.utoronto.ca/vic/index.htm [accessed 2013-03-03] [WebCite Cache ID 6EqKWUF0Z]
36. Huwendiek S, De Lang B. Student questionnaire concerning their learning and clinical reasoning experiences with virtual patients. URL: http://www.virtualpatients.eu/wp-content/uploads/2009/06/evip_evaluation_tool_form1_student_questionnaire_VPs.pdf [accessed 2014-04-06] [WebCite Cache ID 6OdKoh6kW]
37. Bateman J, Allen ME, Kidd J, Parsons N, Davies D. Virtual patients design and its effect on clinical reasoning and student experience: a protocol for a randomised factorial multi-centre study. BMC Med Educ 2012;12:62 [FREE Full text] [doi: 10.1186/1472-6920-12-62] [Medline: 22853706]
38. Royal College of Nursing. Defining nursing. Nursing is URL: http://www.rcn.org.uk/__data/assets/pdf_file/0003/78564/001983.pdf [accessed 2013-03-03] [WebCite Cache ID 6EqK626r7]
39. Clark RC. Scenario-based e-learning. In: Evidence-Based Guidelines for Online Workforce Learning. San Francisco: Pfeiffer; 2013.

Abbreviations

CR: clinical reasoning
eViP: electronic virtual patients
OPT: Outcome-Present State Test
VIC: virtual interactive case
VP: virtual patient
vpNAM: virtual patient Nursing Activity Model
vpNDM: virtual patient Nursing Design Model

Edited by P Bamidis, D Giordano, C Pattichis, N Zary; submitted 18.03.13; peer-reviewed by T Poulton, M Triola, C Bond; comments to author 18.04.13; revised version received 30.01.14; accepted 21.02.14; published 10.04.14.

Please cite as:
Georg C, Zary N
Web-Based Virtual Patients in Nursing Education: Development and Validation of Theory-Anchored Design and Activity Models
J Med Internet Res 2014;16(4):e105
URL: http://www.jmir.org/2014/4/e105/
doi:10.2196/jmir.2556
PMID:24727709

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org