

Journal of Medical Internet Research

Impact Factor (2018): 4.945 - ranked #1 medical informatics journal by Impact Factor
Volume 15 (2013), Issue 6 ISSN: 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Two Fully Automated Web-Based Interventions for Risky Alcohol Use: Randomized Controlled Trial (e110) Marc-Dennan Tensil, Benjamin Jonas, Evelin Strüber.	3
Cardiovascular Health Effects of Internet-Based Encouragements to Do Daily Workplace Stair-Walks: Randomized Controlled Trial (e127) Lars Andersen, Emil Sundstrup, Marianne Boysen, Markus Jakobsen, Ole Mortensen, Roger Persson.	14
A Service-Based Evaluation of a Therapist-Supported Online Cognitive Behavioral Therapy Program for Depression (e121) John Sharry, Ruth Davidson, Orla McLoughlin, Gavin Doherty.	26
Development of an Obesity Management Ontology Based on the Nursing Process for the Mobile-Device Domain (e130) Hyun-Young Kim, Hyeoun-Ae Park, Yul Min, Eunjo Jeon.	41
Collecting Maternal Health Information From HIV-Positive Pregnant Women Using Mobile Phone-Assisted Face-to-Face Interviews in Southern Africa (e116) Alastair van Heerden, Shane Norris, Stephen Tollman, Linda Richter, Mary Rotheram-Borus.	52
Comparison of Physical Activity Measures Using Mobile Phone-Based CalFit and Actigraph (e111) David Donaire-Gonzalez, Audrey de Nazelle, Edmund Seto, Michelle Mendez, Mark Nieuwenhuijsen, Michael Jerrett.	67
Crowdsourcing Participatory Evaluation of Medical Pictograms Using Amazon Mechanical Turk (e108) Bei Yu, Matt Willis, Peiyuan Sun, Jun Wang.	97
Patients' Experiences of Helpfulness in Guided Internet-Based Treatment for Depression: Qualitative Study of Integrated Therapeutic Dimensions (e126) Kjersti Lillevoll, Maja Wilhelmsen, Nils Kolstrup, Ragnhild Høifødt, Knut Waterloo, Martin Eisemann, Mette Risør.	109
Incidence of Online Health Information Search: A Useful Proxy for Public Health Risk Perception (e114) Bo Liang, Debra Scammon.	121
Postmarket Drug Surveillance Without Trial Costs: Discovery of Adverse Drug Reactions Through Large-Scale Analysis of Web Search Queries (e124) Elad Yom-Tov, Evgeniy Gabrilovich.	142
Collection and Visualization of Dietary Behavior and Reasons for Eating Using Twitter (e125) Melanie Hingle, Donella Yoon, Joseph Fowler, Stephen Kobourov, Michael Schneider, Daniel Falk, Randy Burd.	154

Using Online Health Communities to Deliver Patient-Centered Care to People With Chronic Conditions (e115)	
Martijn van der Eijk, Marjan Faber, Johanna Aarts, Jan Kremer, Marten Munneke, Bastiaan Bloem.	170
Video Consultation Use by Australian General Practitioners: Video Vignette Study (e117)	
Moyez Jiwa, Xingqiong Meng.	180
Two New Meal- and Web-Based Interactive Food Frequency Questionnaires: Validation of Energy and Macronutrient Intake (e109)	
Sara Christensen, Elisabeth Möller, Stephanie Bonn, Alexander Ploner, Antony Wright, Arvid Sjölander, Olle Bälter, Lauren Lissner, Katarina Bälter.	193
Silence is Golden: Effect of Encouragement in Motivating the Weak Link in an Online Exercise Video Game (e104)	
Brandon Irwin, Deborah Feltz, Norbert Kerr.	208
The Conceptual and Practical Ethical Dilemmas of Using Health Discussion Board Posts as Research Data (e112)	
Carol Bond, Osman Ahmed, Martin Hind, Bronwen Thomas, Jaqui Hewitt-Taylor.	218
Utilization and Perceived Problems of Online Medical Resources and Search Tools Among Different Groups of European Physicians (e122)	
Marlene Kritz, Manfred Gschwandtner, Veronika Stefanov, Allan Hanbury, Matthias Samwald.	228
Virtual Reality for Enhancing the Cognitive Behavioral Treatment of Obesity With Binge Eating Disorder: Randomized Controlled Study With One-Year Follow-up (e113)	
Gian Cesa, Gian Manzoni, Monica Bacchetta, Gianluca Castelnovo, Sara Conti, Andrea Gaggioli, Fabrizia Mantovani, Enrico Molinari, Georgina Cárdenas-López, Giuseppe Riva.	254

Review

Mobile Health Applications for the Most Prevalent Conditions by the World Health Organization: Review and Analysis (e120)	
Borja Martínez-Pérez, Isabel de la Torre-Díez, Miguel López-Coronado.	78

Viewpoint

Community Management That Works: How to Build and Sustain a Thriving Online Health Community (e119)	
Colleen Young.	239

Letter to the Editor

The Systematic Removal of Participants Post-Randomization Can Lead to Alternate Explanations of the Results (e128)	
John Cunningham.	267

Original Paper

Two Fully Automated Web-Based Interventions for Risky Alcohol Use: Randomized Controlled Trial

Marc-Dennan Tensil^{1*}, Dipl-Psych; Benjamin Jonas^{1*}, Dipl-Psych; Evelin Strüber², Dipl-Soz

¹Delphi-Gesellschaft, Berlin, Germany

²Federal Centre for Health Education (BZgA), Cologne, Germany

*these authors contributed equally

Corresponding Author:

Marc-Dennan Tensil, Dipl-Psych

Delphi-Gesellschaft

Kaiserdamm 8

Berlin, 14057

Germany

Phone: 49 30 39 40 97 85

Fax: 49 30 39 40 97 89

Email: tensil@delphi-gesellschaft.de

Abstract

Background: Excessive alcohol use is a widespread problem in many countries, especially among young people. To reach more people engaging in high-risk drinking behaviors, a number of online programs have been developed in recent years. Change Your Drinking is a German, diary-based, fully automated alcohol intervention. In 2010, a revised version of the program was developed. It is more strongly oriented to concepts of relapse prevention than the previous version, includes more feedback, and offers more possibilities to interact with the program. Moreover, the program duration was extended from 10 to 14 days.

Objective: This paper examines whether the revised version of Change Your Drinking is more effective in reducing alcohol consumption than the original version.

Methods: The effectiveness of both program versions was compared in a Web-based, open, randomized controlled trial with follow-up surveys 6 weeks and 3 months after registration. Participants were recruited online and were randomly assigned to either the original or the revised version of Change Your Drinking. The following self-assessed outcomes were used: alcohol use days, alcohol intake in grams, the occurrence of binge drinking and risky drinking (all referring to the past 7 days prior to each survey), and the number of alcohol-related problems.

Results: A total of 595 participants were included in the trial. Follow-up rates were 58.0% after 6 weeks and 49.6% after 3 months. No significant group differences were found in any of the outcomes. However, the revised version was used by more participants (80.7%) than the original version (55.7%). A significant time effect was detected in all outcomes (alcohol use days: $P=.002$; alcohol intake in grams: $P<.001$; binge drinking: $P<.001$; alcohol-related problems: $P=.004$; risky drinking: $P<.001$).

Conclusions: The duration and complexity of the program played a minor role in reducing alcohol consumption. However, differences in program usage between the versions suggest the revised version was more attractive to participants.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN): 31586428; <http://www.controlled-trials.com/ISRCTN31586428/> (Archived by WebCite at <http://www.webcitation.org/6BFxApCUT>)

(*J Med Internet Res* 2013;15(6):e110) doi:[10.2196/jmir.2489](https://doi.org/10.2196/jmir.2489)

KEYWORDS

alcohol abuse; binge drinking; Internet intervention; relapse prevention; randomized controlled trial

Introduction

Excessive alcohol use is associated with numerous health and social consequences and represents a major challenge for public health activities. It is estimated that approximately 4% of all

deaths worldwide are due to the consumption of alcohol [1]. In Germany, more than 70,000 people die from alcohol use-related consequences each year [2].

Typically, people are introduced to alcohol at a young age, which makes the earliest possible prevention a priority. In

Germany, the average age of the first binge drinking experience is between 15 and 16 years [3]. To a certain extent, experimenting with alcohol can be regarded as a normal part of adolescence [4,5]. However, regular binge drinking not only represents a direct health risk, it also can impair brain development substantially [6,7].

In Germany, the prevalence of excessive drinking is highest among young adults. Forty-two percent of those aged between 18 and 25 years engage in binge drinking at least once a month [3]; 40% of those aged between 18 to 20 years meet the criteria for problematic alcohol use as defined by the Alcohol Use Disorders Identification Test (AUDIT) [8]. Despite high levels of problematic alcohol use, few young people seek professional help. Globally, the World Health Organization (WHO) estimates that 78% of those in need for treatment due to alcohol abuse or dependence remain untreated although effective intervention methods are available [9].

Internet-based self-help programs can help to reduce this gap. Their advantages include easy accessibility and discreteness; therefore, they provide an appealing alternative especially to those who would abstain from face-to-face treatment because of fear of exposure or embarrassment. Meta-analyses show that online self-help programs have a rather small effect size, but because of their scalability they are a cost-effective way to reduce alcohol consumption in the population [10,11]. Online self-help alcohol interventions have been shown to be effective in the general adult population [12-15] as well as in college student samples [16-18], underage drinkers [19], and with young people in the workplace setting [20].

Since 2009, the German Federal Centre for Health Education, Bundeszentrale für gesundheitliche Aufklärung (BZgA), has offered Change Your Drinking, a free online self-help program for young adults with problematic alcohol use. Based on cognitive behavioral principles, it provides a consumption diary for 10 days and 1 brief tailored feedback at the end of this period. Statements from users previously collected in an unpublished Web survey, however, indicated a need for further and more elaborate feedbacks. Moreover, some users regarded the consumption diary as too simple. Based on this information, the intervention was revised and extended to 14 days, aiming to involve the users more deeply into the program and to provide them with a more sophisticated version of the consumption diary. These goals reflect recent findings according to which elaborated self-help alcohol interventions spanning over different sessions are more effective than single-session feedback interventions [12,13], and a higher degree of interactivity is associated with higher effect sizes [21]. Moreover, Web-based interventions with a higher intended usage were found to be more likely to be adhered to [22].

This paper examines whether the revised version of Change Your Drinking is more effective in reducing alcohol consumption than the original version. Outcomes were the alcohol use days, the total intake in grams, the occurrence of binge drinking (yes/no), and binge drinking (yes/no) for the past 7 days prior to each survey as well as the number of alcohol-related problems.

Methods

Study Design

A randomized controlled trial (RCT) was conducted to compare the 2 versions of Change Your Drinking with follow-up surveys 6 weeks and 3 months after registration. Participants were invited by email and reimbursed with a €10 shopping voucher for their participation. The study was approved by the ethics committee of the Department of Applied Human Sciences at the University of Magdeburg-Stendal (Ref 4973-15) and was registered with Current Controlled Trials (ISRCTN: 31586428).

The intervention and the study were purely Web-based on the addiction prevention website Drugcom [23] and the alcohol prevention website Kenn dein Limit (Know Your Limit) [24] both run by the BZgA. Recruitment of participants started in December 2010 and ended in March 2012. All users of the freely accessible Check Your Drinking self-assessment were invited for the study if they met the eligibility criteria described subsequently. The results of the self-assessment were used both for the Change Your Drinking program and as baseline data for the trial.

If the eligibility criteria (see Study Criteria) were met, users were informed about the study. A portable document format (PDF) file containing all relevant study details was offered for download (see [Multimedia Appendix 1](#)). Users who were willing to participate were then asked to register and to provide their informed consent by clicking an "I agree to participate" button. After successfully confirming their email addresses, participants were randomly assigned either to the original (version 1) or the revised program (version 2) by random number generator software. Researchers could not influence nor predict the randomization result. The participants were blind to the results of the randomization because they only received detailed information about the program version they were allocated to.

Those who opted not to participate in the study or who did not meet the eligibility criteria had full access to the original version of the program and were not included in any follow-up surveys.

Measures

Trial data were collected via self-assessment in the baseline survey, as well as 6 weeks and 3 months afterwards. The past 7 days were used as reference period for alcohol consumption. In order to quantify the alcohol intake, participants were first asked to indicate their number of alcohol use days in the previous 7 days. Afterwards, they were requested to specify the number and type of alcoholic beverages consumed on each drinking day. Using these details, the amount of pure alcohol and the number of standard glasses (SG) per day was calculated. According to the BZgA, 1 SG in Germany corresponds to 10 grams of pure alcohol [25]. If 5 or more SG were consumed on any of the previous 7 days, this was classified as binge drinking.

Another outcome was risky consumption in the past 7 days, as defined by the following factors: (1) an average of more than 24 or 12 g (for males and females, respectively) of pure alcohol per day (these are the tolerable upper alcohol intake levels for males and females in Germany according to Burger and colleagues [26]), or (2) more than 5 days of consumption, or

(3) at least 1 incident of binge drinking in the reference period. The definition of risky alcohol use has been derived from the guidelines for low-risk consumption as defined by the BZgA [27].

Alcohol-related problems were measured by a German version of McGee and Kypri's [28] Alcohol Problems Scale (APS). The scale consists of 14 items describing negative consequences of alcohol consumption, such as vomiting, unprotected sexual intercourse, or blackouts, during the last 30 days. Each item is to be answered with yes, no, or "no answer."

The AUDIT was used as part of the baseline survey to test for the study criteria. A cut-off of 8 points as suggested by Babor et al [29] was used to define risky alcohol consumption. As a measure of the program usage we tracked the diary usage (used at least once: yes/no).

Study Criteria

To be invited for the study, participants had to be Internet literate, at least 18 years old, and had either reached the cut-off of 8 points in AUDIT or have consumed more than 24/12 g (male/female) of pure alcohol per day on average in the past week. We did not use stricter study criteria (eg, exclusion of individuals currently in other treatment) because we wanted to increase the generalizability of the results to the regular users of Change Your Drinking.

Interventions

Change Your Drinking is an Internet-based self-help program based entirely on automated tailored feedbacks. A comparison of both versions of the intervention is displayed in Table 1. Exemplary feedbacks of each version in German language are included in Multimedia Appendix 2.

Table 1. Comparison of both Change Your Drinking versions.

Characteristics	Version 1	Version 2
Aim	No alcohol use or low-risk use ^a	No alcohol use or low-risk use ^a
Duration	10 days	14 days
Interventions	Detailed tailored feedback at baseline with advice regarding the participant's alcohol use (Check Your Drinking)	Detailed tailored feedback at baseline with advice regarding the participant's alcohol use (Check Your Drinking)
	General information on control strategies	General information on control strategies
	10-day alcohol use diary	14-day alcohol use diary including a tool to help developing control strategies
		Daily: Short tailored feedback on the individual's alcohol use and graphical display of alcohol use
		On day 7: Detailed tailored feedback on one's alcohol use and tips on how to cope with risk situations
	On day 10: Tailored feedback on the individual alcohol consumption	On day 14: New detailed tailored feedback on one's alcohol use and tips on how to cope with risk situations
		Advice to reflect one's reasons for reducing or abstaining from alcohol
		Advice to reward oneself for achieving the personal goal

^aLow-risk use was defined as (1) no more than 24/12 g (male/female) of pure alcohol per day [26], (2) no more than 5 alcohol use days, and (3) no incidents of binge drinking in past 7 days [25].

Original Version (Version 1)

Change Your Drinking is based on solution-focused brief intervention and methods of cognitive behavioral therapy. Berg and Miller's [30] solution-focused treatment approach concentrates on achieving concrete behavioral goals within relatively narrow timeframes.

The first step in the program is a self-assessment tool (Check Your Drinking) which provides users with tailored feedback on their drinking behavior. If indicated, users are recommended to participate in the Change Your Drinking program.

At the beginning, participants of Change Your Drinking are to choose a use-related goal which must fall within the limits for low-risk consumption (for a definition of low-risk use see Table 1). Afterwards, the participants are given access to an online diary to keep track of their alcohol use over the next 10 days. Static information on alcohol control strategies is also given.

The program aims to develop self-awareness and self-regulatory skills [31,32].

On the tenth day, participants receive feedback based on their current alcohol use and on how well the use-related goal was met. The feedback is based on Miller and Rollnick's principles of Motivational Interviewing [33].

Revised Version (Version 2)

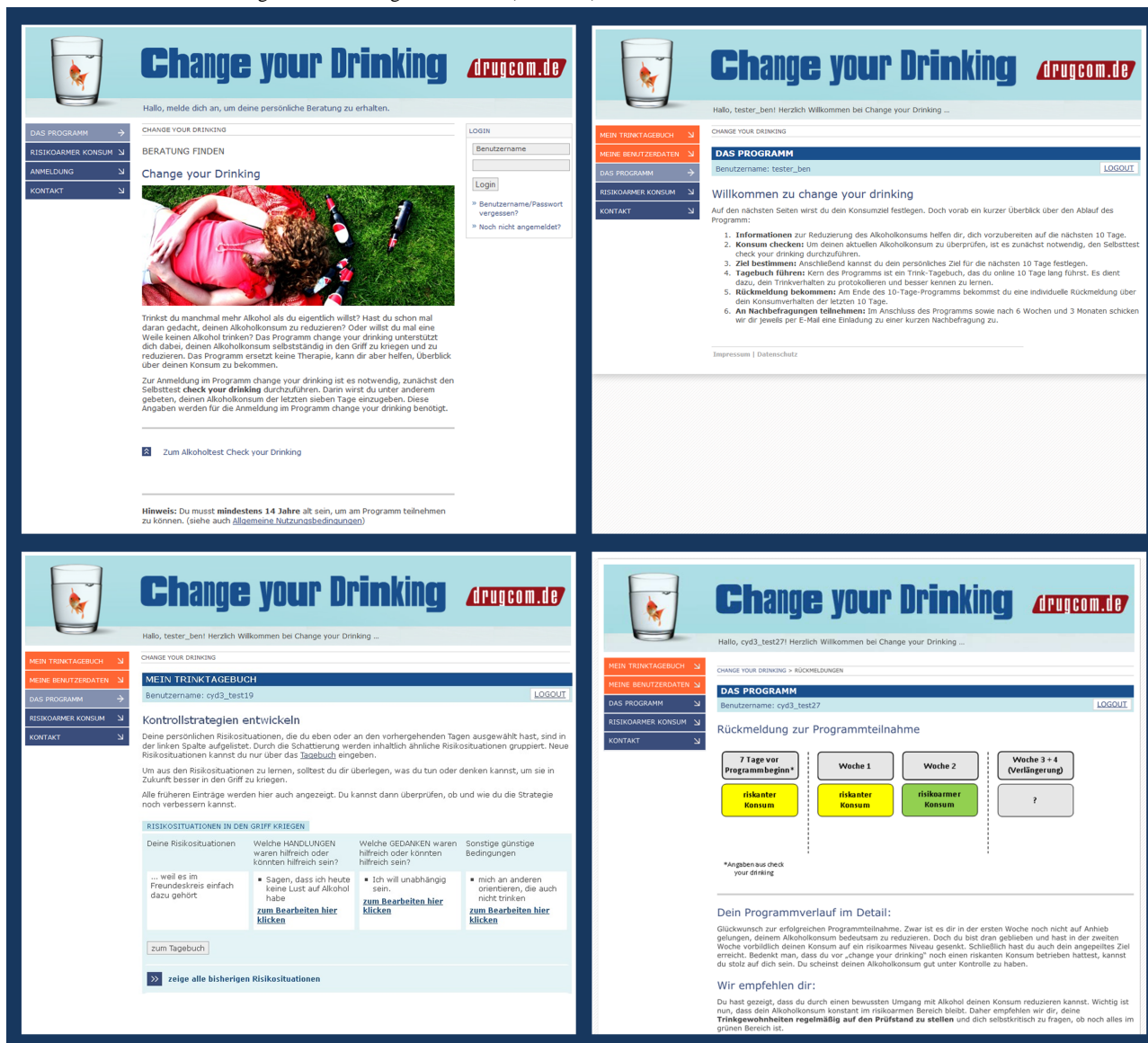
To promote confrontation with one's own consumption pattern, the original version of Change Your Drinking was revised and supplemented with new modules. Based on the principles of relapse prevention [34,35], participants are now asked daily to confront their risk situations and to develop or refine control strategies. Short and motivating feedbacks are provided to reinforce the reflection of one's own alcohol use.

Moreover, 2 tailored and motivating feedbacks after 7 and 14 days were introduced in the intervention, thus extending the

program's length from 10 to 14 days. Both feedbacks address the participant's current consumption levels, comparing those with the initial values (as reported in the Check Your Drinking self-assessment tool), and the previously chosen use-related goals. Tailored tips to cope with risk situations are also provided

in the feedbacks. In addition, data entered in the diary is displayed graphically to give participants a quick look at changes in their alcohol use. Screenshots of version 2 of the intervention are shown in [Figure 1](#).

Figure 1. Screenshots of the Change Your Drinking intervention (version 2).



Statistical Analysis

The comparison of both program versions was conducted with generalized estimating equations (GEE) using Stata 11 (StataCorp LP, College Station, TX, USA), accounting for all 3 data collection points. The GEE analyses were modeled with unstructured correlation matrices between the data collection points. We assumed to have binary distributed data (binge drinking yes/no), Poisson-distributed data (alcohol use days and alcohol-related problems), and a negative binomial distribution with log link in the alcohol intake data (intake in grams).

In the GEE analyses, a group difference was assumed in case of a significant group \times time interaction. To measure the development of both groups, the time effect of each outcome

was examined. In case of any statistically significant result, Cohen's d was calculated.

In a first step of data analysis, we tested whether group differences at baseline or the program usage (operationalized by online diary used: yes/no) moderated the effects of the group assignment on the study outcomes. In case of significance, the respective measure and its interaction with the group factor were included in the GEE; otherwise, it was not considered in the effectiveness testing.

For the intention-to-treat analyses (ITT), missing data was estimated by multiple imputations with Stata's "ICE" command. We performed 10 imputations. The results from the multiple imputations were compared with completer analyses and last observation carried forward (LOCF) analyses. In the completer analyses, missing follow-up data was not imputed, so only those

cases that provided follow-up data were analyzed. In the LOCF analyses, missing data was replaced with data from the preceding data collection point.

To compare both groups at study baseline and to determine whether baseline measures were predicting follow-up participation, logistic regression analyses were performed. In all analyses, we used a 2-sided significance level of $\alpha=.05$. The study was powered to detect a group difference of $d\geq 0.20$. Therefore, we aimed for a sample size of $N=624$ ($\alpha=.05$; power=0.80).

The research is reported in accordance with the E-CONSORT checklist [36] (see [Multimedia Appendix 3](#)).

Results

Flow of Participants

Each time the Change Your Drinking starting page was opened, it was checked whether data for the Check Your Drinking self-assessment were available and whether the study criteria were fulfilled or not. In the trial period, Check Your Drinking was completed 10,887 times. A total of 5823 cases did not meet the study criteria and 4469 users refused to participate. Thus, 595 persons were included in the trial and were randomized, resulting in 2 approximately equally sized groups (see [Figure 2](#)). Participation in the trial ($N=595$) compared to refusing to participate ($n=4469$) was predicted by female gender (odds ratio [OR] 1.68, 95% CI 1.40-2.00, $P<.001$) and higher education (OR 1.39, 95% CI 1.22-1.58, $P<.001$). Thus, in the group of participants, females (38.8%) and individuals with high school education (65.0%) were more highly represented than in the group of nonparticipants (females: 27.4%, participants with high school education: 52.9%). Significant group differences were also found for the utilization of professional treatment (OR 1.83, 95% CI 1.10-3.03, $P=.02$) and age (OR 0.99, 95% CI 0.99-1.00, $P=.04$). That is, participants in the trial tended to use professional help more often (10.3% vs 5.9%) and were slightly younger (between-group $d=0.09$) than individuals who refrained from participating. Baseline alcohol use was not associated with trial participation (use days: OR 1.00, 95% CI 0.96-1.04, $P=.95$; intake in grams: OR 1.00, 95% CI 1.00-1.00, $P=.16$).

In total, 345 persons participated in the first follow-up and 295 in the second, resulting in follow-up rates of 58.0% and 49.6%, respectively. Although loss to follow-up was not predicted by group allocation (OR 0.82, 95% CI 0.56-1.20, $P=.31$), the level

of education, program usage, and alcohol use were significant predictors. Thus, those who took part in the follow-up surveys were more highly educated (OR 1.36, 95% CI 1.14-1.61, $P=.001$), used the diary more often (OR 6.56, 95% CI 4.39-9.81, $P<.001$), and consumed less alcohol (OR 0.97, 95% CI 0.94-1.00, $P=.04$). However, we do not expect any significant bias on that account because these measures were all included in the equations of the multiple imputations.

Sample Description

There were no significant group differences at baseline (see [Table 2](#)). Mean age of participants was approximately 30 years and most participants (364/595, 61.2%) were male. The education level of the participants was relatively high, with 387 of 595 (65.0%) having attended high school. A total of 534 of 595 (89.7%) of the participants were currently not using any other professional help to deal with their alcohol use.

In contrast to the baseline measures, the usage of the diary was significantly higher in the revised version of the program (OR 3.33, 95% CI 2.30-4.81; $P<.001$). According to the analysis plan, the diary usage was included in the first step of the effectiveness analyses.

Effectiveness Results

Diary usage did not moderate the effects of the group assignment on any of the study outcomes (alcohol use days: beta = -0.04, 95% CI -0.22 to 0.14, $P=.66$; alcohol intake: beta = -0.01, 95% CI -0.13 to 0.11, $P=.85$; binge drinking: beta = 0.27, 95% CI -0.32 to 0.85, $P=.37$; alcohol-related problems: beta = 0.04, 95% CI -0.11 to 0.19, $P=.56$; risky drinking: beta = -0.49, 95% CI -1.02 to 0.03, $P=.07$); therefore, it was removed from the analysis.

There were no significant group differences in any of the primary outcomes of the study (see [Table 3](#)). Thus, the number of alcohol use days, the alcohol intake, the frequency of binge drinking, the number of use-related problems, and the indicators for risky consumption follow a similar direction in both groups.

However, significant overall reductions in alcohol use can be noted. Thus, after 3 months, participants in both program versions consumed on average 1.2 days ($d=0.46$) and 133.3 grams ($d=0.59$) less than at baseline. A significant reduction in alcohol-related problems ($d=0.36$), binge drinking incidents (reduction of 31.1%), and risky drinking (reduction of 23.6%) were seen. All the results were confirmed through completer and LOCF analyses.

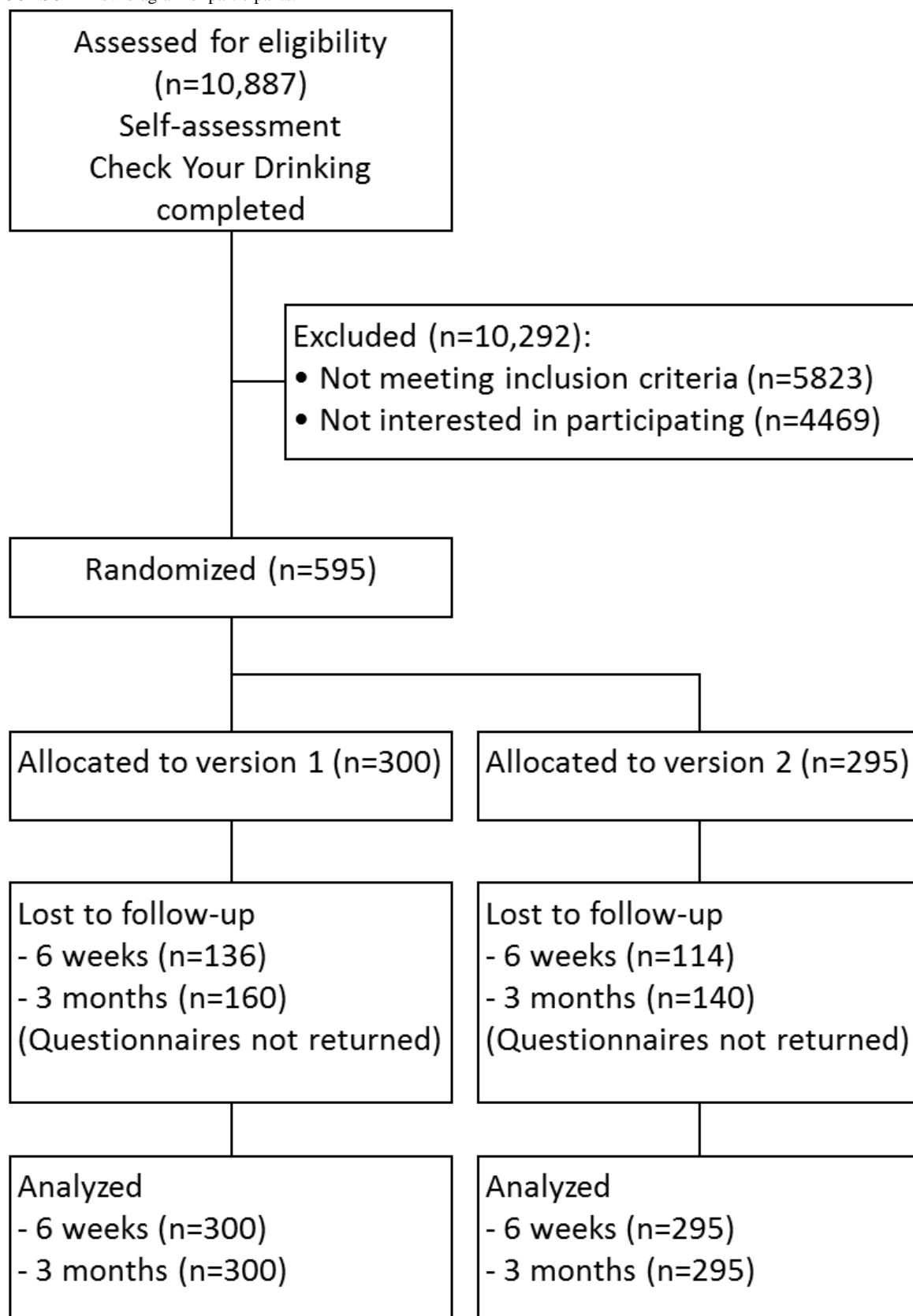
Figure 2. CONSORT flow diagram of participants.

Table 2. Participants' characteristics at baseline and usage of the Change Your Drinking program (N=595).

Variable	Version 1 (n=300)	Version 2 (n=295)
Gender, n (%)		
Male	177 (59.0)	187 (63.4)
Female	123 (41.0)	108 (36.6)
Age, mean (SD)	29.8 (10.3)	29.0 (9.4)
Currently no professional help, n (%)	266 (88.7)	268 (90.8)
Educational level, n (%)		
Basic school (Hauptschule)	25 (8.3)	23 (7.8)
Middle school (Realschule)	57 (19.0)	64 (21.7)
High school (Gymnasium)	203 (67.7)	184 (62.4)
Other school	15 (5.0)	24 (8.1)
Alcohol use		
Alcohol use days ^a , mean (SD)	4.1 (1.8)	4.3 (1.9)
Alcohol intake ^a (g), mean (SD)	313.4 (193.0)	318.5 (194.4)
Binge drinking ^a , n (%)	278 (92.7)	271 (91.9)
Alcohol-related problems, mean (SD)	2.4 (1.8)	2.3 (1.9)
Usage of the program diary, n (%)		
Diary used at least once	167 (55.7)	238 (80.7)
Diary used on all 10/14 days	127 (42.3)	163 (55.3)

^aDuring the past 7 days.**Table 3.** Effectiveness results^a of study primary outcomes.

Outcome	Version 1 (n=300)			Version 2 (n=295)			Group × time interaction ^b		Main effect of time	
	Baseline	6 weeks	3 months	Baseline	6 weeks	3 months	Beta (95% CI)	P value	Beta (95% CI)	P value
Alcohol use ^c (days), mean (SD)	4.1 (1.8)	2.9 (3.1)	2.9 (4.1)	4.3 (1.9)	3.0 (2.9)	3.0 (2.6)	0.01 (−0.09, 0.11)	.88	−0.18 (−0.27, −0.08)	.002
Alcohol intake ^c (grams), mean (SD)	313.4 (193.0)	162.2 (171.6)	183.3 (272.8)	318.5 (194.4)	151.8 (166.6)	181.9 (240.7)	−0.01 (−0.14, 0.11)	.84	−0.14 (−0.22, −0.07)	<.001
Binge drinking ^c (yes), n (%)	278 (92.7)	190 (63.4)	194 (64.5)	271 (91.9)	178 (60.4)	171 (57.8)	−0.11 (−0.39, 0.17)	.44	−0.69 (−0.92, −0.47)	<.001
Alcohol-related problems ^d , mean (SD)	2.4 (1.8)	1.6 (2.1)	1.5 (3.2)	2.3 (1.9)	1.5 (1.9)	1.5 (2.3)	0.04 (−0.11, 0.19)	.56	−0.24 (−0.39, −0.09)	.004
Risky drinking ^c (yes), n (%)	300 (100.0)	231 (77.1)	238 (79.3)	295 (100.0)	215 (72.9)	216 (73.4)	−0.10 (−0.47, 0.27)	.59	−0.64 (−0.92, −0.37)	<.001

^aITT analyses following multiple imputation. Results of complete case and LOCF analyses can be found in [Multimedia Appendix 4](#).^bComparison between version1 and 2 was conducted with the group × time interaction.^cDuring the past 7 days.^dDuring the past 30 days.

Discussion

This study examined the effectiveness of the revised version of the fully automated alcohol intervention Change Your Drinking as compared to the original version of the program. The revised version lasts 14 days instead of 10 days, contains more feedback and interaction options, and is more strongly oriented toward the concept of relapse prevention [34,35]. However, in terms of drinking days, alcohol intake, and other use-related outcomes, the revised version did not yield superior results compared to the original version of the program. Instead, users of both versions reduced their use behavior in a similar way. For example, risky alcohol consumption was reduced in both groups by 23.6% after 3 months. Although trial participants differed from regular users of Change Your Drinking in terms of gender and education level, we assume that the results can largely be generalized to all users of the program.

The results suggest that the higher degree of interactivity, the improved feedback, the stronger emphasis on relapse prevention, and the program extension were not sufficient to significantly enhance the program effects. This is consistent with other findings that the effectiveness of a Web-based alcohol intervention is not increased by more feedback [17] or by the provision of a more elaborate and tailored intervention [37]. The findings are also reflected in the meta-analysis of Rooke and colleagues [10] on the effectiveness of computer-delivered alcohol and tobacco interventions. In this study, no association between the number of intervention sessions and the emphasis on relapse prevention and the intervention's efficacy was found [10]. These results suggest that the effectiveness of self-help alcohol interventions can only be increased to a very limited extent by these features. To significantly enhance their effectiveness it might be necessary to include additional means of interaction, such as personal support by a counselor or therapist [14,38]. Moreover, it has to be noted that Web-based trials using an active comparison group (like ours) often yield very limited effects [10].

These results might question the time and costs spent developing elaborate automated programs. However, if an intervention is less sophisticated, it probably will be less attractive and, thus, be utilized by less people. This point is supported by the usage statistics of Change Your Drinking because the more elaborate version was used by more participants than the original version.

Limitations

A key limitation to this study was the reliance on self-reported data. However, this type of data has repeatedly been shown to be reliable and valid [39,40]. Moreover, verification with urine samples or clinical interviews was not a practical option considering the widely scattered sample.

Because of the financial compensation given for follow-up participation, it cannot be ruled out that several participants tried to register more than once for the study. Although technical measures were taken to prevent this from happening, the

anonymous study setting allowed a participant to sign up with different email addresses. However, we do not expect any bias in favor of any group due to this reason because multiple registrations (if any) presumably were equally distributed among both study groups.

Moreover, it should be noted that a significant number of participants did not take part in the final follow-up surveys. Therefore, we included all relevant participant data in the multiple imputations to estimate their follow-up data and crosschecked those results with completer analyses and LOCF analyses which came to very similar results.

To elucidate the effects of the interventions more thoroughly, it would have been desirable to collect additional data on their acceptability and usability. However, to keep the questionnaires short and the respondent burden as low as possible, we did not include these outcomes in this survey. A future evaluation study will deal with this topic in detail.

We did not include a no-intervention control group; therefore, we cannot determine the *de facto* effectiveness of the intervention. A comparison with other RCTs on Web-based alcohol interventions can provide only rough hints whether the reduction of alcohol use in the Change Your Drinking intervention exceeds mere trial participation. The overall reduction of alcohol intake ($d=0.59$) is considerably stronger than the effects of no-intervention control groups in other trials [20,41] (the difference of weekend drinking between baseline and 30-day follow-up in the trial of Dumas et al [20] and the difference of typical weekly drinking between baseline and 3-month follow-up in the trial of Cunningham et al [41], where effect sizes below $d=0.10$ were achieved). However, other RCTs on Web-based alcohol interventions show that nontreated groups can attain considerable performances up to $d=0.40$ (the difference of drinks last week between baseline and 3-month follow-up in the trial of Blankers et al [14] and the difference of alcohol intake in the past 7 days between baseline and 3-month follow-up in the trial of Jonas et al [42]). So these comparisons only suggest that both versions of Change Your Drinking to be somewhat effective, which leads to no clear conclusion. Hence, we cannot say whether the alcohol use reductions found in both groups are a consequence of the program participation, response bias, regression to the mean, or spontaneous remission. Nevertheless, since the original version of the intervention is freely accessible on the study website, a no-intervention control group was not feasible.

Conclusion

The revised version of Change Your Drinking is not more effective in reducing alcohol use than the original version of the program. The extension of the program and its stronger elaboration presumably played a minor role in reducing alcohol consumption. However, analyses of the program usage suggest that it may be more attractive to participants. Acceptability and usability of both program versions will be examined in a future study.

Acknowledgments

The study was funded by the Federal Centre for Health Education (BZgA).

Conflicts of Interest

Marc-Dennan Tensil and Benjamin Jonas are researchers at Delphi Gesellschaft, which developed Check Your Drinking and Change Your Drinking on behalf of the BZgA. Evelin Strüber is a research consultant at the BZgA.

Multimedia Appendix 1

Information about the trial (German).

[[PDF File \(Adobe PDF File\), 86KB - jmir_v15i6e110_app1.pdf](#)]

Multimedia Appendix 2

Exemplary feedbacks of both program versions.

[[PDF File \(Adobe PDF File\), 221KB - jmir_v15i6e110_app2.pdf](#)]

Multimedia Appendix 3

CONSORT-EHEALTH checklist (V1.6.2) [36].

[[PDF File \(Adobe PDF File\), 996KB - jmir_v15i6e110_app3.pdf](#)]

Multimedia Appendix 4

Results of the complete case & LOCF analyses.

[[PDF File \(Adobe PDF File\), 7KB - jmir_v15i6e110_app4.pdf](#)]

References

1. Rehm J, Mathers C, Popova S, Thavorncharoensap M, Teerawattananon Y, Patra J. Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders. *Lancet* 2009 Jun 27;373(9682):2223-2233. [doi: [10.1016/S0140-6736\(09\)60746-7](#)] [Medline: [19560604](#)]
2. DHS. Jahrbuch Sucht. Lengerich: Pabst; 2012.
3. BZgA. Die Drogenaffinität Jugendlicher in der Bundesrepublik Deutschland 2011. Cologne, Germany: Bundeszentrale für gesundheitliche Aufklärung; 2012 Feb. URL: <http://www.bzga.de/pdf.php?id=04214daa80cc48c7afcdb7d64baf60db> [accessed 2012-11-21] [WebCite Cache ID 6CLH8tplX]
4. Leppin A. Alkoholkonsum und Alkoholmissbrauch bei Jugendlichen: Entwicklungsprozesse und Determinanten. In: Leppin A, Hurrelmann K, Petermann H, editors. Jugendliche und Alltagsdrogen: Konsum und Perspektiven der Prävention. Neuwied, Germany: Luchterhand; 2000:64-94.
5. Weichold K, Bühler A, Silbereisen R. Konsum von Alkohol und illegalen Drogen im Jugendalter. In: Silbereisen R, Hasselhorn M, editors. Enzyklopädie der Psychologie, Bd 5 Entwicklungspsychologie des Jugendalters. Göttingen, Germany: Hogrefe; 2008:537-586.
6. Hickie I. Alcohol and the teenage brain: safest to keep them apart. Camperdown, Australia: Brain & Mind Research Institute at the University of Sydney; 2009. URL: <http://drinkwisewebsite.s3-ap-southeast-1.amazonaws.com/2011/10/Alcohol-and-the-Teenage-Brain-Safest-to-keep-them-apart.pdf> [accessed 2012-11-21] [WebCite Cache ID 6CLLZqR9A]
7. Lubman DI, Yücel M, Hall WD. Substance use and the adolescent brain: a toxic combination? *J Psychopharmacol* 2007 Nov;21(8):792-794. [doi: [10.1177/0269881107078309](#)] [Medline: [17984159](#)]
8. Pabst A, Piontek D, Kraus L, Müller S. Substanzkonsum und substanzbezogene Störungen. Ergebnisse des Epidemiologischen Suchtsurveys 2009. *Sucht* 2010;56(5):327-336. [doi: [10.1024/0939-5911/a000044](#)]
9. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. *Bull World Health Organ* 2004 Nov;82(11):858-866 [FREE Full text] [Medline: [15640922](#)]
10. Rooke S, Thorsteinsson E, Karpin A, Copeland J, Allsop D. Computer-delivered interventions for alcohol and tobacco use: a meta-analysis. *Addiction* 2010 Aug;105(8):1381-1390. [doi: [10.1111/j.1360-0443.2010.02975.x](#)] [Medline: [20528806](#)]
11. Riper H, van Straten A, Keuken M, Smit F, Schippers G, Cuijpers P. Curbing problem drinking with personalized-feedback interventions: a meta-analysis. *Am J Prev Med* 2009 Mar;36(3):247-255. [doi: [10.1016/j.amepre.2008.10.016](#)] [Medline: [19215850](#)]
12. Riper H, Kramer J, Smit F, Conijn B, Schippers G, Cuijpers P. Web-based self-help for problem drinkers: a pragmatic randomized trial. *Addiction* 2008 Feb;103(2):218-227. [doi: [10.1111/j.1360-0443.2007.02063.x](#)] [Medline: [18199300](#)]

13. 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](https://doi.org/10.2196/jmir.2090)] [Medline: [22954459](https://pubmed.ncbi.nlm.nih.gov/22954459/)]
14. Blankers M, Koeter MW, Schippers GM. Internet therapy versus internet self-help versus no treatment for problematic alcohol use: A randomized controlled trial. *J Consult Clin Psychol* 2011 Jun;79(3):330-341. [doi: [10.1037/a0023498](https://doi.org/10.1037/a0023498)] [Medline: [21534652](https://pubmed.ncbi.nlm.nih.gov/21534652/)]
15. Boon B, Risselada A, Huiberts A, Riper H, Smit F. Curbing alcohol use in male adults through computer generated personalized advice: randomized controlled trial. *J Med Internet Res* 2011;13(2):e43 [FREE Full text] [doi: [10.2196/jmir.1695](https://doi.org/10.2196/jmir.1695)] [Medline: [21719412](https://pubmed.ncbi.nlm.nih.gov/21719412/)]
16. Hustad JT, Barnett NP, Borsari B, Jackson KM. Web-based alcohol prevention for incoming college students: a randomized controlled trial. *Addict Behav* 2010 Mar;35(3):183-189 [FREE Full text] [doi: [10.1016/j.addbeh.2009.10.012](https://doi.org/10.1016/j.addbeh.2009.10.012)] [Medline: [19900763](https://pubmed.ncbi.nlm.nih.gov/19900763/)]
17. Kypri K, Langley JD, Saunders JB, Cashell-Smith ML, Herbison P. Randomized controlled trial of Web-based alcohol screening and brief intervention in primary care. *Arch Intern Med* 2008;168(5):530-536.
18. Neighbors C, Lee CM, Lewis MA, Fossos N, Walter T. Internet-based personalized feedback to reduce 21st-birthday drinking: a randomized controlled trial of an event-specific prevention intervention. *J Consult Clin Psychol* 2009 Feb;77(1):51-63 [FREE Full text] [doi: [10.1037/a0014386](https://doi.org/10.1037/a0014386)] [Medline: [19170453](https://pubmed.ncbi.nlm.nih.gov/19170453/)]
19. Spijkerman R, Roek MA, Vermulst A, Lemmers L, Huiberts A, Engels RC. Effectiveness of a web-based brief alcohol intervention and added value of normative feedback in reducing underage drinking: a randomized controlled trial. *J Med Internet Res* 2010;12(5):e65 [FREE Full text] [doi: [10.2196/jmir.1465](https://doi.org/10.2196/jmir.1465)] [Medline: [21169172](https://pubmed.ncbi.nlm.nih.gov/21169172/)]
20. Dumas DM, Hannah E. Preventing high-risk drinking in youth in the workplace: a web-based normative feedback program. *J Subst Abuse Treat* 2008 Apr;34(3):263-271. [doi: [10.1016/j.jsat.2007.04.006](https://doi.org/10.1016/j.jsat.2007.04.006)] [Medline: [17600650](https://pubmed.ncbi.nlm.nih.gov/17600650/)]
21. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of Internet-based psychotherapeutic interventions. *Journal of Technology in Human Services* 2008 Jul 2008;26(2-4):109-160. [doi: [10.1080/15228830802094429](https://doi.org/10.1080/15228830802094429)]
22. 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 Nov;14(6):e152 [FREE Full text] [doi: [10.2196/jmir.2104](https://doi.org/10.2196/jmir.2104)] [Medline: [23151820](https://pubmed.ncbi.nlm.nih.gov/23151820/)]
23. Federal Centre for Health Education. 2012 Dec 13. Drugcom: Change Your Drinking URL: <http://www.drugcom.de/beratung-finden/change-your-drinking-beratung-und-hilfe-zur-reduzierung-des-alkoholkonsums/> [accessed 2012-12-13] [WebCite Cache ID 6CstlbzBI]
24. Federal Centre for Health Education. 2013 May 06. Alcohol?: Know Your Limit [German] URL: <http://www.kenn-dein-limit.info/home.html> [accessed 2013-05-06] [WebCite Cache ID 6GPsBwUgi]
25. Know Your Limit [German]. 2013 Feb 27. What is a standard glass? [German] URL: <http://www.kenn-dein-limit.de/alkohol-beratung/haeufige-fragen/fragen-zu-alkohol/was-ist-ein-standardglas/> [accessed 2013-02-27] [WebCite Cache ID 6EkQ55Ujh]
26. Burger M, Brönstrup A, Pietrzik K. Derivation of tolerable upper alcohol intake levels in Germany: a systematic review of risks and benefits of moderate alcohol consumption. *Prev Med* 2004 Jul;39(1):111-127. [doi: [10.1016/j.ypmed.2003.11.011](https://doi.org/10.1016/j.ypmed.2003.11.011)] [Medline: [15207992](https://pubmed.ncbi.nlm.nih.gov/15207992/)]
27. Kurzintervention bei Patienten mit Alkoholproblemen. Cologne, German: Bundeszentrale für gesundheitliche Aufklärung URL: <http://www.bzga.de/pdf.php?id=15b8ff4e612c49ccb5a4cd354e5d6f99> [accessed 2013-02-27] [WebCite Cache ID 6EkQXFT3I]
28. Mcgee R, Kypri K. Alcohol-related problems experienced by university students in New Zealand. *Aust N Z J Public Health* 2004 Aug;28(4):321-323. [Medline: [15704694](https://pubmed.ncbi.nlm.nih.gov/15704694/)]
29. Babor TF, Higgins-Biddle JC, Saunders JB, Monteiro MG. AUDIT: The Alcohol Use Disorders Identification Test. Geneva: World Health Organization; 2001. URL: http://whqlibdoc.who.int/hq/2001/who_msd_msb_01.6a.pdf [accessed 2012-11-21] [WebCite Cache ID 6CLLziIJ]
30. Berg IK, Miller SD. Kurzzeittherapie bei Alkoholproblemen: Ein lösungsorientierter Ansatz. Heidelberg: Carl-Auer-Verlag; 2009.
31. Kanfer FH. Self-regulation: research issues/speculations. In: *Behavior Modification in Clinical Psychology*. New York: Appleton-Century-Crofts; 1970:178-220.
32. Kanfer FH. Implications of a self-regulation model of therapy for treatment of addictive behaviors. In: *Treating Addictive Behaviors: Processes of Change*. New York: Plenum Press; 1986:29-47.
33. Miller WR, Rollnick S. *Motivational Interviewing, Third Edition: Helping People Change (Applications of Motivational Interviewing)*. New York: The Guilford Press; 2012.
34. Marlatt GA, Gordon JR. *Relapse Prevention: Maintenance Strategies in the Treatment of Addictive Behaviors*. New York: Guilford Press; 1985.
35. Parks GA, Anderson BK, Marlatt GA. Relapse prevention therapy. In: *The Essential Handbook of Treatment and Prevention of Alcohol Problems*. New York: Wiley; 2004:87-104.

36. 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](https://doi.org/10.2196/jmir.1923)] [Medline: [22209829](https://pubmed.ncbi.nlm.nih.gov/22209829/)]
37. Wallace P, Murray E, McCambridge J, Khadjesari Z, White IR, Thompson SG, et al. On-line randomized controlled trial of an internet based psychologically enhanced intervention for people with hazardous alcohol consumption. *PLoS One* 2011;6(3):e14740 [FREE Full text] [doi: [10.1371/journal.pone.0014740](https://doi.org/10.1371/journal.pone.0014740)] [Medline: [21408060](https://pubmed.ncbi.nlm.nih.gov/21408060/)]
38. Postel MG, de Haan HA, ter Huurne ED, Becker ES, de Jong CA. Effectiveness of a web-based intervention for problem drinkers and reasons for dropout: randomized controlled trial. *J Med Internet Res* 2010;12(4):e68 [FREE Full text] [doi: [10.2196/jmir.1642](https://doi.org/10.2196/jmir.1642)] [Medline: [21163776](https://pubmed.ncbi.nlm.nih.gov/21163776/)]
39. Del Boca FK, Darkes J. The validity of self-reports of alcohol consumption: state of the science and challenges for research. *Addiction* 2003 Dec;98 Suppl 2:1-12. [Medline: [14984237](https://pubmed.ncbi.nlm.nih.gov/14984237/)]
40. Babor TF, Steinberg K, Anton R, Del Boca F. Talk is cheap: measuring drinking outcomes in clinical trials. *J Stud Alcohol* 2000 Jan;61(1):55-63. [Medline: [10627097](https://pubmed.ncbi.nlm.nih.gov/10627097/)]
41. Cunningham JA, Wild TC, Cordingley J, Van Mierlo T, Humphreys K. A randomized controlled trial of an internet-based intervention for alcohol abusers. *Addiction* 2009;104(12):2023-2032. [doi: [10.1111/j.1360-0443.2009.02726.x](https://doi.org/10.1111/j.1360-0443.2009.02726.x)]
42. Jonas B, Tossmann P, Tensil M, Leuschner F, Strüder E. Effekte einer einmaligen Chat-Intervention auf problematischen Substanzkonsum. *SUCHT - Zeitschrift für Wissenschaft und Praxis / Journal of Addiction Research and Practice* 2012;58(3):173-182. [doi: [10.1024/0939-5911.a000182](https://doi.org/10.1024/0939-5911.a000182)]

Abbreviations

APS: Alcohol Problems Scale

AUDIT: Alcohol Use Disorders Identification Test

BZgA: Bundeszentrale für gesundheitliche Aufklärung (German Federal Centre for Health Education)

GEE: generalized estimating equations

ITT: intention-to-treat

LOCF: last observation carried forward

RCT: randomized controlled trial

SG: standard glasses

Edited by G Eysenbach; submitted 14.12.12; peer-reviewed by M Blankers, S Rooke; comments to author 25.01.13; revised version received 01.03.13; accepted 19.04.13; published 06.06.13.

Please cite as:

Tensil MD, Jonas B, Strüder E

Two Fully Automated Web-Based Interventions for Risky Alcohol Use: Randomized Controlled Trial

J Med Internet Res 2013;15(6):e110

URL: <http://www.jmir.org/2013/6/e110/>

doi: [10.2196/jmir.2489](https://doi.org/10.2196/jmir.2489)

PMID: [23742808](https://pubmed.ncbi.nlm.nih.gov/23742808/)

©Marc-Dennan Tensil, Benjamin Jonas, Evelin Strüder. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 06.06.2013. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Cardiovascular Health Effects of Internet-Based Encouragements to Do Daily Workplace Stair-Walks: Randomized Controlled Trial

Lars Louis Andersen¹, PhD; Emil Sundstrup¹, MSc; Marianne Boysen¹, MSc; Markus Due Jakobsen¹, MSc; Ole Steen Mortensen², MD, PhD; Roger Persson³, PhD

¹National Research Centre for the Working Environment, Copenhagen, Denmark

²University Hospital Køge, Department of Occupational Medicine, Køge, Denmark

³Lund University, Department of Psychology, Lund, Sweden

Corresponding Author:

Lars Louis Andersen, PhD

National Research Centre for the Working Environment

Lersø Parkalle 105

Copenhagen, 2100

Denmark

Phone: 45 39165300

Fax: 45 39165301

Email: lla@nrcwe.dk

Abstract

Background: Although the hazardous health effects of a sedentary lifestyle are well known, many adults struggle with regular physical activity. Simple and efficient encouragements for increased physical activity are needed.

Objective: To determine the effect on cardiovascular health of email-based encouragements to do daily stair-walks at work together with colleagues among adults in sedentary occupations.

Methods: A single-blind randomized controlled trial was performed at a large administrative company in Copenhagen, Denmark. Participants were 160 office workers (125 women, 35 men; mean age 42 years, SD 10; sitting 89.5% of work time). At baseline, aerobic fitness was 37 mL/min/kg (SD 9), mean blood pressure was 118/79 mmHg (SD 14/9), and mean body mass index (BMI) was 23 kg/m² (SD 4). Participants were randomly assigned (2:1 ratio) to an email group receiving weekly email-based encouragements to walk the stairs for 10 minutes a day or to a control group receiving weekly reminders to continue their usual physical activities. The primary outcome was the change from baseline to 10-week follow-up in aerobic fitness determined from a maximal cycle test. The examiner was blinded to group allocation.

Results: Adherence to the email encouragements was fairly high with 82.7% of the participants performing at least 3 sessions of 10-minute stair-walks per week (mean 3.3, SD 1.3). Mean heart rate reached 167 beats/min (SD 10) during stair-walks. In the intention-to-treat analysis, aerobic fitness increased 1.45 mL/min/kg (95% CI 0.64-2.27) at 10-week follow-up in the email group compared with the control group. In participants with low aerobic fitness at baseline (n=56), aerobic fitness increased 1.89 mL/min/kg (95% CI 0.53-3.24), and systolic and diastolic blood pressure decreased 4.81 mmHg (95% CI 0.47-9.16) and 2.67 mmHg (95% CI 0.01-5.32), respectively, in the email group compared with the control group. Body weight decreased in the email group of those with low aerobic fitness compared with the control group, but this was not statistically significant.

Conclusions: Simple and inexpensive email-based encouragements to do daily stair-walks together with colleagues at work improves cardiovascular health among adults in sedentary occupations. There exists an enormous potential to prevent the hazardous health effects of a sedentary lifestyle through the use of email-based encouragements to do short bouts of physical activity at the workplace.

Trial Registration: Clinicaltrials.gov NCT01293253; <http://clinicaltrials.gov/ct2/show/NCT01293253> (Archived by WebCite at <http://www.webcitation.org/6HWG2jw68>).

(*J Med Internet Res* 2013;15(6):e127) doi:[10.2196/jmir.2340](https://doi.org/10.2196/jmir.2340)

KEYWORDS

physical activity; Internet technology; sedentary; fitness; vigorous activity; intervention; randomized controlled trial; stair-walk; blood pressure; body mass index

Introduction

According to the World Health Organization, lack of physical activity is ranked as the fourth leading cause of all deaths worldwide [1]. To promote and maintain health, the American College of Sports Medicine (ACSM) and the American Heart Association (AHA) recommend that healthy working-age adults perform moderate aerobic physical activity for at least 30 minutes on 5 days each week or vigorous aerobic physical activity for at least 20 minutes on 3 days each week [2]. In spite of this knowledge, surveys from North America and Britain show that only one-quarter to one-half of adults self-report that they meet public recommendations for physical activity [3,4]. Thus, lack of sufficient physical activity in the population remains a problem.

The workplace may provide an optimal setting to encourage a healthier lifestyle because most adults spend the majority of their waking hours at work together with colleagues, many with similar needs for physical activity [5]. Furthermore, in modern society most adult workers have the opportunity to communicate and organize physical activities through emails or Internet-based media. In addition, overcoming the motivational barriers for performing regular physical activity is easier in a social setting, such as the workplace, than individually [6]. Internet technology helps create social networks; thus, social media may be used as a community setting at workplaces to encourage a healthier lifestyle and motivate physical activities. However, the need for special facilities and subsequent showering limits the feasibility of prolonged and strenuous physical exercise programs at the workplace. Further, lack of time is often cited as the major reason among adults for not being physically active [7]. Thus, short bouts of vigorous physical activity at the workplace that can easily be organized by using Internet technology and without the need for special facilities and subsequent showering may be preferred [8]. Robroek and coworkers [9] showed that employees receiving monthly email prompts were 6 times more likely than those not receiving email prompts to continue using an Internet-based physical activity and healthy nutrition program at the workplace. Thus, emails and Internet-based media may be an efficient and inexpensive method to organize and motivate participation in such physical activities.

Stair-walking is a vigorous form of physical activity requiring 5 to 10 times the energy expenditure of rest [10]. As most office buildings have accessible stairways, stair-walking may be a feasible form of physical activity at the workplace that can be encouraged and organized using Internet technology. Most stair-walk interventions at the workplace have reported positive effects on aerobic fitness, blood pressure, and/or blood lipids [11-15], which is important from a preventive perspective because a high level of aerobic fitness is associated with decreased risk for cardiovascular and all-cause mortality [16,17]. However, methodological limitations were present in many of the previous stair-walk studies, for example, no control group

or no blinding of examiners. Further, none of the aforementioned studies used Internet technology as part of the intervention. Thus, high-quality randomized controlled trials investigating the health effects of Internet-based encouragements to do daily physical activity at the workplace are needed. Companies play an important role in health promotion, but randomized studies have failed to show positive return of investment of health-promoting activities at the workplace probably because the cost of interventions are often high [18]. Consequently, as many companies have limited resources for health-promoting activities, the implementation must be simple and inexpensive. Email-based encouragements are cheap, easy to administer, and have the ability to reach virtually all employees at office workplaces; thus, they may result in mutual benefit for employees and employers. One option could be to send out emails to encourage using the stairs as part of daily routine, for example, when having meetings on another floor, and using the stairs instead of elevators [14]. Encouragements could also be delivered through emails to form groups with colleagues and do planned stair-walking as an active break to supplement overall physical activity. The present study uses this approach.

The aim of the present study is to determine the effect of email-based encouragement to do daily stair-walks together with colleagues on cardiovascular health among adults in sedentary occupations. It was hypothesized that email-based encouragements to do daily stair-walking improves aerobic fitness (primary outcome).

Methods**Ethical Approval and Trial Registration**

The Local Ethical Committee approved the study protocol (H-3-2010-062). The trial was registered prior to enrollment of participants (Clinicaltrials.gov NCT01293253), which ensured that the study aim, hypothesis, and primary outcome were predefined. There were no changes in outcomes after the trial commenced. All participants gave written informed consent in agreement with the Declaration of Helsinki.

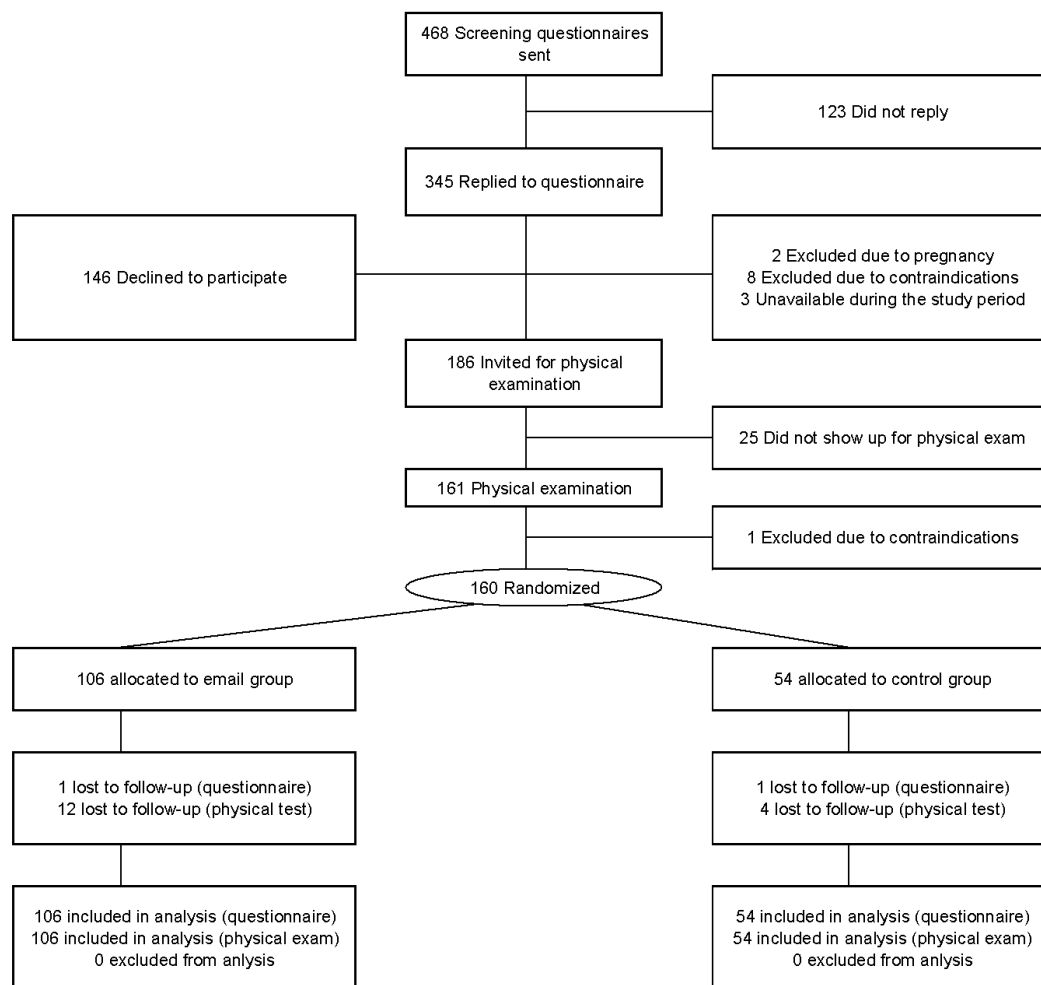
Study Design and Flow of Participants

A randomized controlled trial was performed in Copenhagen, Denmark, from February to June 2011. The CONSORT-EHEALTH checklist was followed to ensure transparent and standardized reporting of the trial [19]. Figure 1 shows the flow of participants through the trial. A screening questionnaire on health, physical activity, and working conditions went out by email to 468 office workers in a large administrative company (ie, a single workplace), and 345 replied. All participants had computer/Internet literacy. The inclusion criterion were willingness to participate in the study (N=199). Exclusion criteria were (1) a medical history of life-threatening disease (n=8), (2) current pregnancy (n=2), (3) unavailable during the study period (n=3), and (4) blood pressure greater than 160/100 mmHg (determined later during physical

examination). An invitation for a physical examination was sent to the remaining 186 employees, and 161 showed up. During the physical examination, 1 employee was excluded because of blood pressure greater than 160/100 mmHg. Thus, the final

sample size was 160 participants. A blinded examiner measured blood pressure and aerobic fitness of the employees at baseline and at 10-week follow-up as described subsequently.

Figure 1. CONSORT flow diagram of participants through the study.



Randomization and Blinding

Using a computer-generated random numbers table, the 160 participants were randomly allocated to the email group or control group (2:1 ratio) following the baseline examination. Simple randomization was used. The 2:1 ratio was chosen to ensure an adequate number of participants in the email intervention to form stair-walking groups together with colleagues. The examiner remained blinded to group allocation, and participants were instructed via email not to reveal their group allocation during follow-up examination. Author LLA performed the randomization and informed participants via email about group allocation.

Primary Outcome: Aerobic Fitness

Aerobic fitness was estimated from a maximal cycle ergometer test, which has previously shown very good validity with direct measurements of oxygen uptake [20]. Men and women started the test with workloads of 105 and 70 watts, respectively, on a Monark cycle ergometer (model 874E, Monark AB, Stockholm, Sweden) and maintained a pedaling frequency of 70 revolutions

per minute. The examiner added 0.5 kg resistance (35 watts) every other minute until the participant was not able to maintain a pedaling frequency of 70 revolutions per minute. The final workload (watts) and time using the final workload (seconds) determined maximal power output (MPO), calculated as $MPO \text{ (watts)} = \text{final workload} - 35 + (35 \times \text{number of seconds at final workload} / 120)$. From this, the maximal oxygen uptake (VO_{2max}) was estimated using the equation $VO_{2max} \text{ (L/min)} = 0.16 + (0.0117 \times MPO)$ [20], and then divided by body weight (kg) and multiplied by 1000 to determine aerobic fitness ($mL/min/kg$).

Subgroup With Low Aerobic Fitness

For exploratory subgroup analysis of employees with low aerobic fitness at baseline, participants were categorized according to the methods of Åstrand [21], defining women aged <30, 30-39, 40-49, 50-59, and >60 years with aerobic fitness of <35, <34, <32, <29, and <29 $mL/min/kg$, respectively, and men aged <30, 30-39, 40-49, 50-59, and >60 years with aerobic fitness of <44, <40, <36, <32, and <27 $mL/min/kg$, respectively, as having low aerobic fitness.

Secondary Outcomes

Blood pressure (systolic/diastolic) was measured during the physical examination after 15 minutes rest. The examiner performed 3 measurements, and the average of these measurements was calculated. Participants were weighed on a Tanita scale (model Tanita Innerscan 543) providing information on body weight and body fat percentage. Height was measured with a Soehnle Foldable Ultrasound Height Rod. Body mass index (BMI) was calculated as $BMI (kg/m^2) = \text{body weight}/(\text{height})^2$.

Interventions

Before the study, a meeting was held with the management of the company who agreed to let employees in the email group walk the stairs for 10 minutes a day during work time. The management of the company felt that 10 minutes of group-based physical activities per day would not impair productivity and could perhaps contribute to improved social climate. Information about the aim and content of the project was posted on the intranet of the company. Further, the employees were invited via email to participate in an informational meeting about the project.

After baseline testing and randomization, participants in the email group were advised and reminded via email each Monday at 9 am over 10 weeks (ie, a total of 10 emails) to walk the stairs for 10 minutes a day during working days (typically 5 days a week). We chose this type of encouragement because emails are cheap, easy to administer, and have the ability to reach virtually all employees at office workplaces. Thus, even companies with limited resources would be able to implement such an intervention as opposed to physical activities requiring gyms, training equipment, and instructors. Furthermore, emails are the preferred source of information among employees [22]. However, no previous studies have documented the optimal frequency of emails to encourage physical activity. Whereas email overload may be a problem for some people [23], emails can be dealt with at one's own convenience and, therefore, have limited impact on the normal workflow [24]. The frequency of once per week was chosen to not unnecessarily disturb the employees with too many emails, yet still frequently remind the employees to walk the stairs. This frequency of emails was based on experience from a previous workplace exercise study in office workers [25].

Participants in the email group were encouraged via email to form small groups with their colleagues in the same group and, if possible, to schedule the daily stair-walks. In detail, participants were advised to (1) walk at the same time each day to make it a habit, for example, before lunch or when meeting at work in the morning, (2) make an appointment in their work calendar (Outlook) with colleagues to use and improve social networks at the workplace, and (3) vary the stair-walks sometimes by changing the speed between flights of stairs and alternating between single and double steps to make it more challenging and avoid boredom. The content of the emails did not build on any specific behavioral theories, but varied slightly from week to week with the following 3 progression phases to ensure organization, intensity, and variation: (1) during the first

weeks, the goal was to organize the stair-walking by scheduling and forming groups, (2) during the middle weeks, the goal was to increase exercise intensity by increasing the speed of stair-walking to ensure effectiveness, and (3) during the later weeks, the goal was also to introduce variation in the stair-walking (eg, varying between single and double steps) to avoid boredom.

The office building had 3 staircases with 3 to 4 floors. Stair-climbing is generally considered a vigorous type of physical activity [10]. In the present study, participants were offered to participate in heart rate measurements during stair-walking on the Wednesday of week 5 between 10 am and 12 am, during which time a research assistant presented with heart rate monitors at the main stair of the company. These measurements confirmed that the intensity of activity was high, with mean heart rate reaching 167 beats/min (SD 10) after 10 minutes of stair-walking ($n=23$), corresponding to 90% (SD 9%) of the heart rate reserve, calculated as $(\text{heart rate during stair-walk} - \text{resting heart rate})/(\text{max heart rate} - \text{resting heart rate}) \times 100\%$. Figure 2 shows the recorded heart rate from 1 of the participants of the email group during 1 of the daily stair-walks. Participants in the control group were reminded each Monday at 9 am over 10 weeks via email to continue their usual physical activities during the project.

Feasibility

After the first week of the intervention, the participants replied to an Internet-based questionnaire about sweating during stair-walking: "Did you experience sweating during the 10-minute stair-walks?" Participants had 3 response options: (1) no, (2) yes, but not to an extent that it bothered me, and (3) yes, to an extent that it bothered me. Participants also replied to an Internet-based questionnaire at 10-week follow-up. Adherence was evaluated by the question: "How many days per week during the last 10 weeks have you walked the stairs for at least 10 minutes at a time" (0, 1, 2, ..., 7 days per week). Asking specific questions about adherence to workplace physical exercise in retrospect has shown good validity to day-to-day training diary registrations [26]. Participants also replied to the question: "During the last week, did you walk the stairs (1) together with colleagues, (2) alone, (3) alone and together with colleagues to an equal extent." Multiple-choice questions concerning reasons for not participating as often as required (due to lack of time, lack of interest, illness, etc) were also given. Further, to determine the potential for long-term implementation of daily stair-walk, participants were asked whether they wished to continue daily stair-walking after termination of the research project, with the response options (1) yes, (2) yes, maybe, and (3) no.

Other Physical Activities

We included questions to assess for initiation of other physical activities during the study period that could affect the primary outcome. At baseline and follow-up, participants filled in a modified version of the Saltin and Grimby [27] questionnaire concerning low-, medium-, and high-intensity leisure time physical activity. On an exploratory basis, we included this question as a covariate in the analyses on changes in fitness and blood pressure. Further, participants from both the control and

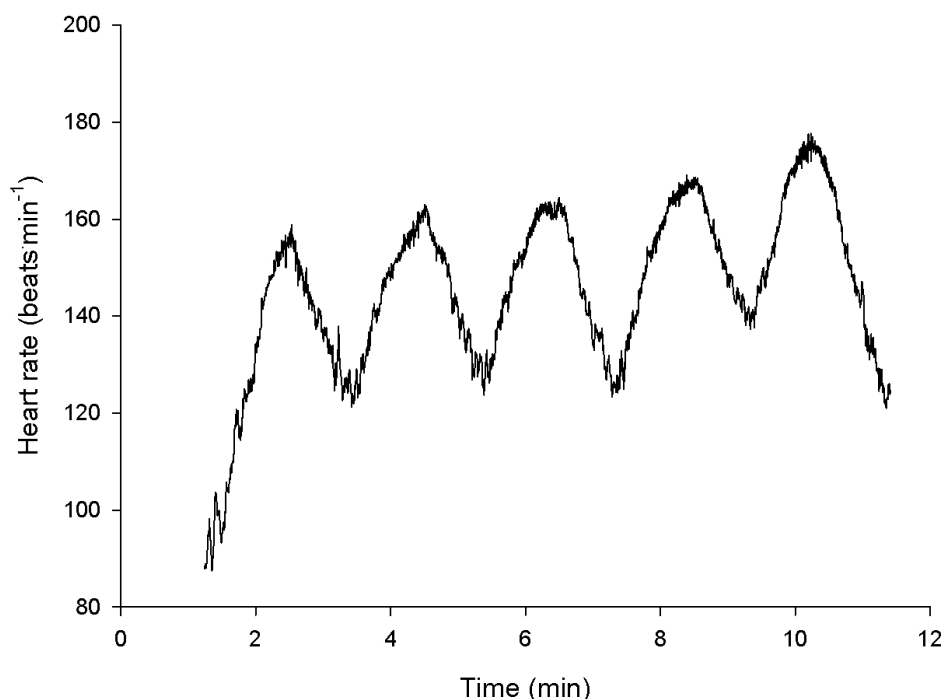
email groups replied to the follow-up question regarding weekly frequency of 10-minute stair-walks.

Statistics

The main outcomes were analyzed according to the intention-to-treat principle using a 2×2 mixed-factorial design, with time and group as independent categorical variables (fixed factors). The last-observation-carried-forward was not used because all methods of imputation have limitations. Instead,

efforts were made at follow-up to explain to all participants, including dropouts, that their data were still required regardless of their level of actual participation, and we used the PROC GLIMMIX (general linear mixed models) of SAS version 9.2. (SAS Institute, Inc, Cary, NC, USA), which inherently accounts for missing values. Baseline values are reported as means (SD) and differences from baseline to follow-up as means (95% confidence interval).

Figure 2. Heart rate recording from one of the participants during one of the daily stair-walks. This participant walked five times from the ground floor to the fourth floor of the building in 10 minutes.



Results

Adherence and Feasibility

In the email group, 82.7% (86/104) of the participants performed at least 3 stair-walking sessions per week. On average, the number of weekly stair-walking sessions was 3.3 (SD 1.3). During the last week of the intervention, participants from the email group replied that they had walked together with colleagues (65.3%, 68/104), alone (31.7%, 33/104), and alone and together with colleagues to an equal extent (2.9%, 3/104), respectively. Regarding the question after the first week of the intervention whether experiencing sweating during the 10-minute stair-walks, in the email group 5.2% (5/96) replied no, 71.8% (69/96) replied yes, but not to an extent that it bothered me, and 22.9% (22/96) replied yes, to an extent that it bothered me. In the email group, reasons for not participating as often as required were lack of time (51.0%, 53/104), illness (18.3%, 19/104), lack of interest (11.5%, 12/104), lack of motivation to start the program after vacation (7.7%, 8/104), difficulties in starting the program after illness (2.9%, 3/104), lack of benefit from the stair-walk program (1.9%, 2/104), and lack of acceptance from nearest leader (1.0%, 1/104). None of the participants cited lack of acceptance from colleagues as a

reason. In the email group, 26.2% (27/103) wished to continue stair-walking after the research project had ended, 53.4% (55/103) stated that maybe they wished to continue, and 20.4% (21/103) did not wish to continue stair-walking.

Attrition

In the email group, 11 participants (10.4%) dropped out during the intervention period because of pregnancy (n=1), leaving the company (n=2), long-term vacation (n=1), stress (n=1), lack of time (n=2), and knee or ankle/foot injuries (n=4) of which 2 reported stair-walking as the cause of injury. In the control group, 5 participants (9.2%) dropped out during the intervention period because of long-term sickness absence (n=1) and with no reason given (n=4). Three and 2 of the dropouts from the email group and control group, respectively, presented for the follow-up examination, and 10 and 5, replied to the follow-up questionnaire, respectively. All dropouts were included in the intention-to-treat analyses.

Intention-to-Treat Analysis

Table 1 shows baseline demographics of all participants of the 2 groups. Table 2 shows results from the intention-to-treat analysis for the change in fitness, blood pressure, and other variables from baseline to follow-up. A priori hypothesis testing

showed a significant group \times time interaction for aerobic fitness ($P<.001$). The email group improved aerobic fitness significantly more than the control group with a between-group difference of 1.45 mL/min/kg (95% CI 0.64-2.27). For the remainder of variables, there were no significant group \times time interactions. We also performed exploratory analyses adjusting for the baseline level of physical activity on the changes in fitness and blood pressure. This had only minor influence on the results and the between-group difference for the change in fitness was 1.34 mL/min/kg (95% CI 0.51-2.16), and still nonsignificant for the changes in blood pressure.

Subgroup With Low Aerobic Fitness

Table 3 shows baseline demographics of participants with low aerobic fitness of the 2 groups. Table 4 shows results from the subgroup analysis among individuals with low aerobic fitness at baseline for the change in fitness, blood pressure, and other variables from baseline to follow-up. There was also a significant group \times time interaction for aerobic fitness ($P=.008$),

and systolic ($P=.03$) and diastolic blood pressure ($P=.04$). There was also a tendency for the email group to decrease body weight ($P=.08$) compared with the control group, although this did not reach statistical significance.

Other Physical Activities

The weekly duration of leisure time physical activity did not change significantly in any of the groups during the intervention period. Based on the follow-up questionnaire replies, 4 participants from the control group had performed 1 ($n=2$) and 2 ($n=2$) weekly sessions of 10-minute stair-walks.

Test-Retest Reliability

Using the baseline and 10-week follow-up data from the control group ($n=50$), test-retest reliability was determined as the intraclass correlation coefficient (ICC) for the different outcomes, with ICC 0.98 for aerobic fitness, ICC 0.90 for diastolic and systolic blood pressure, and ICC 0.99 for weight, BMI, and body fat percentage.

Table 1. Baseline characteristics of participants in the email group and control group ($N=160$).

Variables	Email group ($n=106$)	Control group ($n=54$)
Demographics		
Age (years), mean (SD)	42 (10)	43 (11)
Height (cm), mean (SD)	171 (9)	172 (10)
Weight (kg), mean (SD)	69 (14)	71 (11)
Body mass index (kg/m^2), mean (SD)	23 (4)	23 (3)
Body fat %, mean (SD)	28 (8)	29 (9)
Gender (women), n (%)	84 (79.2%)	41 (75.9%)
Cardiovascular, mean (SD)		
Aerobic fitness (mL/min/kg)	36 (8)	38 (10)
Systolic blood pressure (mmHg)	118 (14)	118 (14)
Diastolic blood pressure (mmHg)	79 (9)	78 (8)
Work-related, mean (SD)		
Sitting (% of work time)	89 (18)	90 (19)
Weekly working hours	38 (3)	39 (4)
Leisure time physical activity (hours/week), mean (SD)		
Low intensity	3.2 (1.7)	3.3 (1.7)
Moderate intensity	1.8 (1.5)	1.9 (1.6)
Vigorous intensity	0.3 (0.8)	0.6 (1.3)
Other, n (%)		
Smokers	8 (7.5%)	6 (11.1%)

Table 2. Changes from baseline to 10-week follow-up (N=160).

Variables	Within-group differences from baseline to 10-week follow-up				Between-group differences from baseline to 10-week follow-up		
	Email group (n=106)		Control group (n=54)		Email vs Control		<i>P</i> value
	Mean	95% CI	Mean	95% CI	Mean	95% CI	
Cardiovascular							
Aerobic fitness (mL/min/kg)	2.32	1.84, 2.79	0.86	0.20, 1.52	1.45	0.64, 2.27	<.001
Systolic blood pressure (mmHg)	−2.62	−4.32, −0.93	−1.04	−3.36, 1.28	−1.58	−4.45, 1.28	.28
Diastolic blood pressure (mmHg)	−2.85	−3.90, −1.81	−2.17	−3.61, −0.74	−0.68	−2.45, 1.10	.45
Leisure time physical activity (hours/week)							
Low intensity	−0.15	−0.48, 0.18	−0.15	−0.61, 0.31	0.01	−0.57, 0.56	.98
Moderate intensity	0.00	−0.30, 0.30	0.03	−0.38, 0.45	−0.04	−0.55, 0.47	.89
Vigorous intensity	0.10	−0.05, 0.25	−0.04	−0.25, 0.17	0.13	−0.12, 0.39	.31
Other							
Weight (kg)	−0.24	−0.60, 0.13	−0.12	−0.62, 0.38	−0.11	−0.73, 0.50	.72
Body fat percentage (%)	−0.44	−0.71, −0.17	−0.47	−0.85, −0.10	0.03	−0.43, 0.49	.89

Table 3. Baseline characteristics of participants with low aerobic fitness in the email group and control group (n=56).

Variables	Email group (n=38)	Control group (n=18)
Demographics		
Age (years), mean (SD)	44 (11)	46 (12)
Height (cm), mean (SD)	170 (8)	166 (7)
Weight (kg), mean (SD)	76 (14)	71 (10)
Body mass index (kg/m ²), mean (SD)	26 (5)	25 (4)
Body fat %, mean (SD)	33 (7)	33 (8)
Gender (women), n (%)	32 (84.2%)	15 (83.3%)
Cardiovascular, mean (SD)		
Aerobic fitness(mL/min/kg)	28 (5)	28 (5)
Systolic blood pressure (mmHg)	119 (17)	119 (14)
Diastolic blood pressure (mmHg)	80 (10)	79 (8)
Work-related, mean (SD)		
Sitting (% of work time)	85 (22)	86 (18)
Weekly working hours	38 (3)	38 (3)
Leisure time physical activity (hours/week), mean (SD)		
Low intensity	3.0 (2.0)	3.0 (2.0)
Moderate intensity	1.0 (1.0)	1.0 (1.0)
Vigorous intensity	0.0 (1.0)	0.0 (0.0)
Other, n (%)		
Smokers	5 (13.2%)	4 (22.2%)

Table 4. Changes from baseline to 10-week follow-up in participants with low aerobic fitness at baseline (n=56).

Variables	Within-group differences from baseline to 10-week follow-up				Between-group differences from baseline to 10-week follow-up		
	Email group (n=38)		Control group (n=18)		Email vs Control		<i>P</i> value
	Mean	95% CI	Mean	95% CI	Mean	95% CI	
Cardiovascular							
Aerobic fitness (mL/min/kg)	2.80	2.04, 3.62	0.95	−0.15, 2.05	1.89	0.53, 3.24	.008
Systolic blood pressure (mmHg)	−3.19	−5.70, −0.68	1.63	−1.92, 5.17	−4.81	−9.16, −0.47	.03
Diastolic blood pressure (mmHg)	−3.46	−4.99, −1.93	−0.79	−2.96, 1.37	−2.67	−5.32, −0.01	.04
Leisure time physical activity (hours/week)							
Low intensity	0.11	−0.51, 0.72	0.00	−0.88, 0.88	0.11	−0.96, 1.18	.84
Moderate intensity	0.16	−0.25, 0.58	−0.06	−0.65, 0.54	0.21	−0.51, 0.94	.55
Vigorous intensity	0.00	−0.22, 0.22	−0.06	−0.37, 0.26	0.06	−0.32, 0.43	.77
Other							
Weight (kg)	−0.81	−1.56, −0.07	0.33	−0.72, 1.38	−1.1	−2.43, 0.15	.08
Body fat percentage (%)	−0.68	−1.16, −0.21	−0.14	−0.81, 0.54	−0.54	−1.37, 0.28	.19

Discussion

Principal Findings

In the present study, email-based encouragements to do 10 minutes of daily stair-walks together with colleagues resulted in 82.7% of the participants walking the stairs regularly, ie, at least 3 times per week. This resulted in improved aerobic fitness. This fairly high adherence shows that email-based encouragements to do daily stair-walks is a feasible way of implementing vigorous physical activity among employees in sedentary occupations.

The intention-to-treat analysis showed that aerobic fitness increased approximately 6% in the email group and 2% in the control group. Because aerobic fitness is an independent predictor of future cardiovascular disease [16,17], the results of the present study may be relevant from a preventive perspective. Although other studies using more comprehensive exercise protocols have found larger improvement in aerobic fitness [28], the simplicity and low cost of email-based encouragements to do daily stair-walks add positively to the arsenal of health-promoting initiatives to improve aerobic fitness of adults in sedentary occupations.

Most participants adhered to the email-based encouragement to do daily stair-walks, with 83% performing at least 3 sessions of 10 minutes per week, and a weekly average of more than 3 sessions, equivalent to 30 additional minutes of vigorous physical activity per week. The major reason for not participating as often as required was lack of time. Although this observation is consistent with Kruger et al's [29] general observation concerning participation in health promotion programs, we have no underlying knowledge whether the reports reflect an actual lack of time or a more socially accepted way to express a disinterest or even laziness. Interestingly, few participants reported colleagues or the nearest leader as reasons

for not participating. This seems to imply that important organizational circumstances were present and that the email-based encouragements to do daily stair-walks were generally accepted. Thus, a feeling of support from the employer may be important to successfully implement physical exercise by means of email-based encouragements at the workplace. The high acceptance level is also implied by the fact that almost 80% of the participants stated that they wished to continue with daily stair-walking after the project.

For health promotion, the ACSM and the AHA recommend moderate aerobic physical activity for at least 30 minutes on 5 days each week or vigorous aerobic physical activity for at least 20 minutes on 3 days each week [2]. However, many working-age adults lack the time for being physically active during leisure [7], and short bouts of physical activity at the workplace have been suggested as a solution [8]. The present study supports this notion and shows that email-based encouragements to do daily stair-walks at the workplace with colleagues can partly contribute toward reaching the recommended weekly dose of vigorous physical activity.

The participants in the subgroup with low fitness showed a 10% increase in aerobic fitness over 10 weeks of email-based encouragements to do daily stair-walks. This subgroup also showed a reduction in both systolic and diastolic blood pressure. Thus, the health benefits of email-based encouragements to increase physical activity appear to be higher among people with low fitness. Importantly, even small changes in blood pressure reduce the risk for cardiovascular mortality [30].

Adverse events in terms of knee or ankle/foot injuries directly related to the daily stair-walks were reported by 2 participants. Thus, a small risk for injury during stair-walks exists, which should also be considered by companies and employees when implementing such types of physical activity at the workplace. Nevertheless, stair-walking appears to be a relatively safe type of physical activity compared with different types of sports [31].

The intention of the weekly emails was to encourage short bouts of vigorous physical activity during the workday without the need for changing clothes and subsequent showering. Heart rate measurements clearly showed that the exercise intensity of stair-walking was high with heart rate reaching an average of 167 beats per minute. Approximately 3 of 4 participants (72%) did not experience sweating to a bothersome extent during stair-walking. However, only 5% did not experience sweating at all, whereas 23% experienced sweating to a bothersome extent. This may limit the feasibility of stair-walking at the workplace for approximately one-quarter of individuals because bothersome sweating may decrease motivation and require additional time for bathing. Moderate-intensity physical activities, for example, walking outside, which does not evoke sweating provides similar health-protective effects as more vigorous types of activity provided that the energy expenditure is equivalent [32]. Moderate-intensity activities, such as walking outside, although more time-consuming, may be preferred when sweating from vigorous activity is bothersome. Such activities can also be organized using Internet technology and reminder emails for the mutual benefit of employees and employers.

In the email group, 26% replied yes, 53% replied yes, maybe, and 20% replied no to the question about whether they wished to continue daily stair-walking after the research project. This indicates a certain potential for long-term sustainability of email-reminders to stimulate brief group-based physical activities at the workplace, but also shows that not all employees find stair-walking an attractive solution. Preferences and available time should be considered when implementing health-promoting physical activities at the workplace.

Strengths and Limitations

There are both strength and limitations of the present study. The simplicity and low cost of email-based encouragements to form groups with colleagues to do physical activity strengthens the potential for wide-scale implementation of the research results at other companies. The randomized controlled design, blinding of examiners, the high follow-up percentage, and the low level of initiation of other physical activities during the intervention period improves the validity of the present results. The test methods were highly reliable over a 10-week period, allowing for very precise estimates of the intervention. However, because of the study design, participants were not blinded. Also, the risk of cross-contamination exists because all participants

were from the same company. Based on the follow-up questionnaire, 7% of the participants from the control group participated in the 10-minute stair-walks, which may have slightly decreased between-group differences. Self-selection, ie, due to ethical reasons only volunteers are included, is a general limitation of randomized controlled trials. Thus, adherence may have been lower at the company level if all employees were included. Further, because of the email-based encouragement to form groups to do stair-walks, unknown cluster effects may have occurred. Further, not all office companies have accessible stairways, limiting the generalizability of the present findings to companies with accessible stairways. Finally, a limitation may be that the present study focused primarily on vigorous physical activity in terms of email-based encouragements to do stair-walks, and not specifically on breaking sedentary behavior several times during the working day. Future studies should combine both aspects (ie, reducing sitting time and increasing physical activity in sedentary populations) because they present distinct risk factors for adverse health outcomes [33]. Web or mobile applications to remind the employees to break sedentary behavior after a certain period, for example, every 30 minutes of continuous keyboard typing or mouse use, may be developed and tested in randomized controlled trials. Another target for future studies may be to investigate the effect of different theory-based motivators toward stair-walking (self-efficacy, goal setting, action planning, etc). This could be combined with research in delivering the encouragements in different ways (eg, emails, websites, text messages, smartphones) or with different frequencies (eg, daily, twice weekly, once weekly). Another possibility could be to develop smartphone applications interacting with the built-in 3D accelerometer of the phone to activate the buzz function after a certain period of inactivity to remind the employee to break sedentarism.

Conclusions

In conclusion, email-based encouragements to do 10 minutes of daily stair-walks together with colleagues improve aerobic fitness among office workers. The high adherence shows that email-based encouragements to do daily stair-walks is a feasible way of implementing vigorous physical activity among employees in sedentary occupations. The weekly duration of stair-walking performed in the present study corresponds to approximately half of the ACSM and AHA recommended weekly amounts of vigorous physical activity.

Acknowledgments

Thanks to Charlotte Saervoll, Mette Justesen, and Mikkel Brandt Petersen for practical help during the project.

Authors' Contributions

LLA designed and led the study. ES, MDJ, and MB collected the data. OSM was involved in the clinical examination. All authors were involved in the interpretation and analysis of data. LLA drafted the manuscript and all coauthors revised it critically for important intellectual content. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH V1.6.2 checklist [19].

[\[PDF File \(Adobe PDF File\), 985KB - jmir_v15i6e127_app1.pdf\]](#)**References**

1. World Health Organization. 2010. WHO and the International Olympic Committee sign agreement to improve healthy lifestyles URL: http://www.who.int/mediacentre/news/releases/2010/ioc_20100721/en/ [accessed 2013-06-14] [WebCite Cache ID 6HNW16qD4]
2. Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, Franklin BA, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Med Sci Sports Exerc* 2007 Aug;39(8):1423-1434. [doi: [10.1249/mss.0b013e3180616b27](https://doi.org/10.1249/mss.0b013e3180616b27)] [Medline: [17762377](#)]
3. Allender S, Peto V, Scarborough P, Boxer A, Rayner M. Physical activity. In: *Diet, Physical Activity and Obesity Statistics*. London: British Heart Foundation; 2006:42-66.
4. Centers for Disease Control and Prevention. 2009. US physical activity statistics URL: <http://www.cdc.gov/nccdphp/dnpa/physical/stats/index.htm> [accessed 2013-06-14] [WebCite Cache ID 6HNWckz73]
5. Bevan S, Quadrello T, McGee R, Mahdon M, Vovrovsky A, Barham L. Fit For Work Europe. London: The Work Foundation; 2009 Sep 1. Musculoskeletal disorders in the European workforce URL: <http://www.fitforworkeurope.eu/default.aspx.locid-0afnew009.htm> [accessed 2013-06-14] [WebCite Cache ID 6HNWnZztH]
6. Kahn EB, Ramsey LT, Brownson RC, Heath GW, Howze EH, Powell KE, et al. The effectiveness of interventions to increase physical activity. A systematic review. *Am J Prev Med* 2002 May;22(4 Suppl):73-107. [Medline: [11985936](#)]
7. Trost SG, Owen N, Bauman AE, Sallis JF, Brown W. Correlates of adults' participation in physical activity: review and update. *Med Sci Sports Exerc* 2002 Dec;34(12):1996-2001. [doi: [10.1249/01.MSS.0000038974.76900.92](https://doi.org/10.1249/01.MSS.0000038974.76900.92)] [Medline: [12471307](#)]
8. Barr-Anderson DJ, AuYoung M, Whitt-Glover MC, Glenn BA, Yancey AK. Integration of short bouts of physical activity into organizational routine a systematic review of the literature. *Am J Prev Med* 2011 Jan;40(1):76-93. [doi: [10.1016/j.amepre.2010.09.033](https://doi.org/10.1016/j.amepre.2010.09.033)] [Medline: [21146772](#)]
9. Robroek SJ, Lindeboom DE, Burdorf A. Initial and sustained participation in an internet-delivered long-term worksite health promotion program on physical activity and nutrition. *J Med Internet Res* 2012;14(2):e43 [FREE Full text] [doi: [10.2196/jmir.1788](https://doi.org/10.2196/jmir.1788)] [Medline: [22390886](#)]
10. Teh KC, Aziz AR. Heart rate, oxygen uptake, and energy cost of ascending and descending the stairs. *Med Sci Sports Exerc* 2002 Apr;34(4):695-699. [Medline: [11932581](#)]
11. Fardy PS, Ilmarinen J. Evaluating the effects and feasibility of an at work stairclimbing intervention program for men. *Med Sci Sports* 1975;7(2):91-93. [Medline: [1152630](#)]
12. Ilmarinen J, Rutenfranz J, Knauth P, Ahrens M, Kylian H, Siuda A, et al. [The effect of an on the job training program -- stairclimbing -- on the physical working capacity of employees (author's transl)]. *Eur J Appl Physiol Occup Physiol* 1978 Feb 21;38(1):25-40. [Medline: [631117](#)]
13. Ilmarinen J, Ilmarinen R, Koskela A, Korhonen O, Fardy P, Partanen T, et al. Training effects of stair-climbing during office hours on female employees. *Ergonomics* 1979 May;22(5):507-516. [doi: [10.1080/00140137908924634](https://doi.org/10.1080/00140137908924634)] [Medline: [520283](#)]
14. Meyer P, Kayser B, Kossovsky MP, Sigaud P, Carballo D, Keller PF, et al. Stairs instead of elevators at workplace: cardioprotective effects of a pragmatic intervention. *Eur J Cardiovasc Prev Rehabil* 2010 Oct;17(5):569-575. [doi: [10.1097/HJR.0b013e328338a4dd](https://doi.org/10.1097/HJR.0b013e328338a4dd)] [Medline: [20299999](#)]
15. Boreham CA, Kennedy RA, Murphy MH, Tully M, Wallace WF, Young I. Training effects of short bouts of stair climbing on cardiorespiratory fitness, blood lipids, and homocysteine in sedentary young women. *Br J Sports Med* 2005 Sep;39(9):590-593 [FREE Full text] [doi: [10.1136/bjsm.2002.001131](https://doi.org/10.1136/bjsm.2002.001131)] [Medline: [16118293](#)]
16. Blair SN, Kohl HW, Paffenbarger RS, Clark DG, Cooper KH, Gibbons LW. Physical fitness and all-cause mortality. A prospective study of healthy men and women. *JAMA* 1989 Nov 3;262(17):2395-2401. [Medline: [2795824](#)]
17. Mora S, Redberg RF, Cui Y, Whiteman MK, Flaws JA, Sharrett AR, et al. Ability of exercise testing to predict cardiovascular and all-cause death in asymptomatic women: a 20-year follow-up of the lipid research clinics prevalence study. *JAMA* 2003 Sep 24;290(12):1600-1607. [doi: [10.1001/jama.290.12.1600](https://doi.org/10.1001/jama.290.12.1600)] [Medline: [14506119](#)]
18. van Dongen JM, Proper KI, van Wier MF, van der Beek AJ, Bongers PM, van Mechelen W, et al. Systematic review on the financial return of worksite health promotion programmes aimed at improving nutrition and/or increasing physical activity. *Obes Rev* 2011 Dec;12(12):1031-1049. [doi: [10.1111/j.1467-789X.2011.00925.x](https://doi.org/10.1111/j.1467-789X.2011.00925.x)] [Medline: [21883870](#)]
19. 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](https://doi.org/10.2196/jmir.1923)] [Medline: [22209829](#)]
20. Andersen LB. A maximal cycle exercise protocol to predict maximal oxygen uptake. *Scand J Med Sci Sports* 1995 Jun;5(3):143-146. [Medline: [7552756](#)]

21. Åstrand I. Aerobic work capacity in men and women with special reference to age. *Acta Physiol Scand Suppl* 1960;49(169):1-92. [Medline: [13794892](#)]
22. Persson R, Cleal B, Bihal T, Hansen SM, Jakobsen M, Villadsen E, et al. Why do people with suboptimal health avoid health promotion at work? *Am J Health Behav* 2013 Jan;37(1):43-55. [doi: [10.5993/AJHB.37.1.5](#)] [Medline: [22943100](#)]
23. Whittaker S, Sidner C. Email overload: exploring personal information and management of email. In: CHI 96. 1996 Presented at: CHI 96; April 13-18, 1996; Vancouver, BC, Canada p. 276-283.
24. Kraut RE, Attewell P. Media use in a global corporation: electronic mailorganizational knowledge. In: Kiesler S, editor. *Culture of the Internet*. Mahwah, NJ: Erlbaum; 1997:323-342.
25. Andersen LL, Saervoll CA, Mortensen OS, Poulsen OM, Hannerz H, Zebis MK. Effectiveness of small daily amounts of progressive resistance training for frequent neck/shoulder pain: randomised controlled trial. *Pain* 2011 Feb;152(2):440-446. [doi: [10.1016/j.pain.2010.11.016](#)] [Medline: [21177034](#)]
26. Andersen LL, Jørgensen MB, Blangsted AK, Pedersen MT, Hansen EA, Sjøgaard G. A randomized controlled intervention trial to relieve and prevent neck/shoulder pain. *Med Sci Sports Exerc* 2008 Jun;40(6):983-990. [doi: [10.1249/MSS.0b013e3181676640](#)] [Medline: [18461010](#)]
27. Saltin B, Grimby G. Physiological analysis of middle-aged and old former athletes. Comparison with still active athletes of the same ages. *Circulation* 1968 Dec;38(6):1104-1115. [Medline: [5721960](#)]
28. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee IM, American College of Sports Medicine. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. *Med Sci Sports Exerc* 2011 Jul;43(7):1334-1359. [doi: [10.1249/MSS.0b013e318213feff](#)] [Medline: [21694556](#)]
29. Kruger J, Yore MM, Bauer DR, Kohl HW. Selected barriers and incentives for worksite health promotion services and policies. *Am J Health Promot* 2007;21(5):439-447. [Medline: [17515009](#)]
30. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R, Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* 2002 Dec 14;360(9349):1903-1913. [Medline: [12493255](#)]
31. Maffulli N, Longo UG, Gougoulas N, Caine D, Denaro V. Sport injuries: a review of outcomes. *Br Med Bull* 2011;97:47-80. [doi: [10.1093/bmb/ldq026](#)] [Medline: [20710023](#)]
32. Brown WJ, Burton NW, Rowan PJ. Updating the evidence on physical activity and health in women. *Am J Prev Med* 2007 Nov;33(5):404-411. [doi: [10.1016/j.amepre.2007.07.029](#)] [Medline: [17950406](#)]
33. Thorp AA, Owen N, Neuhaus M, Dunstan DW. Sedentary behaviors and subsequent health outcomes in adults a systematic review of longitudinal studies, 1996-2011. *Am J Prev Med* 2011 Aug;41(2):207-215. [doi: [10.1016/j.amepre.2011.05.004](#)] [Medline: [21767729](#)]

Abbreviations

ACSM: American College of Sports Medicine

AHA: American Heart Association

BMI: body mass index

ICC: intraclass correlation coefficient

MPO: maximal power output

VO_{2max}: maximal oxygen uptake

Edited by G Eysenbach; submitted 04.09.12; peer-reviewed by M Kwan; comments to author 25.09.12; revised version received 02.10.12; accepted 08.06.13; published 21.06.13.

Please cite as:

Andersen LL, Sundstrup E, Boysen M, Jakobsen MD, Mortensen OS, Persson R

Cardiovascular Health Effects of Internet-Based Encouragements to Do Daily Workplace Stair-Walks: Randomized Controlled Trial *J Med Internet Res* 2013;15(6):e127

URL: <http://www.jmir.org/2013/6/e127/>

doi: [10.2196/jmir.2340](#)

PMID: [23793032](#)

©Lars Louis Andersen, Emil Sundstrup, Marianne Boysen, Markus Due Jakobsen, Ole Steen Mortensen, Roger Persson. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 21.06.2013. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of

Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

A Service-Based Evaluation of a Therapist-Supported Online Cognitive Behavioral Therapy Program for Depression

John Sharry¹, BA(Hons), MSc, DPsych; Ruth Davidson², MLitt, BA(Hons); Orla McLoughlin³, BA(Hons), MSc; Gavin Doherty⁴, BA(mod), DPhil

¹Mater Misericordiae University Hospital, Dublin, Ireland

²University of Limerick, Limerick, Ireland

³Counselling Service, Trinity College Dublin, Dublin, Ireland

⁴School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland

Corresponding Author:

Gavin Doherty, BA(mod), DPhil
School of Computer Science and Statistics
Trinity College Dublin
O'Reilly Institute
Dublin, D2
Ireland
Phone: 353 1 8963858
Fax: 353 1 6772204
Email: Gavin.Doherty@tcd.ie

Abstract

Background: Evidence suggests that Internet-delivered cognitive behavioral therapy (CBT) may be as effective as face-to-face delivery for depression, but attrition and engagement rates remain a challenge.

Objective: This service-based study aimed to evaluate an online, therapist-supported, CBT-based program for depression. The program was specifically designed to address engagement issues, most notably by integrating online therapist support and communication within the platform.

Methods: Participants were 80 adults who were registered university students. Participants used the modular online program over 8 weeks, supported by a therapist. Engagement information was gathered automatically by the online system, and analyzed for all participants. Severity of participants' self-reported symptoms of depression were assessed preintervention and postintervention using the Beck Depression Inventory-II (BDI-II). Postintervention measures were completed by 53 participants.

Results: A high level of engagement was observed compared to a previous study within the same service, along with extensive use of a range of program features. A statistically significant ($P<.001$) decrease in self-reported depressive symptomatology from preintervention (mean BDI-II 25.47) to postintervention (mean BDI-II 15.53) with a large effect size ($d=1.17$) was also observed.

Conclusions: The results indicate the potential of unintrusive and easily provided online support to enhance engagement with online interventions. The system described in the paper also illustrates how such online support can be tightly integrated with interactive online programs by using a range of design strategies intended to improve the user experience.

(*J Med Internet Res* 2013;15(6):e121) doi:[10.2196/jmir.2248](https://doi.org/10.2196/jmir.2248)

KEYWORDS

Internet; user-computer interface; depression; cognitive behavioral therapy; patient adherence; online interventions; guided online program; online therapist support; user experience

Introduction

Background

Many adults experience depression in their lifetime, but a relatively small number seek or receive effective or

evidence-based treatment for their difficulties [1]. This may be due to a number of factors, including stigma associated with seeking help for mental health difficulties, a lack of knowledge about different treatment options, as well as logistical or fiscal reasons. Computerized or Internet-delivered treatment programs are the latest and most innovative approach to improving access

to psychological treatments for mental health difficulties, including depression. A large number of programs have been developed and evaluated to date, targeting a variety of disorders, including depression, panic disorder, generalized anxiety disorder, phobias, obsessive compulsive disorder, eating disorders, and addictions [2]. Computerized programs have used a variety of therapeutic approaches, although cognitive behavioral therapy (CBT) is the most common treatment approach in such programs. CBT is perhaps particularly suited to adaptation to a computerized medium because of its well-structured protocol and clearly delineated activities and homework exercises [3].

Existing Evidence Base

The integration of computerized interventions into mainstream mental health services offers a number of benefits, including the means to increase capacity in mental health services, reduce costs, and improve convenience to therapist and client [4,5]. In addition, online interventions have the ability to surmount geographical barriers as well as barriers associated with stigma that may prevent clinic attendance [6,7]. Not only is it important to establish that computerized interventions do what they are intended for, but it is also crucial that they are designed to be engaging, usable, and acceptable to those who need them [8], otherwise these factors may mask the effectiveness of the intervention. Much of the literature is concerned with 2 themes of enquiry which are central to the development of a clinically useful body of knowledge related to computerized psychotherapeutic interventions.

Clinical Efficacy

The first theme relates to the actual clinical efficacy of the computerized intervention. This line of enquiry seeks to ascertain if completion of the intervention leads to any significant increase in well-being or decrease in distress for the user. Typically, this manifests as an outcome study that assesses symptomatology and other clinically relevant variables preintervention and postintervention. Results are promising; in a comprehensive meta-analysis of the effectiveness of Internet-based psychotherapeutic interventions, Barak et al [2] reported a similar effect size to that expected with more traditional face-to-face therapy. A review of the literature concerned solely with depression tallies with this [9]. Multiple studies and meta-analyses conclude that computerized and online psychotherapeutic interventions (most CBT-based) offer a promising and potentially effective treatment of depression and anxiety [6,7,10-12]. Many researchers highlight, however, that caution is advised in generalizing these conclusions too widely or too prematurely, and emphasize that more systematic research is needed to clearly delineate the precise factors that contribute to the effectiveness of online and computerized interventions. For instance, efficacy can vary according to factors, such as level of therapist support, the presenting difficulty of the client, the length of treatment, the age of the client, the media and technology used to deliver the program, and the interactivity of the program [2,6]. Even seemingly small differences among programs and their delivery can have implications for outcomes. In particular, research suggests that even a minimal amount of therapist support can increase the efficacy of an online or

computerized program [7,9], particularly in the treatment of depression [5]. Some research suggests both supported and unsupported online interventions may be effective [13,14]. A recent meta-analysis of online interventions for depression highlighted the superiority of supported interventions in terms of outcomes, but noted that unsupported interventions may also have a useful role to play in treatment provision [15]. At present, it seems reasonable to conclude that some form of support has an added beneficial effect for program outcomes, but the absolute necessity of therapist support, the optimum level of such support, or its precise nature remains to be established [5,13,15]. It is unclear which elements of the guidance or support are responsible for any observed benefits, with Andersson et al [16] suggesting that positive reinforcement for work completed by the client may be the most important contribution of the person supporting the online intervention.

Engagement

The second theme of enquiry relates to the issue of how engaging computerized interventions are and how willing clients are to use them as they are intended. Related to this, researchers have also begun to examine the characteristics of the individuals most likely to both engage with and benefit from a computerized intervention [1,17,18]. Although attrition from psychotherapy is an issue that affects services offered in more traditional ways [19,20], it has been repeatedly highlighted as a significant issue in online interventions in both research and practice [14,21-23]. There are a number of factors that may predict engagement with and adherence to online interventions, including factors related to the program itself and how it is delivered [2,23,24], as well as characteristics of the program users, such as age, gender, education, duration and severity of psychological distress, and personal psychological factors [20,23-26]. To improve engagement, a number of researchers have successfully included direct therapist support alongside the Internet intervention, such as motivational interviewing and regular telephone conversations with therapists [7,9], suggesting that human contact is vital for achieving high levels of engagement. A meta-analysis of the literature has suggested that although overall dropout rates from online programs for depression may exceed 50% [15], this can be substantially reduced by the addition of some form of support or guidance [14].

A Therapist-Supported Online Program

We can see from the literature that the efficacy of online programs can vary as a function of delivery and context, and that difficulties remain surrounding engagement and attrition. This motivates development and exploration of innovative strategies for improving engagement in realistic service environments. To this end, we present an initial clinical evaluation of a highly interactive therapist-supported Internet-based intervention for depression, namely the Mind Balance program. We report the engagement patterns of clients using the therapist-supported program, and outcome data on changes in self-reported depression symptomatology following use of the program.

Program Content

Mind Balance is a 7-module online CBT-based intervention for depression, delivered on a Web 2.0 platform using media-rich interactive content. The structure and content of the program modules follow evidence-based principles of a traditional CBT program, incorporating ideas from mindfulness. The content of each module is described briefly in [Table 1](#). Each module is structured in an identical way and incorporates introductory quizzes, videos, informational content, interactive activities, as well as homework suggestions and summaries. In addition, personal stories and accounts from other clients are incorporated into the presentation of the material.

Engagement Strategies Employed Within the Program

The Mind Balance program employs a number of innovative engagement strategies for improving the user experience, described subsequently. These features are also presented in the SilverCloud Platform Overview video ([Multimedia Appendix 1](#)). The design emerged from an extensive user-centered design process and was refined through a number of formative evaluations.

Personal

The client has their own secure home page ([Figure 1](#)), which is about who they are and where they are in the program. For example, the client can fill in a profile with basic information about themselves, such as age and interests. This establishes a sense of ownership, and provides useful information for the therapist, allowing them to provide feedback that is more

personal. The home page is intended to provide a reflective space; the client can document their thoughts and feelings, and these can be elaborated on within the journal application, which also acts as the vehicle for therapeutic writing exercises. The user has actions suggested to them, and as they complete modules of the program, their achievements are noted. Users are free to access the modules in any order they wish, in either a linear or nonlinear manner, contributing to a sense of empowerment. A range of satellite applications are provided along with the central content, such as a goal-setting application, which can be used independently of the program content. Applications are released as the user goes through the program, with the intention of maintaining engagement by introducing new features over time and not overwhelming the client initially. Clients can also control which applications appear on their home page.

Interactive

The program includes a number of interactive elements and graphical exercises (an example is shown in [Figure 2](#)), which are aimed at engaging clients with the therapeutic content, for example, reflecting on their own thinking. Clients also have the ability to respond to content, indicating whether they like it, and also to comment on it. The client is provided with immediate feedback wherever possible; for example, when a charting exercise, such as a mood chart, is completed, the application item is graphically updated on the home page. Likewise, items are ticked off on the to-do list when completed, and achievements unlocked in each module summary.

Table 1. Mind Balance: description of module contents.

Module name	Brief description
Getting Started	Outlines the basic premise of CBT, some information about depression, and introduces some of the key ideas of Mind Balance. Users are encouraged to begin to chart their own current difficulties with depression.
Tune In 1: Getting to Grips with Mood	Focuses on mood monitoring and emotional literacy. Users can explore different aspects of emotions, physical reactions, action and inaction, and how they are related.
Tune in 2: Spotting Thoughts	This module focuses on noting and tracking thoughts. Users can explore the connection between their cognitions and their mood, and record them graphically.
Change It 1: Boosting Behavior	This module focuses on behavioral change as a way to improve mood. Ideas about behavioral activation are included, and users can plan and record activities, and chart their relationship with their mood.
Change It 2: Challenge Your Thoughts	This module supports users to challenge distorted or overly negative thinking patterns, with thought records, as well as helpful coping thoughts.
Change It 3: Core Beliefs	This module outlines the role that deeply held core beliefs can play in mood and depression. Users can use a range of interactive activities to identify, challenge, and balance any unhelpful core beliefs.
Bringing It All Together	Users are encouraged to bring together all the skills and ideas they have gathered, note their personal warning signs, and make a plan for staying well.

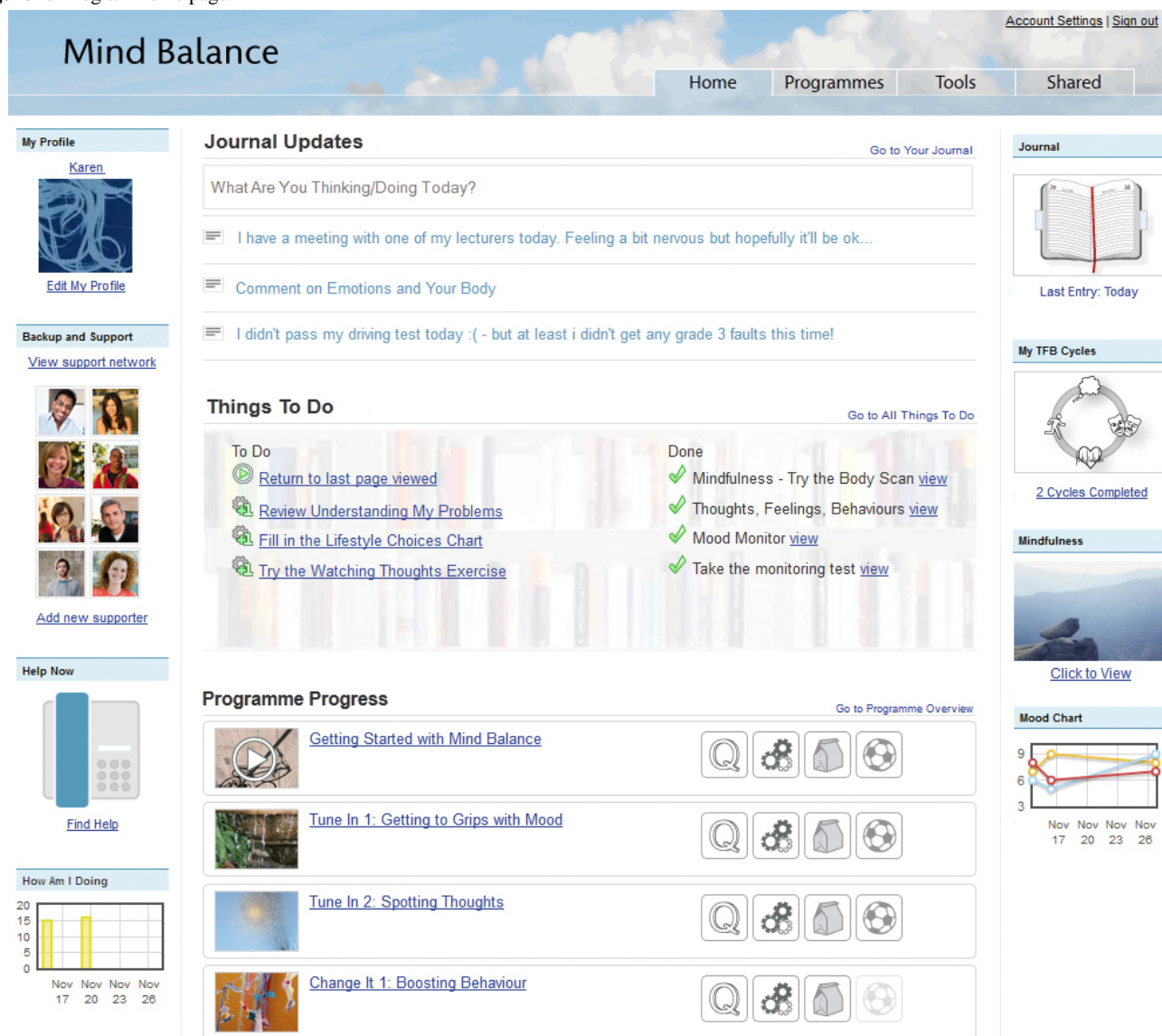
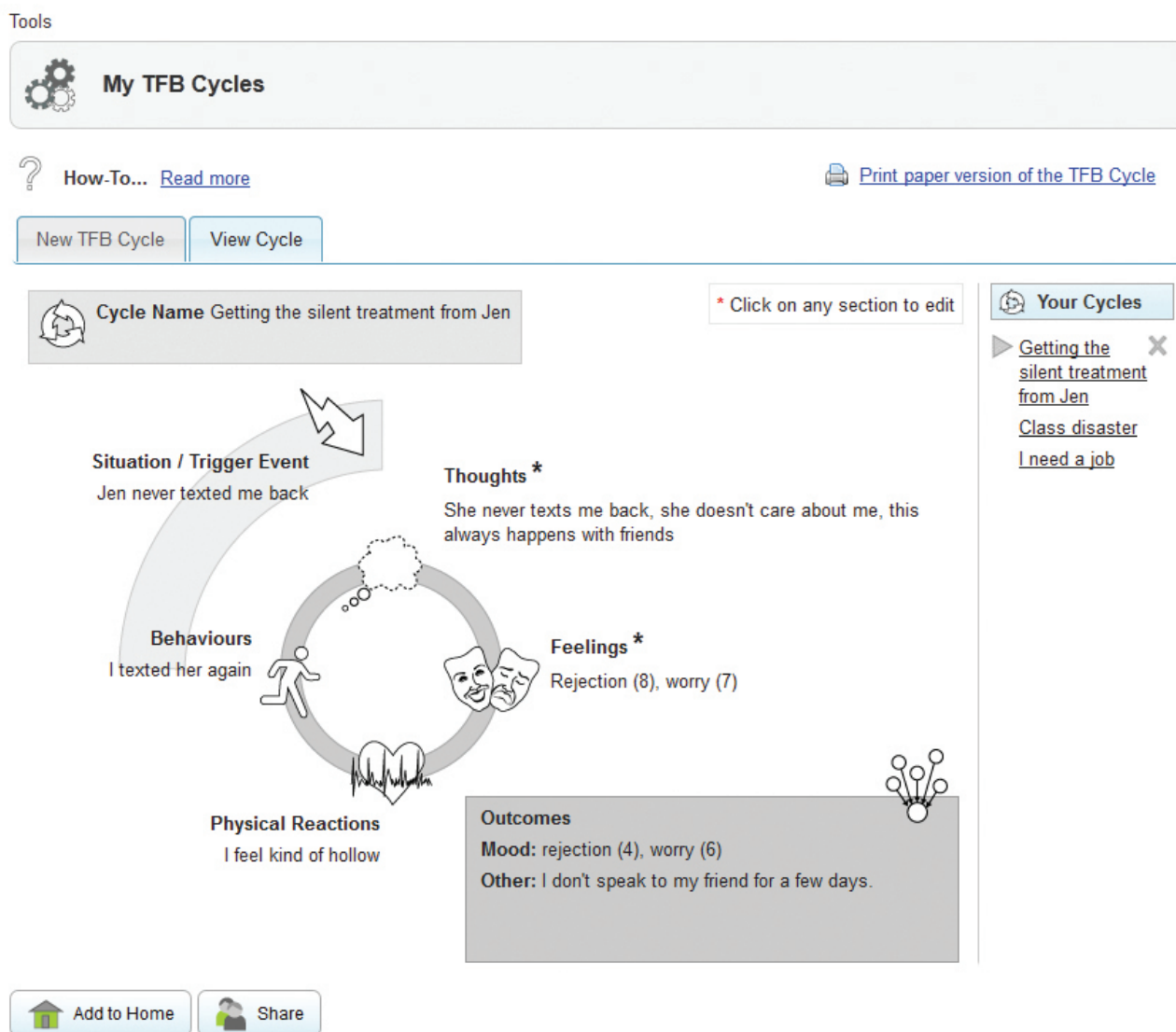
Figure 1. Program home page.

Figure 2. Example of an interactive exercise: thoughts, feelings, and behaviors.

Supportive

Each client has an assigned therapist, who provides weekly reviews of their progress on the program. This support is asynchronous, whereby the therapist sets a date to review their client's progress, and they do not provide feedback, support, or contact outside this time. This is an important program feature in terms of maintaining the therapeutic boundary in the online space. The therapist can support multiple clients, logging in once weekly for instance, and reviewing the work of all their online clients within an allocated time period. Such asynchronous online contact may be logistically easier to implement for many services compared with motivational interviewing and telephone support, and should be more attractive to users who are unwilling to have direct contact with a service. The system supports the exchange of messages between the client and therapist, but goes beyond email because the client is encouraged to share their content (such as completed exercises and comments) with their therapist (see [Figure 2](#)). This shared content allows the therapist to respond in a more personal way and provide guidance as well as encouragement to keep using the program. Adherence information is also

available to the supporter, and they can keep track of the client's progress. This is all personally sensitive information; for this reason, a shared view is provided in the client interface where they can see the therapist's view of their data. By making the visibility of their data to the therapist more transparent, as well as the ability to explicitly change the sharing status of data, the client is provided with a greater sense of control while facilitating a meaningful interaction with the therapist.

Social

Although group therapy and peer group support are well established, introducing contact with other clients within any online system raises a number of ethical concerns regarding the possibility for unhelpful or negative content or communications. As a first step, the client can see anonymous indications of other people in the system. The intention is to reassure clients that they are not alone in experiencing difficulties and that many other people have experienced similar problems and overcome them. Users can respond to content by indicating that they "like" it, and can see how many other people liked it, helping to reduce the sense of isolation. More detailed shared content (such as tips and ideas) is subject to therapist moderation.

A key innovative aspect of the program is the way in which these different strategies are combined within the overall design. For example, the interactive features of the program generate content which can be shared with the therapist (enhancing the quality of feedback which can be provided), and in some cases with peers (further enhancing the experience of other users).

Methods

A service-based design was used to assess engagement with the Mind Balance program delivered by the mental health service of an Irish university (Trinity College, Dublin). As a first exploratory trial, it was felt that a service-based design would be informative regarding the potential of the program to engage clients, provide initial evidence regarding the potential effectiveness of the treatment, and establish the feasibility of incorporating low-intensity online therapist support in a real-world environment. Part of the appeal of the program to this population was expected to be the relative anonymity of the online mode of delivery; therefore, face-to-face contact was not required to sign up for the program. The protocol for use was kept as close as possible to the actual practice for use of online interventions within the service; this degree of fidelity was borne out by the subsequent incorporation of the program into the day-to-day practice of the service following the same protocol. Client log-ins and use of program features were automatically monitored. Changes in symptoms of depression were also assessed through measures taken before and after use of the program. Ethical approval was secured from the Research Ethics Committee of Trinity College prior to the study.

Recruitment

Participants were recruited in 2 ways as the program was rolled out, in parallel with the usual practice of the counseling service.

Self-Sign-Up

Recruitment began in January 2011, with all first- and third-year students emailed through the university emailing system to inform them of the opportunity to receive this online intervention as part of a research study. In September 2011, all fourth-year and postgraduate students were similarly emailed. Finally, in

January 2012, all first years were again emailed to inform them of this opportunity. This rolling recruitment model was used to manage the numbers of students simultaneously availing of the intervention because of limited therapist resources. Specific-year groups were chosen at various times based on existing records of service engagement patterns in the student counseling service. Participants were informed that they would be required to register with the student counseling service in order to avail of the program. Interested students were instructed to follow a Web link embedded in the text of the email and were directed from there to the sign-up page. Students could also access this Web link from the University's online mental health portal. Once an individual had navigated to the sign-up page, they were presented with a comprehensive information sheet and an electronic consent form. Following confirmation of informed consent, students were required to complete the preprogram measures electronically and were then free to explore the online program as they wished. Postintervention measures were sent out automatically after completion of the final review.

Directly Added by Face-to-Face Therapist

A number of participants were referred directly to the program by staff in the counseling service. Those receiving face-to-face therapy in conjunction with the program were excluded from the research sample.

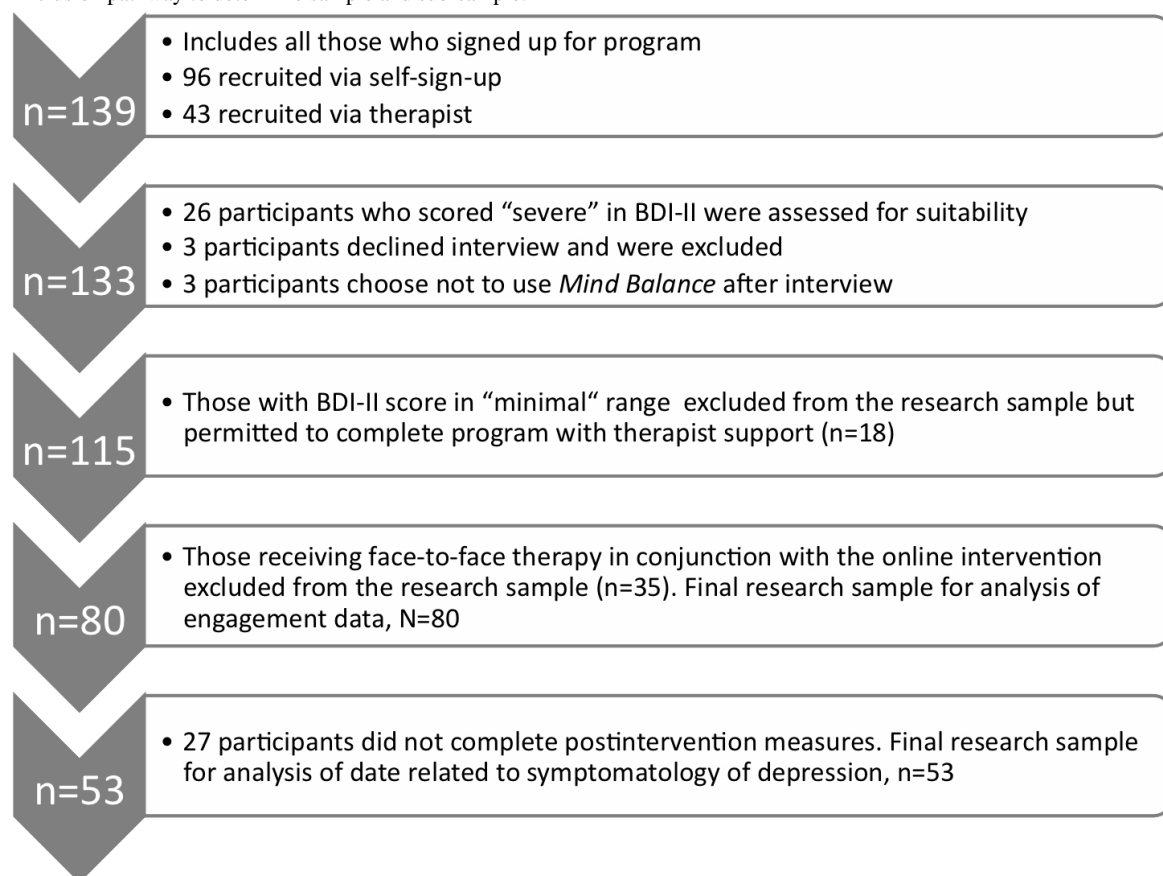
Participants

Participants were 80 students who identified themselves as having an existing difficulty with low mood, mean age 23.29 years (SD 4.84) and 69% female, who were all part of the population supported by the student counseling service. All of the participants were registered students. Demographic details for the main and subsample are detailed in [Table 2](#).

Inclusion criteria were as follows: (1) at least 18 years of age, (2) having at least self-reported mild symptoms of depression with a Beck Depression Inventory-II (BDI-II) score ≥ 14 , and (3) have received no face-to-face therapy since program commencement. A breakdown of how this final sample was reached through exclusions from the initial sample is provided in [Figure 3](#).

Table 2. Demographic details of the sample and subsample for pre-post analysis.

Group	Age, mean (SD)	Age range	Male, n (%)	Female, n (%)
Overall sample (N=80)	23.28 (4.84)	18-46	25 (31%)	55 (69%)
Pre-post subsample (n=53)	22.69 (5.17)	18-46	13 (25%)	40 (75%)

Figure 3. Exclusion pathway to determine sample and sub-sample.

Ethical Considerations

Telephone and email contact details for support services (including the counseling service, public emergency services, and relevant non-government organizations) for crises or out-of-hours emergencies were provided to users in the "help now" section of the website, which was visible at all times regardless of the page the user had navigated to. Any individual who scored 29 or more on the BDI-II or who scored 2 or more on BDI-II item 9 (suicidality) in their online screening was automatically alerted that a counselor would get in touch with them. They were telephoned within 1 working day to invite them for an interview, and contacted by email if they could not be reached by telephone. The screening interviews were carried out by a counseling psychologist and focused on the client's own description of the problem, including duration of the problem and their understanding of the causes, their psychiatric or mental health history, current alcohol and drug use, current health and functioning (sleeping, eating, attending lectures, keeping up with commitments, etc), and current support network. A clinical decision was made on the basis of this assessment in relation to whether the client could go ahead and use the online program on a standalone basis. Alternatively, it was recommended that they attend face-to-face counseling instead of or in conjunction with the online program. Three clients declined to attend for interview and were excluded. Three others chose not to use the program following the interview and were excluded. Those receiving face-to-face therapy combined with the program were not included in the research sample presented here. Although the primary concern was client safety,

the approach taken was one which would be applied in the normal practice of the service.

Measures

Engagement

The online system collected anonymized descriptive information relating to engagement and usage. Data collected included the number of sessions completed, mean time spent on the program, average number of sessions per client, and average length of a session. A session was defined as an instance when a client logged on to the system. Session time estimation will always be an imperfect calculation because users may be interrupted or take breaks within a session, and may not formally log out of the system. All client activity within the system, such as reading a content page, saving a journal entry, or updating an activity, was logged with a time stamp. Starting with the log entry of the client logging on, the total time was calculated by adding up the time that elapsed between each subsequent log record. On its own, this yields a result vulnerable to overestimation of session time. To avoid counting periods when the user was not actively engaged with the system, any interactions taking longer than 30 minutes were counted as 1 minute. Any period of inactivity longer than 3 hours started the count on a new session, rather than extending the time of the current session. Use of different program components was also measured. Data related to therapist reviews were also collected. In the interests of examining engagement and usage patterns in a fair and meaningful way, all 80 participants were included in that analysis regardless of whether or not they had completed

the BDI-II postintervention. Dropout is analyzed in terms of the last time a participant used the system.

Symptomatology of Depression

The BDI-II was used to assess the severity of self-reported symptomatology of depression preintervention and postintervention. Cut-off scores suggested for the BDI-II are minimal symptoms: 0-13; mild symptoms: 14-19; moderate symptoms: 20-28; and severe symptoms: ≥ 29 [27]. Reliability, validity, and test-retest reliability have been established across various populations and cultural groups [27-31].

All data were collected and stored electronically on a secure server.

Program Delivery

Following initial sign-up, clients were assigned a therapist who would support them online. This happened automatically and randomly in the case of those clients recruited via self sign-up. In the case of those who signed up following a recommendation by a face-to-face therapist, the same therapist was assigned as their online supporter. While the program suggested that clients complete 1 module per week, clients were free to dictate the pace and order of their work. Therapists were asked to provide 7 weekly reviews of each client's work. Allowing for 1 final session for the client to consider the final review, this corresponds to a target of 8 sessions per client, and also matches the target of 8 sessions in the program previously used within the service evaluated in Richards [32]. The service scheduled 1 hour per week for the 6 participating therapists to complete their Mind Balance reviews. The program content was identical for each client, but each therapist was free to provide reviews congruent with their own therapeutic persuasion. Therapists were given training and a treatment manual containing a broad guide of how to respond in their reviews (supporting progress, giving encouragement, specific feedback on activities shared), but could be flexible if they wanted to add to this. For the most part, this was only expected to happen when the users shared large amounts of personal content. In these cases, therapists were advised that this was not email counseling, and so very lengthy detailed responses were not advised. A dashboard interface was provided to therapists that gave an overview of their clients and the overall level of engagement of each, with additional detailed engagement information and shared data available for each client. In the interests of transparency, this information was available to clients under a "shared" tab in the interface. When the client was not engaging with the system, more generic messages could be used, and a template system

was provided to facilitate this. All therapists held postgraduate qualifications and consisted of 3 counseling psychologists; 1 CBT therapist; 1 marriage, family, and child counselor; and 1 psychotherapist. Clients were informed at each review of the precise date for the next review. If the therapist wished, they were free to continue to provide reviews for their client past the seventh recommended review, if they saw this as clinically justified, again with the aim of allowing a natural service-based pattern of use. Email technical support was available to both clients and therapists.

Results

Engagement

Figure 4 illustrates the level of engagement with the program for the overall sample (N=80). This graph depicts the number of users who were engaged at a given week or later, and conversely the week at which users dropped out. Looking at the graph, it is evident that relatively few users dropped out in the initial weeks of the program, and 79% were engaged at the target week 8 or later.

Table 3 presents key descriptive statistics relating to program usage. An examination of the log data reveals that 50% of reviews were read by the recipient (client) within 24 hours. Therapists reported that reviewing a client's work took 10 to 15 minutes per review on average, varying by the degree of engagement and content shared. The template feature allowed therapists to quickly send more generic encouragement to clients not logging in to the program.

Viewing the engagement data by BDI-II category (Table 4), clients in the severe category spent more time on average in the program than clients in the mild and moderate categories. A demographic breakdown of those reaching the target of 8 sessions is presented in Table 5.

Usage of different program features is illustrated in Figure 5. These refer to the percentage of clients sending notes to the therapist, sharing content with the therapist, viewing the therapist review, commenting on the content, performing a charting exercise, completing a psychoeducational quiz, setting a goal within the goal-setting tool, giving feedback on what they thought of a module, entering their own take-home point, "liking" content, sending brief updates via short message service (SMS) text messaging, writing within the journal, making brief updates from the home page, filling in the backup and support network, and filling in personal information in the profile.

Table 3. Usage statistics for overall sample (N=80).

Statistic description	Mean (SD)	Range
Time (minutes:seconds)		
Length of time per session	11:35 (18:00)	0:00-125:11
Total time spent on program	151:09 (141:48)	0:00-712:09
Sessions		
Number of sessions per user	11.95 (9.79)	1-53
Reviews		
Number of reviews per client	7.4 (4.2)	2-33

Table 4. Usage by Beck Depression Inventory-II (BDI-II) category for the overall sample (N=80).

BDI-II category	Mild (n=19)	Moderate (n=41)	Severe (n=20)
Time (minutes:seconds), mean	120:33	132:43	231:37
Sessions, n	9	10.6	18
Target of 8 sessions, %	47	54	95

Table 5. Demographics of session target completers and noncompleters.

Number of sessions	Age (years), mean (SD)	Age range	Male, n (%)	Female, n (%)
≥8 sessions	22.6 (5.1)	18-46	13 (26%)	37 (74%)
<8 sessions	24.5 (4.4)	18-34	12 (40%)	18 (60%)

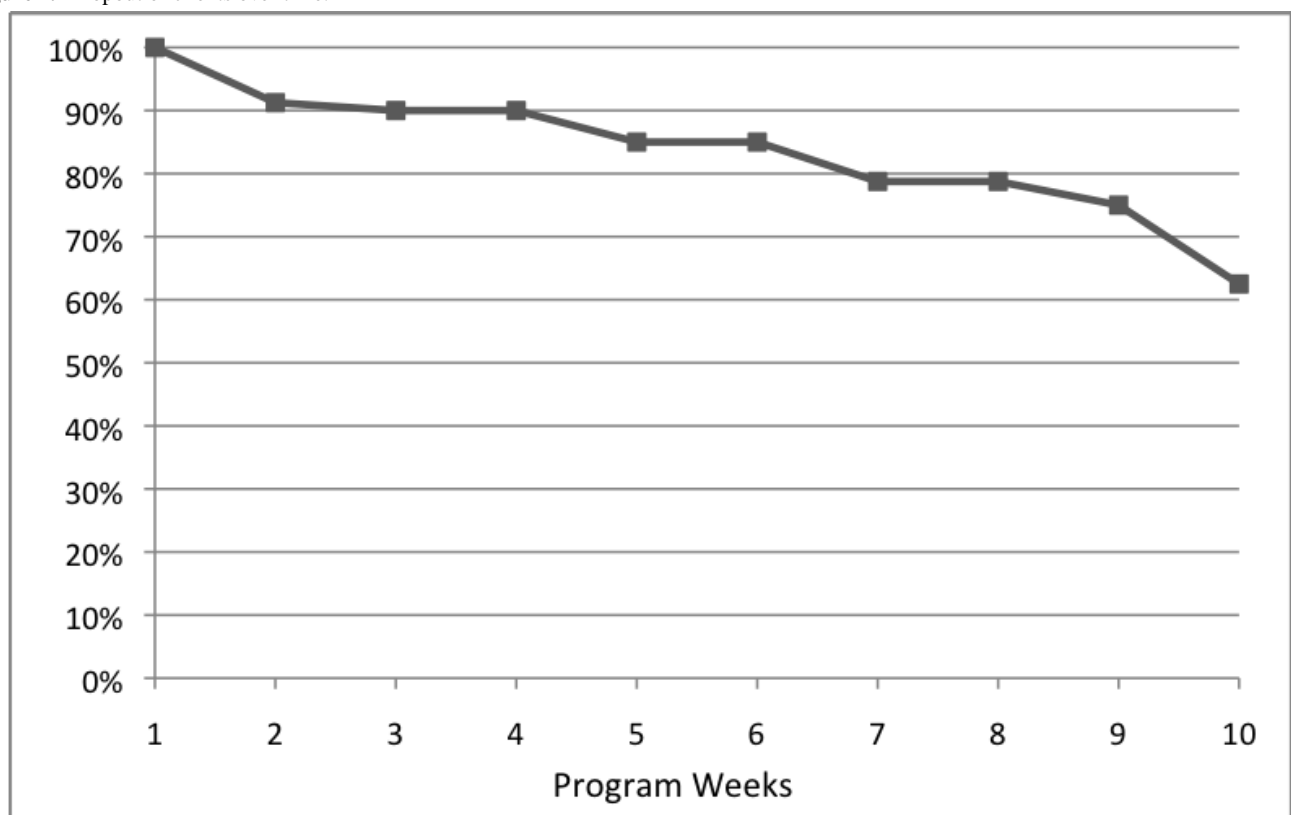
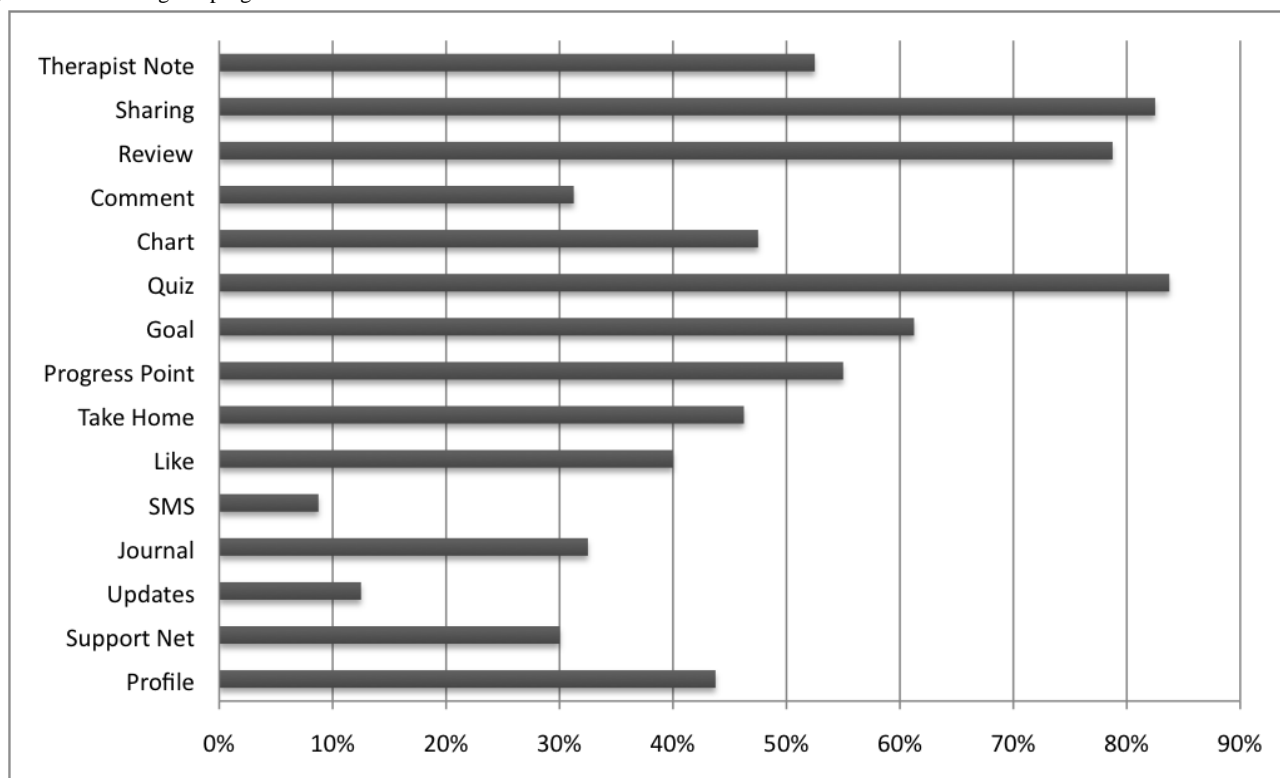
Figure 4. Dropout of clients over time.

Figure 5. Client usage of program features.

Change in Symptomatology of Depression

The reliability of the BDI-II was assessed using Cronbach alpha and was found to be in the acceptable range for both the overall sample ($N=80$) and the subsample ($n=53$), being Cronbach $\alpha=.78$ and Cronbach $\alpha=.79$, respectively. This indicates satisfactory internal reliability for the BDI-II for the present sample. Table 6 presents the mean BDI-II scores for the present sample both preintervention and postintervention, as well as results of the repeated measures t tests of difference.

As illustrated in Table 6, results indicate a statistically significant ($P<.001$) decrease in symptomatology of depression,

as assessed by the BDI-II, from preintervention to postintervention. Cohen's d was calculated as $d=1.17$, suggesting a large pre-post effect size for BDI-II [33]. An independent samples t test revealed no significant difference between preintervention BDI-II scores of those who completed postintervention measures ($n=53$) and those who did not ($n=27$), $t_{78}=-0.382$, $P=.70$. Table 7 outlines the number and percentage of participants in each category of severity of symptomatology preintervention and postintervention for the subsample ($n=53$) only. Postintervention, 74% of participants' ($n=39$) BDI-II scores had reduced and placed them in a lower category of severity (eg, from moderate to mild) than that suggested by their preintervention BDI-II score.

Table 6. Comparison of BDI-II scores preintervention (pre) and postintervention (post) and repeated measures t test.

BDI-II score	Pre, mean (SD)	Post, mean (SD)	t_{52}	P value
Overall sample ($n=80$)	25.71 (7.80)	—	—	—
Subsample ($n=53$)	25.47 (7.95)	15.53 (9.06)	7.70	<.001
No post-BDI ($n=27$)	26.19 (7.64)	—	—	—

Table 7. Participants in each severity category preintervention and postintervention (subsample only).

BDI-II category (score range)	Preintervention, n (%)	Postintervention, n (%)
Minimal (0-13)	0 (0)	28 (53)
Mild (14-19)	13 (25)	7 (13)
Moderate (20-28)	23 (43)	11 (21)
Severe (29+)	17 (32)	7 (13)

Discussion

Engagement

Poor treatment adherence and dropout have been repeatedly highlighted in the literature as significant issues to overcome in the delivery of Internet interventions [20,21]. Looking at the present study, it is evident that dropout for this interactive therapist-supported program is gradual up to the end of the supported treatment period. Viewing dropout in terms of the number of participants not engaged at session 8, the dropout rate for the present study stands at 37.5%. This compares favorably with the reported dropout rate of 74% for an unsupported program [32]. Given that the interventions were delivered in the same setting and with a similar population, this implies that delivering online programs incorporating online therapist support may enhance engagement and completion rates. This concurs with previous research [23,24] that suggests characteristics of program delivery, such as the presence of therapist support, may significantly influence client engagement with online programs. As a further point of reference, internal statistics within the service put average session attendance for face-to-face treatment at 3.49 and 3.73 sessions over the 2 years that this study took place. Approximately 12% of clients attended for 8 sessions or more.

On the demographic breakdown, we must note that this is a relatively homogeneous population compared with what may be seen in other environments. The student population supported by the service has a majority of women (59% in 2010/2011), and the greater number of women signing up for the program mirrors the use of the service generally. The incidence of depression also differs by gender. The size of the sample for male participants precludes us from making any statements regarding the lower number of males completing the target, but this issue would be worth investigating in future studies.

We can see from Figure 5 that key features regarding therapist support, such as sharing, therapist notes, and therapist reviews, were used by most users. However, social features, such as the like button, also received significant numbers of users, and the progress point, take-home point, and goal-setting features, which required more active participation, were also used by many users. The SMS text messaging feature that allowed users to make brief updates via mobile phone was only used by 7 of 80 users (9%). The Mind Balance program did not place emphasis on these updates, and further work is needed on how best to integrate mobile support into this type of intervention. Another interesting direction for analysis would be to investigate the relationship between viewing the content of particular modules and usage of the other features of the program. Overall, we can see from this data that the interactive features of the program were used to a significant extent, and that the system was not used simply as a multimedia delivery medium. Features within the program intended to improve engagement cover many, but not all, of the persuasive framework elements presented in Kelders et al [34]. As such, there are further design opportunities to be explored within this form of intervention, particularly in the area of social support.

Therapist Support

The primary aim of online interventions is to maintain clinical gains while reducing the amount of therapist time required per client. It is worth reflecting on the precise nature of the support needed to maintain an optimum pre-post clinical effect as well as satisfactory completion rates [5]. A number of studies have used therapist support that is reasonably time consuming (eg, motivational interviewing or weekly telephone calls), which may counteract some of the benefits of using online interventions. Even for low-intensity interventions using telephone support in which the overall amount of therapist contact time is similar (eg, 64 minutes [35]), it can be argued that the overhead in the online case is likely to be lower because of the lack of scheduling problems, and the flexibility available to the therapist in terms of when to perform the reviews. The system integrated therapist support into the same online space to maximize efficiency and ease-of-use for the therapist, while still maintaining the clinical benefits of therapist input. The fact that pre-post outcome effect size and completion rates for the present study were satisfactory despite using a relatively low level of therapist input is encouraging. Investigating the cost-effectiveness of this form of support would be an interesting direction for future work.

The nature of the support given by the therapist also deserves further exploration. Taking persuasive technology as encompassing both direct interaction with a computer system and computer-mediated communication [34], there is a need to explore and validate models for Internet interventions [36] that are useful for design and that encompass this human support. An initial theoretical model for human-supported eHealth interventions has been proposed by Mohr et al [37] that provides a number of testable hypotheses. This model also generates a number of useful suggestions regarding the nature of the support or coaching provided to clients. Exploring and refining such a model together with emerging best practice surrounding this program constitutes a promising direction for future work. However isolating the effects of a large number of interdependent program features is a significant challenge and motivates the development of new approaches that leverage finer grained analysis of usage data. Qualitative exploration of the issues surrounding the perception of a therapeutic relationship with both the system and the supporter, and the most appropriate forms of feedback would also be valuable. Previous work suggests the impact is spread across several factors that both reduce and enhance engagement [38].

Outcomes

A main focus of the development of the program was to improve client engagement; therefore, the outcome data are encouraging, with a significant reduction in depressive symptomatology following program completion, with a pre-post effect size estimate of $d=1.17$. This concurs with the current body of literature, which does suggest a reasonably robust pre-post outcome effect size for online CBT-based interventions [7]. As Andersson and Cuijpers [9] point out, it is important to distinguish between interventions that are therapist supported and those that are not when estimating efficacy. Pre-post clinical outcome effect sizes for therapist-supported programs have been

estimated at $d=0.61$ and $d=1.0$ in meta-analyses by Andersson and Cuijpers [9] and Spek et al [7], respectively. Thus, the pre-post effect sizes for the present study are consistent with those found with existing comparable interventions.

Recruitment

Recruitment for the study also illustrates the appeal that online interventions may have for some clients. All students are eligible for free face-to-face treatment, but the offer of online treatment led many people with clinically significant symptoms of depression to make contact with the service for the first time. This suggests that therapist-supported online interventions, such as Mind Balance, may attract people who are reluctant to engage with face-to-face treatment; hence, improving access to psychological interventions for those in need of them.

Limitations

The findings of the study are promising, but they should be interpreted with caution. Although the results were statistically significant, the modest sample size motivates continued long-term data gathering. Further replication of the research is also needed. Similarly, the absence of a control group does not allow us to draw specific conclusions with regards the efficacy of the program in relation to spontaneous symptom improvement [39]. As a repeated measures experiment, care should also be taken when comparing to independent group effect sizes [40]. A further controlled study should also examine the maintenance of gains at follow-up, as this was a further limitation of this initial study. Given the service-based nature of the study, clients included those who had contact with the service before commencing the program, which is an additional potential confound. However, in terms of representing a realistic client mix, this is a group that needs to be supported by the service, and for whom online treatment options should be available.

The study did not automatically exclude potential participants who reported symptoms of depression in the severe range. Although this results in a sample that may differ from that in

several comparable studies, it mirrors a service-based pattern of use. The level of engagement with the program exhibited by this group was high, and together with the positive outcome data, suggests that clients along the full span of the symptom severity spectrum may benefit from supported online programs. This service-based study provides an example of how online interventions may be used and fit in with existing mental health services in this setting. The program is now embedded within the day-to-day practice of the service, following the same procedure described here, illustrating the ecological validity of the deployment within the study. Knowledge gained through the study has also been fed into the development of a more detailed treatment manual for supporters to be used in future trials. All data related to time spent on the program, modules viewed, exercises completed, and so on were collected electronically. This ensured a high level of accuracy in determining the level of system usage because there was no reliance on self-reports of time spent on the program.

Conclusion

In this paper, we have presented a study of a new and innovative program for online CBT integrating a range of features intended to increase client engagement. Both client and therapist users were able to use the system successfully, and they made use of a range of program features. The findings suggest that a low level of asynchronous online therapist support is a promising avenue for the development of online interventions, when appropriately integrated into the delivery of the online intervention. As a new mode for delivery of online guided self-help, it contributes to the exploration of the optimum level and nature of therapist input needed to achieve clinical gains in an accessible manner and provides an indication of the many possibilities within the design space for such systems. It strongly motivates exploration of issues of clinical efficacy within a service-based randomized controlled trial and of the potential effectiveness of this form of interactive delivery for other types of intervention.

Acknowledgments

The authors are very grateful for the efforts and support of James Bligh, David Coyle, Karen Tierney, Ella McCabe, Deirdre Flynn, Erica Finn, Derek Richards, and the therapists of the Counseling Service.

Conflicts of Interest

The program and study were supported by the National Digital Research Centre, a government-funded agency that supports translational research and technology transfer from the university sector. Following the study, the platform was developed into a commercial product in which authors JS and GD have an interest. None of the therapists who acted as supporters to the program have a financial interest in the program.

Multimedia Appendix 1

Demonstration of innovative program features.

[[MOV File, 37MB](#) - [jmir_v15i6e121_app1.mov](#)]

References

1. Spek V, Nyklíček I, Cuijpers P, Pop V. Predictors of outcome of group and internet-based cognitive behavior therapy. *J Affect Disord* 2008 Jan;105(1-3):137-145. [doi: [10.1016/j.jad.2007.05.001](#)] [Medline: [17543392](#)]

2. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of Internet-based psychotherapeutic interventions. *Journal of Technology in Human Services* 2008 Jul 03;26(2-4):109-160 [FREE Full text] [doi: [10.1080/15228830802094429](https://doi.org/10.1080/15228830802094429)]
3. Przeworski A, Newman MG. Efficacy and utility of computer-assisted cognitive behavioural therapy for anxiety disorders. *Clinical Psychologist* 2006;10(2):43-53. [doi: [10.1080/13284200500378779](https://doi.org/10.1080/13284200500378779)]
4. Anderson P, Jacobs C, Rothbaum BO. Computer-supported cognitive behavioral treatment of anxiety disorders. *J Clin Psychol* 2004 Mar;60(3):253-267. [doi: [10.1002/jclp.10262](https://doi.org/10.1002/jclp.10262)] [Medline: [14981790](https://pubmed.ncbi.nlm.nih.gov/14981790/)]
5. Newman MG, Szkodny LE, Llera SJ, Przeworski A. A review of technology-assisted self-help and minimal contact therapies for anxiety and depression: is human contact necessary for therapeutic efficacy? *Clin Psychol Rev* 2011 Feb;31(1):89-103. [doi: [10.1016/j.cpr.2010.09.008](https://doi.org/10.1016/j.cpr.2010.09.008)] [Medline: [21130939](https://pubmed.ncbi.nlm.nih.gov/21130939/)]
6. Reger MA, Gahm GA. A meta-analysis of the effects of internet- and computer-based cognitive-behavioral treatments for anxiety. *J Clin Psychol* 2009 Jan;65(1):53-75. [doi: [10.1002/jclp.20536](https://doi.org/10.1002/jclp.20536)] [Medline: [19051274](https://pubmed.ncbi.nlm.nih.gov/19051274/)]
7. Spek V, Cuijpers P, Nyklíček I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. *Psychol Med* 2007 Mar;37(3):319-328. [doi: [10.1017/S0033291706008944](https://doi.org/10.1017/S0033291706008944)] [Medline: [17112400](https://pubmed.ncbi.nlm.nih.gov/17112400/)]
8. Doherty G, Coyle D, Matthews M. Design and evaluation guidelines for mental health technologies. *Interacting with Computers* 2010 Jul 10;22(4):243-252. [doi: [10.1016/j.intcom.2010.02.006](https://doi.org/10.1016/j.intcom.2010.02.006)]
9. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. *Cogn Behav Ther* 2009;38(4):196-205. [doi: [10.1080/16506070903318960](https://doi.org/10.1080/16506070903318960)] [Medline: [20183695](https://pubmed.ncbi.nlm.nih.gov/20183695/)]
10. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. *PLoS One* 2010;5(10):e13196 [FREE Full text] [doi: [10.1371/journal.pone.0013196](https://doi.org/10.1371/journal.pone.0013196)] [Medline: [20967242](https://pubmed.ncbi.nlm.nih.gov/20967242/)]
11. Griffiths KM, Christensen H. Review of randomised controlled trials of internet interventions for mental disorders and related conditions. *Clinical Psychologist* 2006;10(1):16-29. [doi: [10.1080/13284200500378696](https://doi.org/10.1080/13284200500378696)]
12. Cavanagh K, Shapiro DA, Van Den Berg S, Swain S, Barkham M, Proudfoot J. The effectiveness of computerized cognitive behavioural therapy in routine care. *Br J Clin Psychol* 2006 Nov;45(Pt 4):499-514. [doi: [10.1348/014466505X84782](https://doi.org/10.1348/014466505X84782)] [Medline: [17076960](https://pubmed.ncbi.nlm.nih.gov/17076960/)]
13. Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F. Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. *Cogn Behav Ther* 2011;40(4):251-266. [doi: [10.1080/16506073.2011.616531](https://doi.org/10.1080/16506073.2011.616531)] [Medline: [22060248](https://pubmed.ncbi.nlm.nih.gov/22060248/)]
14. Titov N, Andrews G, Davies M, McIntyre K, Robinson E, Solley K. Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. *PLoS One* 2010;5(6):e10939 [FREE Full text] [doi: [10.1371/journal.pone.0010939](https://doi.org/10.1371/journal.pone.0010939)] [Medline: [20544030](https://pubmed.ncbi.nlm.nih.gov/20544030/)]
15. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. *Clin Psychol Rev* 2012 Jun;32(4):329-342. [doi: [10.1016/j.cpr.2012.02.004](https://doi.org/10.1016/j.cpr.2012.02.004)] [Medline: [22466510](https://pubmed.ncbi.nlm.nih.gov/22466510/)]
16. Andersson G, Paxling B, Wiwe M, Vernmark K, Felix CB, Lundborg L, et al. Therapeutic alliance in guided internet-delivered cognitive behavioural treatment of depression, generalized anxiety disorder and social anxiety disorder. *Behav Res Ther* 2012 Sep;50(9):544-550. [doi: [10.1016/j.brat.2012.05.003](https://doi.org/10.1016/j.brat.2012.05.003)] [Medline: [22728647](https://pubmed.ncbi.nlm.nih.gov/22728647/)]
17. Cavanagh K, Shapiro DA, Van Den Berg S, Swain S, Barkham M, Proudfoot J. The acceptability of computer-aided cognitive behavioural therapy: a pragmatic study. *Cogn Behav Ther* 2009;38(4):235-246. [doi: [10.1080/16506070802561256](https://doi.org/10.1080/16506070802561256)] [Medline: [19306147](https://pubmed.ncbi.nlm.nih.gov/19306147/)]
18. Titov N, Andrews G, Kemp A, Robinson E. Characteristics of adults with anxiety or depression treated at an internet clinic: comparison with a national survey and an outpatient clinic. *PLoS One* 2010;5(5):e10885 [FREE Full text] [doi: [10.1371/journal.pone.0010885](https://doi.org/10.1371/journal.pone.0010885)] [Medline: [20526371](https://pubmed.ncbi.nlm.nih.gov/20526371/)]
19. Brawley LR, Culos-Reed SN. Studying adherence to therapeutic regimens: overview, theories, recommendations. *Control Clin Trials* 2000 Oct;21(5 Suppl):156S-163S. [Medline: [11018570](https://pubmed.ncbi.nlm.nih.gov/11018570/)]
20. Melville KM, Casey LM, Kavanagh DJ. Dropout from Internet-based treatment for psychological disorders. *Br J Clin Psychol* 2010 Nov;49(Pt 4):455-471. [doi: [10.1348/014466509X472138](https://doi.org/10.1348/014466509X472138)] [Medline: [19799804](https://pubmed.ncbi.nlm.nih.gov/19799804/)]
21. Eysenbach G. The law of attrition. *J Med Internet Res* 2005;7(1):e11 [FREE Full text] [doi: [10.2196/jmir.7.1.e11](https://doi.org/10.2196/jmir.7.1.e11)] [Medline: [15829473](https://pubmed.ncbi.nlm.nih.gov/15829473/)]
22. Christensen H, Griffiths K, Farrer L. Adherence in internet interventions for anxiety and depression. *J Med Internet Res* 2009;11(2):e13 [FREE Full text] [doi: [10.2196/jmir.1194](https://doi.org/10.2196/jmir.1194)] [Medline: [19403466](https://pubmed.ncbi.nlm.nih.gov/19403466/)]
23. Nicholas J, Proudfoot J, Parker G, Gillis I, Burckhardt R, Manicavasagar V, et al. The ins and outs of an online bipolar education program: a study of program attrition. *J Med Internet Res* 2010;12(5):e57 [FREE Full text] [doi: [10.2196/jmir.1450](https://doi.org/10.2196/jmir.1450)] [Medline: [21169169](https://pubmed.ncbi.nlm.nih.gov/21169169/)]
24. Neil AL, Batterham P, Christensen H, Bennett K, Griffiths KM. Predictors of adherence by adolescents to a cognitive behavior therapy website in school and community-based settings. *J Med Internet Res* 2009;11(1):e6 [FREE Full text] [doi: [10.2196/jmir.1050](https://doi.org/10.2196/jmir.1050)] [Medline: [19275982](https://pubmed.ncbi.nlm.nih.gov/19275982/)]

25. de Graaf LE, Hollon SD, Huibers MJ. Predicting outcome in computerized cognitive behavioral therapy for depression in primary care: A randomized trial. *J Consult Clin Psychol* 2010 Apr;78(2):184-189. [doi: [10.1037/a0018324](https://doi.org/10.1037/a0018324)] [Medline: [20350029](https://pubmed.ncbi.nlm.nih.gov/20350029/)]
26. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. *Psychol Med* 2009 May;39(5):705-712. [doi: [10.1017/S0033291708004224](https://doi.org/10.1017/S0033291708004224)] [Medline: [18812006](https://pubmed.ncbi.nlm.nih.gov/18812006/)]
27. Beck AT, Steer RA, Ball R, Ranieri W. Comparison of Beck Depression Inventories -IA and -II in psychiatric outpatients. *J Pers Assess* 1996 Dec;67(3):588-597. [doi: [10.1207/s15327752jpa6703_13](https://doi.org/10.1207/s15327752jpa6703_13)] [Medline: [8991972](https://pubmed.ncbi.nlm.nih.gov/8991972/)]
28. Beck AT, Steer RA, Brown GK. Beck Depression Inventory Manual, 2nd edition. Boston, MA: Harcourt Brace; 1996.
29. Dozois DJA, Dobson KS, Ahnberg JL. A psychometric evaluation of the Beck Depression Inventory-II. *Psychological Assessment* 1998;10(2):83-89. [doi: [10.1037/1040-3590.10.2.83](https://doi.org/10.1037/1040-3590.10.2.83)]
30. Kapci EG, Uslu R, Turkcapar H, Karaoglan A. Beck Depression Inventory II: evaluation of the psychometric properties and cut-off points in a Turkish adult population. *Depress Anxiety* 2008;25(10):E104-E110. [doi: [10.1002/da.20371](https://doi.org/10.1002/da.20371)] [Medline: [17876817](https://pubmed.ncbi.nlm.nih.gov/17876817/)]
31. Steer RA, Clark DA. Psychometric characteristics of the Beck Depression Inventory-II, with college students. *Measurement and Evaluation in Counseling and Development* 1997;10:128-136.
32. Richards D, Timulak L, Hevey D. A comparison of two online cognitive-behavioural interventions for symptoms of depression in a student population: The role of therapist responsiveness. *Counselling and Psychotherapy Research* 2012 Oct 12:1-10. [doi: [10.1080/14733145.2012.733715](https://doi.org/10.1080/14733145.2012.733715)]
33. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Erlbaum; 1988.
34. 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 Nov;14(6):e152 [FREE Full text] [doi: [10.2196/jmir.2104](https://doi.org/10.2196/jmir.2104)] [Medline: [23151820](https://pubmed.ncbi.nlm.nih.gov/23151820/)]
35. Marks IM, Mataix-Cols D, Kenwright M, Cameron R, Hirsch S, Gega L. Pragmatic evaluation of computer-aided self-help for anxiety and depression. *Br J Psychiatry* 2003 Jul;183:57-65. [Medline: [12835245](https://pubmed.ncbi.nlm.nih.gov/12835245/)]
36. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. *Ann Behav Med* 2009 Aug;38(1):18-27 [FREE Full text] [doi: [10.1007/s12160-009-9133-4](https://doi.org/10.1007/s12160-009-9133-4)] [Medline: [19802647](https://pubmed.ncbi.nlm.nih.gov/19802647/)]
37. 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](https://doi.org/10.2196/jmir.1602)] [Medline: [21393123](https://pubmed.ncbi.nlm.nih.gov/21393123/)]
38. Donkin L, Glozier N. Motivators and motivations to persist with online psychological interventions: a qualitative study of treatment completers. *J Med Internet Res* 2012 Jun;14(3):e91 [FREE Full text] [doi: [10.2196/jmir.2100](https://doi.org/10.2196/jmir.2100)] [Medline: [22743581](https://pubmed.ncbi.nlm.nih.gov/22743581/)]
39. Harris AD, McGregor JC, Perencevich EN, Furuno JP, Zhu J, Peterson DE, et al. The use and interpretation of quasi-experimental studies in medical informatics. *J Am Med Inform Assoc* 2006 Feb;13(1):16-23. [doi: [10.1197/jamia.M1749](https://doi.org/10.1197/jamia.M1749)] [Medline: [16221933](https://pubmed.ncbi.nlm.nih.gov/16221933/)]
40. Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychol Methods* 2002 Mar;7(1):105-125. [Medline: [11928886](https://pubmed.ncbi.nlm.nih.gov/11928886/)]

Abbreviations

BDI-II: Beck Depression Inventory-II
CBT: cognitive behavioral therapy
SMS: short message service

Edited by G Eysenbach; submitted 04.07.12; peer-reviewed by H Riper, P Lam, C Warren, J Wright, R Anderson, C Donovan; comments to author 25.09.12; revised version received 19.11.12; accepted 23.04.13; published 27.06.13.

Please cite as:

Sharry J, Davidson R, McLoughlin O, Doherty G
A Service-Based Evaluation of a Therapist-Supported Online Cognitive Behavioral Therapy Program for Depression
J Med Internet Res 2013;15(6):e121
URL: <http://www.jmir.org/2013/6/e121/>
doi: [10.2196/jmir.2248](https://doi.org/10.2196/jmir.2248)
PMID: [23807565](https://pubmed.ncbi.nlm.nih.gov/23807565/)

License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Development of an Obesity Management Ontology Based on the Nursing Process for the Mobile-Device Domain

Hyun-Young Kim¹, PhD, RN; Hyeoun-Ae Park^{2,3}, PhD, RN, FAAN; Yul Ha Min^{2,3}, MS, RN; Eunjoo Jeon², MS, RN

¹College of Nursing, Eulji University, Daejeon, Republic of Korea

²College of Nursing, Seoul National University, Seoul, Republic of Korea

³Research Institute of Nursing Science, Seoul National University, Seoul, Republic of Korea

Corresponding Author:

Hyeoun-Ae Park, PhD, RN, FAAN

College of Nursing

Seoul National University

103 Daehak-ro

Jongno-gu

Seoul, 110-799

Republic of Korea

Phone: 82 2 7408827

Fax: 82 2 7654103

Email: hapark@snu.ac.kr

Abstract

Background: Lifestyle modification is the most important factor in the management of obesity. It is therefore essential to enhance client participation in voluntary and continuous weight control.

Objective: The aim of this study was to develop an obesity management ontology for application in the mobile-device domain. We considered the concepts of client participation in behavioral modification for obesity management and focused on minimizing the amount of information exchange between the application and the database when providing tailored interventions.

Methods: An obesity management ontology was developed in seven phases: (1) defining the scope of obesity management, (2) selecting a foundational ontology, (3) extracting the concepts, (4) assigning relationships between these concepts, (5) evaluating representative layers of ontology content, (6) representing the ontology formally with Protégé, and (7) developing a prototype application for obesity management.

Results: Behavioral interventions, dietary advice, and physical activity were proposed as obesity management strategies. The nursing process was selected as a foundation of ontology, representing the obesity management process. We extracted 127 concepts, which included assessment data (eg, sex, body mass index, and waist circumference), inferred data to represent nursing diagnoses and evaluations (eg, degree of and reason for obesity, and success or failure of lifestyle modifications), and implementation (eg, education and advice). The relationship linking concepts were “part of”, “instance of”, “derives of”, “derives into”, “has plan”, “followed by”, and “has intention”. The concepts and relationships were formally represented using Protégé. The evaluation score of the obesity management ontology was 4.5 out of 5. An Android-based obesity management application comprising both agent and client parts was developed.

Conclusions: We have developed an ontology for representing obesity management with the nursing process as a foundation of ontology.

(*J Med Internet Res* 2013;15(6):e130) doi:[10.2196/jmir.2512](https://doi.org/10.2196/jmir.2512)

KEYWORDS

obesity management; ontology; nursing process

Introduction

Overview

Advancements in socioeconomic status and lifestyle changes in Korea have led to an increasing number of obese people and a consequent increase in the importance of obesity management to prevent illness and promote health among the population [1]. There has been an emphasis on the development and implementation of various programs to help clients who wish to manage obesity effectively. Since lifestyle modification is the most important aspect of obesity management, such programs must be able to enhance the client's participation in voluntary and continuous weight control [2-4].

This study investigated two strategies for empowering clients to effectively manage their obesity: (1) use of mobile devices, which are now an integral part of everyday life in many countries [5], and (2) use of evidence-based knowledge. It is important to provide a nursing service whenever a client requires it, without time or space limitations, and especially for a problem such as obesity that requires continuous self-monitoring. Previous research has demonstrated that mobile devices such as smartphones are very effective tools for self-monitoring in obesity management [6,7]. In addition, mobile devices allow users to input and communicate data in real time, regardless of their physical location [8].

Some clients lack knowledge regarding how to manage obesity, and the information provided through mobile devices should be based on the available evidence. The best sources for evidence-based knowledge are published clinical practice guidelines. The content for obesity management should vary with the health behavior and diet patterns of the individual client. The effectiveness of obesity management can be improved if time-stamped behavioral data are collected and tailored coaching is provided remotely, based on published clinical practice guidelines, using mobile devices [2,9].

Since these guidelines are written in natural language, it is necessary to translate them into a computer-interpretable format. Ontology is a computer-interpretable knowledge model for formalizing and representing shared concepts in a specific domain of interest. It is considered to be a highly effective way of improving the integration, interoperability, and sharing of data. Moreover, the enhanced reusability of clinical data and evidence-based knowledge support clinical decision making by explicitly defining and delivering semantic concepts in a specific domain [10-12].

Background

This project is a part of the Health Avatar Project in Korea, which promotes health and manages health problems using personal data and knowledge in virtual space. The health avatar consists of an agent avatar, which is a representation of the expert knowledge or published knowledge (such as clinical practice guidelines), and a client avatar, which is a representation of personal data from molecular to community levels (Figure 1). The health avatar collects data from the client avatar and provides it with tailored care solutions based on the personal data and knowledge or evidence provided by the agent avatar.

The client avatar collects and sends personal data to the agent avatar. Upon receiving that personal data, the agent avatar makes judgments and provides the client avatar with tailored recommendations.

The obesity management application is a type of health avatar. The client avatar can be represented by data such as weight, height, abdominal circumference, physical activity, and diet, while the agent avatar can be represented by the clinical practice guidelines published by the National Institute for Health and Clinical Excellence (NICE). Personal data such as weight and height, and tailored recommendations, alarms, or reminders such as "educate about low-calorie diet" are transferred between the client and agent avatars. The front end serves as a user interface to collect these personal data and display tailored recommendations suggested by the agent avatar.

Objectives

This study considered two objectives during the development of the obesity management ontology. The first objective was to identify concepts related to client participation in behavioral modification for obesity management, since it is important to identify the factors that motivate and encourage participation in the process. The possibility of using the nursing process as a foundation of ontology was explored, since that process is a patient-focused clinical reasoning method for nursing problems, and each phase of the nursing process emphasizes the active participation of the client [13]. The second objective was to minimize information exchange between the application and the database when providing tailored interventions. The obesity management application program developed in this study operates on a platform that requires real-time, two-way data communication with the database, storing an enormous volume of individual life-log data. In addition, this application program could ultimately share the database and platform with other health care management services, such as those for diabetes and dyslipidemia. Thus, it is important to distinguish between the types of data used in different phases of the service to minimize data communication between the database and the application.

With this background, a clinical practice guideline-based obesity management ontology was developed by identifying the concepts and relationship between concepts according to the nursing process, using ontology development methods that are used in the biomedical arena. In addition, concepts describing client participation in the application program, which is the most important aspect of obesity management, were identified and reflected in the ontology, and finally, the ontology was evaluated.

Methods

Overview

An obesity management ontology was developed using the General Formal Ontology method [14], which comprises seven phases: (1) defining the scope of obesity management, (2) selecting a foundational ontology, (3) extracting the concepts, (4) assigning relationships among these concepts, (5) evaluating representative layers of ontology content, (6) representing the ontology formally using Protégé (an open-source ontology editor

and knowledge-based framework), and (7) developing a prototype Android application for obesity management.

Defining the Scope for Obesity Management

To determine the scope of obesity management services for adult clients, a guideline developed by NICE [15] was referenced as an evidence-based clinical practice guideline. This guideline was chosen because there is a version for patients/clients. Five experts (1 physician who cares for obese patients, 2 nurses who have participated in obesity-related research for over 2 years, and 2 informatics nurses with more than 5 years of work experience) participated in this process. Intervention strategies were selected from the guideline using a consensus method.

Selecting a Foundational Ontology for Obesity Management

We reviewed top-level ontologies such as Basic Formal Ontology and a theoretical framework to determine the most appropriate foundational ontology for obesity management, thus minimizing errors during the creation of a domain-specific ontology [10,16,17]. The first criterion used for selecting a foundational ontology was whether it meets two of the purposes of an obesity management ontology: (1) enhancing client participation and (2) minimizing real-time data traffic when providing client-specific interventions. The second criterion used was whether it can be used as a basic framework when translating text-based guidelines into computer-interpretable knowledge [10,12]. An initial ontology was constructed according to the information types used during the care process for obesity management. This initial ontology was elaborated using concrete concepts extracted in the next phase.

Extracting Concepts From the Guidelines on Obesity Management

The concepts needed to assess the client, make a diagnosis, determine client-specific interventions, and evaluate the outcome of those interventions within the scope of obesity management were extracted. These concepts included those used to represent not only personal information but also client-specific interventions from the guidelines based on personal data. This phase considered the concepts needed to enhance the client participation, together with ways to operationalize those concepts.

Assigning Relationships Among These Concepts

The extracted concepts were specified as classes, individuals, and relationships between classes and individuals in order to define the obesity management ontology according to the foundational ontology determined in the second phase [18]. The “OBO relationship ontology of the Open Biological and Biomedical Ontologies” (OBO) [19,20] and the linkage concept hierarchy of SNOMED CT (Systematized Nomenclature of Medicine Clinical Terms) were referred to when defining the relationships.

Evaluation of Representative Layers of Ontology Content for Obesity Management

The content of the obesity management ontology was evaluated using scenarios via peer-review by 3 ontology and domain experts [12,21,22]. This was a necessary modeling process to evaluate the quality and consistency of the ontology [23]. We evaluated whether the selected foundational ontology was sufficiently valid to represent the obesity management ontology. In addition, we determined whether the obesity management ontology developed with concepts extracted from clinical practice guideline was representative of the concepts identified from real clinical cases. Representation/semantic-layer evaluation criteria were used as one of the basic internal dimension evaluations. Nine-item Likert scales with 5-point scores (from “strongly agree” to “strongly disagree”) were used as evaluation tools. The nine items were as follows: match between formal and cognitive semantics, consistency (no formal contradictions), clarity (context and background knowledge), explicitness (understanding the concept semantics), interpretability (formal, informal, logical specification, and documentation), accuracy (fit between ontology and corpus terms), comprehensiveness (extent of the target domain covered), granularity (fine-grained coverage vs loose coverage), and relevance (for users) [22]. Three experts with experience in ontology development participated in the evaluation, with five case studies on obesity management published in Korean used as cases for evaluation [24,25].

Formal Representation of the Ontology With Protégé

The extracted concepts and relationships were formalized using Protégé. The feasibility of the representation of the nursing-process ontology based on Protégé was assessed.

Development of a Prototype Application for Obesity Management

A prototype application was developed based on an obesity management ontology incorporating the two key features of client participation and minimizing information exchange between the application and the database.

Results

The Scope for Obesity Management

The clinical practice guidelines on obesity reported by NICE (2006) [15] include lifestyle interventions, behavioral interventions, physical activity, dietary advice, pharmacological interventions, and surgical interventions as strategies for the management of overweight and obesity. In this study, lifestyle interventions, behavioral interventions, physical activity, and dietary advice were proposed as obesity management strategies. With respect to behavioral interventions, keeping a diary of physical activity and diet, and cognitive restructuring such as realistic goal setting were included according to the scope and sustainability of clients' behavioral change. Pharmacological and surgical interventions were excluded because these strategies cannot be self-managed by clients, as they require the intervention of a health care provider.

A Foundational Ontology: the Nursing Process

The nursing process met the requirements of promoting client participation and minimizing data transactions between databases and the obesity management application in this study. The nursing process comprises five cyclical phases: assessment, nursing diagnosis, outcome identification, implementation, and evaluation. The nursing process as a framework for nursing practice is a systematic, patient-centered, and goal-oriented method with the following phases: (1) identifying client status and need, (2) making nursing diagnoses based on the available client data, (3) determining the expected outcome for a nursing problem for the client, (4) determining and implementing the best nursing interventions to reach the goal, and (5) evaluating whether the expected outcome has been reached. This framework was selected as a foundation of ontology because it describes well both the obesity management process and the client participation therein. In obesity management, the expected outcome is determined by the clients, the evaluation is performed repeatedly to determine whether the outcome has been reached, and the tailored best intervention is implemented to reach expected outcomes based on the client's assessment data.

Figure 2 illustrates the nursing process used for obesity management as a foundation of ontology. The nursing process can be subdivided into five phases according to the time taken to access the application when specifying the obesity management ontology. This was represented as a dynamic ontology with a spiral reflecting two cyclical nursing processes that are connected to each other. The blue color in the figure (ie, boxes 1, 2, 3, and 4) indicates the initial process of the obesity management service from the beginning to the end, while the green color (ie, boxes 5 and 6) indicates the repetitive process of obesity management from the beginning of the revisit to the service until a client reaches his or her identified outcome. These two cyclical nursing processes contribute to minimizing the amount of data transactions. The nursing diagnosis inferred using the personal data and evidence-based knowledge in a single, long-interval process can be reused in short-interval periodic processes; therefore, data needed in short-interval, periodic processes would be minimized.

Figure 3 shows the obesity management information, which in part reflects the cyclical characteristics of the service provision and login information, with client information guiding that cyclic process. This initial nursing process ontology was elaborated based on these information characteristics. The model has three levels of category: temporal, nursing process, and analytical. The temporal level was an abstract category

corresponding to the obesity management. This was divided into assessment, nursing diagnosis, goal, outcome evaluation, and implementation phases, according to the nursing process. The analytical category shows the data linkage to each phase of the nursing process. In the assessment phase, data are collected from a client, and the nursing diagnosis phase defines the degree of obesity using the data collected in the assessment phase. A target weight loss and the date by which to achieve that goal are set in the goal phase, while the intervention for obesity management is provided in the implementation phase. Finally, the state of the client after the intervention is defined in the outcome evaluation phase by comparing the goal with the reassessment data.

Concepts Extracted From the Guidelines on Obesity Management

The concepts required to make nursing diagnoses, generate client-specific interventions to enable the individual to reach the goal that they have set, and evaluate outcomes in the obesity management process were extracted. In total, 127 concepts were identified ([Multimedia Appendix 1](#)). The extracted concepts included data collected in the assessment, the degree of obesity, interventions, and outcomes categories. Examples include personal information such as sex, height, weight, waist circumference, intention to lose weight, and risk factors that the client entered when they accessed the application for the first time, and the calorie gap calculated using the diet and physical activity monitoring data. Other examples are a nursing diagnosis such as “obese due to lack of physical activity” made by the application using data entered by the client, and an intervention such as “educate about physical activity” recommended by the application based on the nursing diagnosis.

Relationships Between Classes and Individuals and Relationships Based on These Concepts

The 127 concepts extracted in the previous phase were arranged as classes and individuals depending on the foundational ontology, using the relationship-linking classes listed in [Table 1](#).

Results of Evaluation of the Obesity Management Ontology

The evaluation scores of the nursing-process-based obesity management ontology for various criteria are presented in [Table 2](#). All nine items scored above 4, with interpretability and consistency items scoring 5 points. The category with the lowest grade was the granularity criterion. Ontology and domain experts recommended that the concepts representing activity level and diet should be defined in more detail.

Table 1. Relationship-linking classes in the obesity management ontology.

Relationship name	Defined at concept
Part of	Relationship between class and subclasses
Instance of	Relationship between instances and class
Derives from	Relationships of nursing diagnosis and evaluation to assessment data
Derives into	Relationships of assessment data to nursing diagnosis and evaluation
Has plan	Relationships of nursing diagnosis and evaluation to implementation
Followed by	Relationships of implementation to nursing diagnosis and evaluation
Intention	Relationship of nursing diagnosis to goal identification

Table 2. Grades of evaluation on the representation layer.

Criterion	Score		
	Mean	Minimum	Maximum
Match between formal and cognitive semantics	4.7	4	5
Clarity	4.0	4	5
Explicitness	4.3	4	5
Interpretability	5.0	4	5
Accuracy	4.7	4	5
Consistency	5.0	5	5
Comprehensiveness	4.3	4	5
Granularity	4.1	4	4
Relevance	4.8	4	5
Total	4.5		

Artifact Using Protégé

Figure 4 shows some of the relationships among the classes represented by obesity management using Protégé. The rectangles represent classes and the arrows represent the relationship between the 127 concepts. The color of the arrow reflects the relationship among the classes and individuals. For example, red arrows (n=1) represent the relationship between each individual and class, which is “instance of”; green arrows (n=2) mean the relationship of “part of”; purple arrows (n=3) represent “followed by”, which is the relationship between implementation and evaluation in this figure; and gray arrows (n=4) represent “derives into”, which is the relationship between the assessment data and evaluation.

A Prototype Application for Obesity Management

The prototype application consisted of two parts: an agent avatar and a client avatar as described in the Figure 1. The agent avatar

as a knowledge base was developed using an XML schema made from the obesity management ontology including 37 classes and 90 individuals in Protégé. The XML schema was translated into database schema with 11 tables for assessment, diagnosis, and outcome identification. In addition, classes such as diagnosis, implementation, and evaluation in XML were converted into Java classes with methods to call recommendations according to data entered, which were linked to database. The client avatar as a user interface consisted of 41 screens to collect client’s health data and deliver recommendations provided from the agent avatar. With JDK (Java development kit) 1.7.0, Eclipse having ADT (Android Development Tool) Plugin for Eclipse, and the Android SDK platform 4.0.3, the client avatar and the agent avatar were integrated. Figure 5 is a screen shot providing one of tailored recommendations.

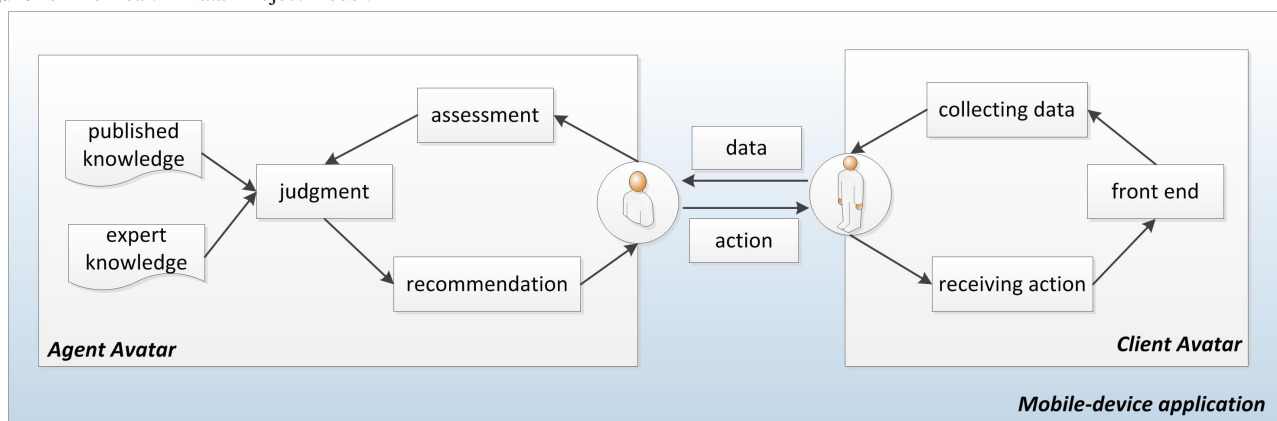
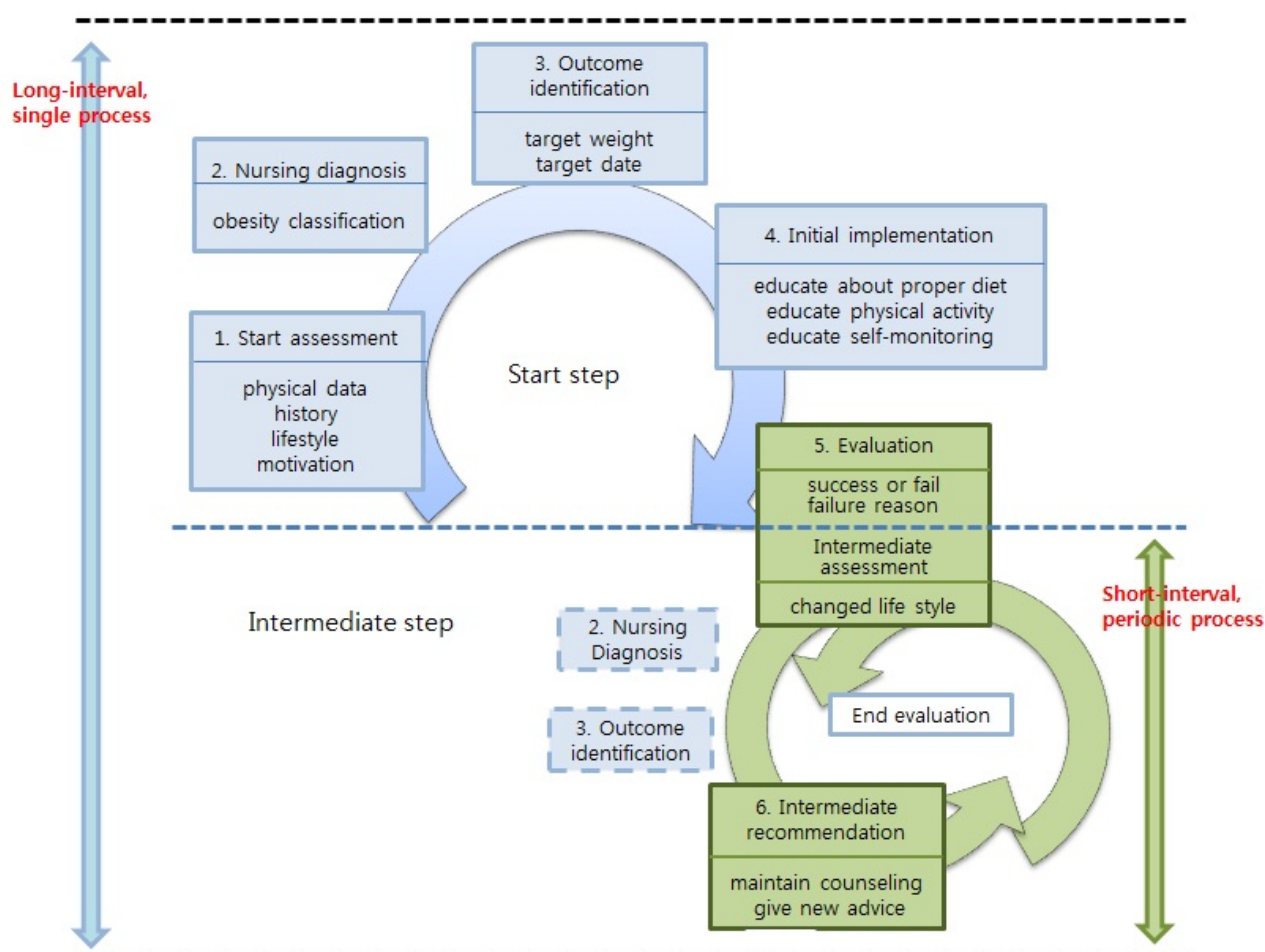
Figure 1. The Health Avatar Project model.**Figure 2.** Cyclical process for obesity management.

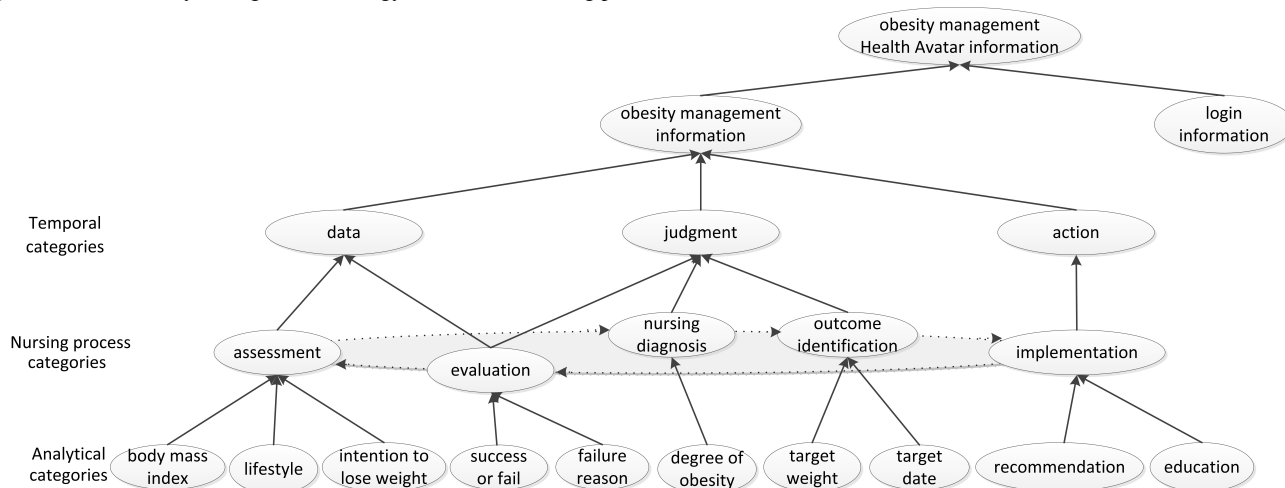
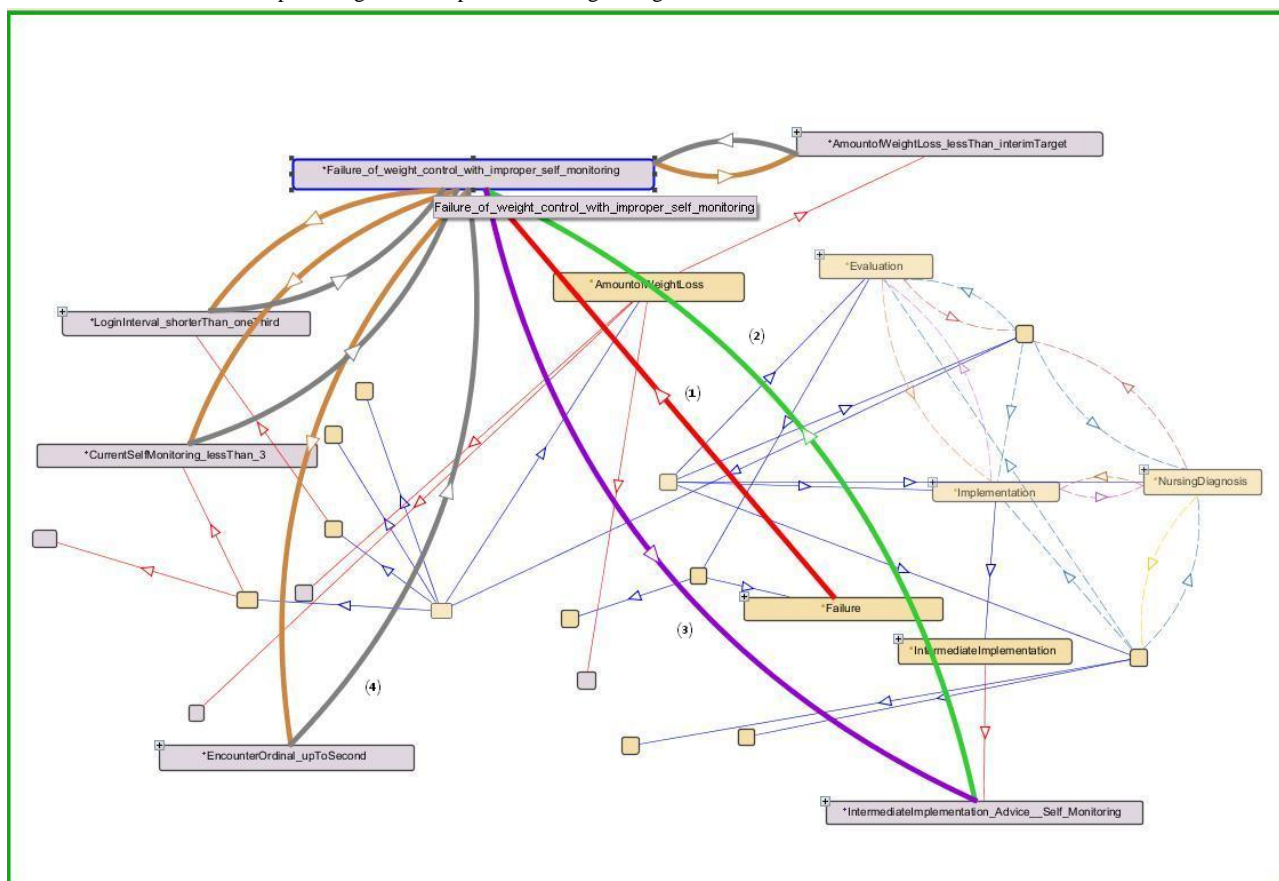
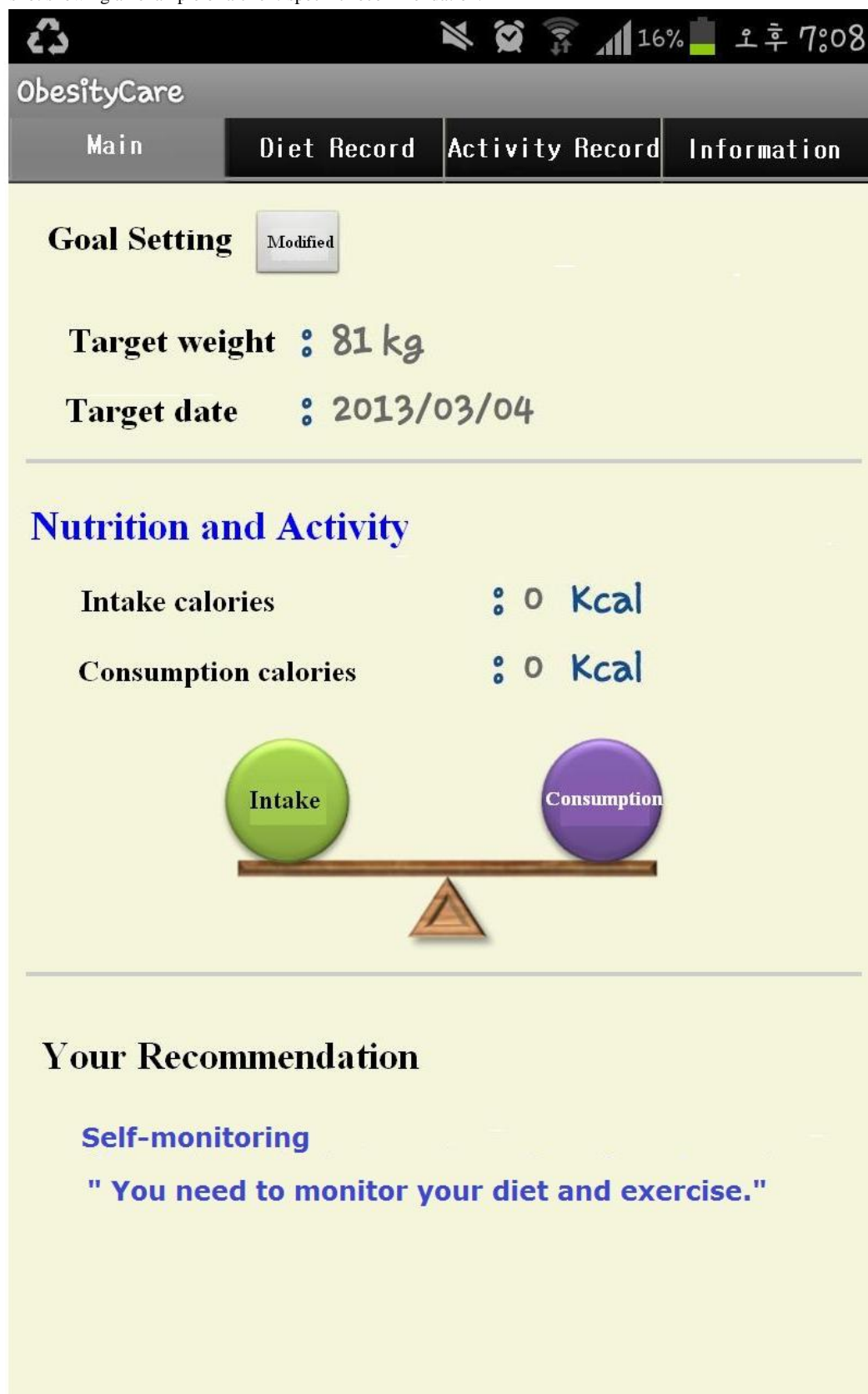
Figure 3. Initial obesity management ontology based on the nursing process.**Figure 4.** Some of the relationships among classes represented using Protégé.

Figure 5. Screenshot showing an example of a client-specific recommendation.

Discussion

Principal Findings and Conclusions

Ontology is an effective way of representing specific domain knowledge in the field of biomedicine. It is also used for knowledge model development to support the clinical decision making of health care providers [18,26,27]. Most of the published ontologies have been developed to aid health care providers in clinical decision making [28,29]. However, the ontology developed in this study was designed to be used with a mobile-device-based application to provide tailored interventions to clients who want to manage their obesity by themselves.

The nursing process was used as a foundation of ontology to describe the client participation in obesity management. Since the nursing process is a scientific and systematic problem-solving framework that allows the client to participate in their own care [13], it was considered a valid representation of the client participation in obesity management. The findings of this study confirm that belief.

The client was considered to be the principal agent of goal setting and outcome evaluation in obesity management. The nursing-process-based obesity management ontology developed in this study allows the client to participate in setting a goal, such as a target weight loss and duration to reach to the target weight, and to monitor and evaluate their behavioral modification toward obesity management. To this end, a set of concepts was introduced to describe whether the goal and rate of weight loss were realistic and adequate, and another set was designed to describe whether the obesity management behavior was adequate based on the self-monitoring data.

One of the success factors in the existing Web-based intervention programs is login intervals by clients reflecting client participation [3]. The ontology developed in this study reflects this by introducing a concept to describe the continuity of the obesity management process based on the interval between two logins. If a client does not use the application within one-third of the target interval, which is the criterion used to judge the continuity of the care process, it was considered that the continuity of the obesity management process was deficient. In that case, the client must restart the long-interval, single process from the assessment point and will have to set a new goal. In order to promote the continuity of the obesity management process, reminder and alarm functions were implemented.

Another key characteristic of the nursing process is that it comprises cyclical phases. In this study, two cyclical nursing processes were connected to each other based on the sequences of logins to the application in our obesity management ontology: the first and second cycles of the nursing process correspond to the first visit and revisits, respectively. At the first visit, the degree of obesity is inferred based on height, weight, abdominal circumference, sex, and past history, and a goal is set. From the second visit, obesity management is evaluated by comparing current weight with the defined outcome. Based on the login information, which is used to guide the cyclical process of

obesity management, if the client is either a first-time user of the application or a previous user who did not meet the continuity criteria, a long-interval, single process will start; otherwise, a short-interval, periodic process starts. This feature minimizes the real-time data traffic.

The obesity management ontology developed in this study was compatible with the Initial Clinical Information Ontology of the Danish General Electronic Patient Journal Conceptual Model [17]. This Danish model was introduced to clearly explain the clinical information created during the general care process using ontology. Even though the nursing process model used in this study as one of the care-process models was developed specifically to describe an obesity management process, it is compatible with the Danish model because the nursing process can be regarded as a type of general care process. However, some of the processes differ between the model developed in this study and the Danish model. The Danish model categorized the process into two different groups, ie, observation and intervention, whereas the model developed in this study included five different phases of the care process: assessment, nursing diagnosis, outcome identification, implementation, and evaluation.

Concepts extracted based on the nursing process are specified as classes and individuals. The relationships between these concepts were extracted from the OBO [19,20] and the linkage concept hierarchy of SNOMED CT. No new relationships were defined in this study. Thus, the ontology developed in the present study based on the nursing process has proved to be sufficiently universal to be integrated with others, since all of these concepts describing relationships are from the OBO and SNOMED CT linkage concept hierarchy. This is important because this study is a part of the Health Avatar Project. The design of this mobile-device-based ontology was based on three major criteria: practical relevance, incorporation of existing standards and ontologies, and extensibility towards future relevant domains [30].

The obesity management ontology developed herein was evaluated at a representation layer of the basic internal layers from various dimensions of the ontology [22]. The representation layer was evaluated for ontology semantics, with nine items. It was found that the concepts and the relationships between concepts used for obesity management were well represented, with a score of 4.5 out of 5. Furthermore, the nursing process was found to be adequate as a foundational ontology for obesity management ontology development. A nursing process with five distinctive phases with unique characteristics may have helped to improve the consistency of the ontology.

Limitations

This study was subject to the following limitations. First, the developed ontology cannot be used for other types of nursing problems because the concepts used in its development were extracted specifically from obesity management guidelines. We recommend further development of this ontology for other types of nursing problems to determine whether the nursing process can be used as a foundation of ontology to represent nursing care in general. Second, the effectiveness of the prototype

Android application developed in this study was not tested. The study in the near future. effectiveness of the application will be established in another

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korean government (MSIP) (Nos. 2012-012257 and 2010-0028631).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Some of the concepts extracted from the guidelines.

[[PDF File \(Adobe PDF File\), 66KB - jmir_v15i6e130_app1.pdf](#)]

References

1. Bae N, Kwon I, Cho Y. Ten year change of body mass index in Korean: 1997-2007. *Korean J Obes* 2009;18(1):24-30.
2. Cussler EC, Teixeira PJ, Going SB, Houtkooper LB, Metcalfe LL, Blew RM, et al. Maintenance of weight loss in overweight middle-aged women through the Internet. *Obesity* 2008;16:1502-1060. [doi: [10.1038/oby.2008.19](#)] [Medline: [18309301](#)]
3. 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;11(4):306-321. [doi: [10.1111/j.1467-789X.2009.00646.x](#)] [Medline: [19754633](#)]
4. Watson A, Bickmore T, Cange A, Kulshreshtha A, Kvedar J. An internet-based virtual coach to promote physical activity adherence in overweight adults: randomized controlled trial. *J Med Internet Res* 2012;26(14(1)):e1 [FREE Full text] [doi: [10.2196/jmir.1629](#)] [Medline: [22281837](#)]
5. Boulos MN, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. *Biomed Eng Online* 2011;10(24):1-14 [FREE Full text] [doi: [10.1186/1475-925X-10-24](#)] [Medline: [21466669](#)]
6. Patrick K, Griswold WG, Raab F, Intille SS. Health and the mobile phone. *Am J Prev Med* 2008;35(2):177-181 [FREE Full text] [doi: [10.1016/j.amepre.2008.05.001](#)] [Medline: [18550322](#)]
7. Mosa AS, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. *BMC Med Inform Decis Mak* 2012;12:67 [FREE Full text] [doi: [10.1186/1472-6947-12-67](#)] [Medline: [22781312](#)]
8. Burdette SD, Herchline TE, Oehler R. Surfing the web: practicing medicine in a technological age: using smartphones in clinical practice. *Clin Infect Dis* 2008;47(1):117-122 [FREE Full text] [doi: [10.1086/588788](#)] [Medline: [18491969](#)]
9. Spring B, Schneider K, McFadden HG, Vaughn J, Kozak AT, Smith M, et al. Multiple behavior changes in diet and activity: a randomized controlled trial using mobile technology. *Arch Intern Med* 2012;172(10):789-796. [doi: [10.1001/archinternmed.2012.1044](#)] [Medline: [22636824](#)]
10. Burek P, Hoehndorf R, Loebe F, Visagie J, Herre H, Kelso J. A top-level ontology of functions and its application in the Open Biomedical Ontologies. *Bioinformatics* 2006;22(14):e66-e73 [FREE Full text] [doi: [10.1093/bioinformatics/btl266](#)] [Medline: [16873523](#)]
11. Aranguren ME, Bechhofer S, Lord P, Sattler U, Stevens R. Understanding and using the meaning of statements in a bio-ontology: recasting the Gene Ontology in OWL. *BMC Bioinformatics* 2007;8:57 [FREE Full text] [doi: [10.1186/1471-2105-8-57](#)] [Medline: [17311682](#)]
12. Bright TJ, Yoko Furuya E, Kuperman GJ, Cimino JJ, Bakken S. Development and evaluation of an ontology for guiding appropriate antibiotic prescribing. *J Biomed Inform* 2012;45(1):120-128 [FREE Full text] [doi: [10.1016/j.jbi.2011.10.001](#)] [Medline: [22019377](#)]
13. Craven RF, Hirnle CJ. In: Hirnle CJ, editor. *Fundamentals of Nursing: Human Health and Function*. Philadelphia: Lippincott Williams & Wilkins; 2009.
14. Herre HB, Burek P, Hoehndorf R, Loebe F, Michalek H. General Formal Ontology: Part 1 Basic principles, Version 1.0. In: *Onto-Med Report*. Germany: Institute of Medical Informatics, Statistics and Epidemiology, University of Leipzig; 2006.
15. National Institute for Health and Clinical Excellence. NICE Clinical Guidelines. 2006 Dec. Obesity: the prevention, identification, assessment, and management of overweight and obesity in adults and children URL: <http://www.nice.org.uk/CG43> [accessed 2013-06-06] [WebCite Cache ID 6H9zXF7bU]
16. Herre H. General Formal Ontology. A foundational ontology for conceptual modelling. In: *Theory and Applications of Ontology: Computer Applications*. Netherlands: Springer Netherlands; 2010:297-345.
17. Beale T, Heard S. An ontology-based model of clinical information. *Stud Health Technol Inform* 2007;129(Pt 1):760-764. [Medline: [17911819](#)]

18. Chi Y. Rule-based ontological knowledge base for monitoring partners across supply networks. *Expert Syst Appl* 2010;37:1400-1407 [[FREE Full text](#)]
19. Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, et al. Relations in biomedical ontologies. *Genome Biol* 2005;6(5):R46 [[FREE Full text](#)] [doi: [10.1186/gb-2005-6-5-r46](https://doi.org/10.1186/gb-2005-6-5-r46)] [Medline: [15892874](#)]
20. The Open Biological and Biomedical Ontology Foundry. The Open Biomedical and Biomedical Ontologies. URL: <http://www.obofoundry.org/> [accessed 2012-08-30] [[WebCite Cache ID 6H87eOpX1](#)]
21. Elhadad M, Gabay D, Netzer Y. Automatic evaluation of search ontologies in the entertainment domain using text classification. In: *Applied Semantic Technologies: Using Semantics in Intelligent Information Processing*. Edited by Gulla VSaJA. London: Taylor & Francis; 2010.
22. Kehagias DD, Papadimitriou I, Hois J, Tzovaras D, Bateman J. A methodological approach for ontology evaluation and refinement. 2008 Presented at: ASK-IT International Conference; June 26-27, 2008; Nuremberg, Germany.
23. Hartmann J, Spyns P, Giboin A, Maynard D, Cuel R, Suárez-Figueroa M, et al. Methods for ontology evaluation. In: *Knowledge Web Deliverable D1.2.3. EU-IST Network of Excellence: EU-IST Network of Excellence (NoE) IST-2004-507482 KWEB*; 2004.
24. Kang J. Clinical case of obesity clinic in Inje University Paik Hospital. *Health & Sports Medicine* 2000;2:191-194.
25. Jung J, Yang S, Kang J. The case study of prescribed exercise for obese adults. *Health & Sports Medicine* 2007;9(1):77-84.
26. Yu AC. Methods in biomedical ontology. *J Biomed Inform* 2006;39(3):252-266. [doi: [10.1016/j.jbi.2005.11.006](https://doi.org/10.1016/j.jbi.2005.11.006)] [Medline: [16387553](#)]
27. Riaño D, Real F, López-Vallverdú JA, Campana F, Ercolani S, Mecocci P, et al. An ontology-based personalization of health-care knowledge to support clinical decisions for chronically ill patients. *J Biomed Inform* 2012;45(3):429-446. [doi: [10.1016/j.jbi.2011.12.008](https://doi.org/10.1016/j.jbi.2011.12.008)] [Medline: [22269224](#)]
28. Haghighi M, Koeda M, Takai T, Tanaka H. Development of clinical ontology for mood disorder with combination of psychomedical information. *J Med Dent Sci* 2009;56(1):1-15. [Medline: [19697514](#)]
29. Mabotuwana T, Warren J. An ontology-based approach to enhance querying capabilities of general practice medicine for better management of hypertension. *Artif Intell Med* 2009;47(2):87-103. [doi: [10.1016/j.artmed.2009.07.001](https://doi.org/10.1016/j.artmed.2009.07.001)] [Medline: [19709865](#)]
30. Villalonga C, Strohbach M, Snoeck N, Sutterer M, Belaunde M, Kovacs E, et al. Mobile ontology: towards a standardized semantic model for the mobile domain. In: *Proceedings of the First International Workshop on Telecom Service Oriented Architectures*. 2007 Presented at: First International Workshop on Telecom Service Oriented Architectures (TSOA-07); Sept. 17, 2007; Vienna, Austria.

Abbreviations

NICE: National Institute for Health and Clinical Excellence

OBO: Open Biological and Biomedical Ontologies

SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms

XML: Extensible Markup Language

Edited by G Eysenbach; submitted 24.12.12; peer-reviewed by J Lee, J Choi, M Svensson; comments to author 23.01.13; revised version received 03.03.13; accepted 27.04.13; published 28.06.13.

Please cite as:

Kim HY, Park HA, Min YH, Jeon E

Development of an Obesity Management Ontology Based on the Nursing Process for the Mobile-Device Domain

J Med Internet Res 2013;15(6):e130

URL: <http://www.jmir.org/2013/6/e130/>

doi: [10.2196/jmir.2512](https://doi.org/10.2196/jmir.2512)

PMID: [23811542](https://pubmed.ncbi.nlm.nih.gov/23811542/)

©Hyun-Young Kim, Hyeoun-Ae Park, Yul Ha Min, Eunjoo Jeon. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 28.06.2013. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Collecting Maternal Health Information From HIV-Positive Pregnant Women Using Mobile Phone-Assisted Face-to-Face Interviews in Southern Africa

Alastair van Heerden^{1,2}, PhD; Shane Norris³, PhD; Stephen Tollman^{4,5,6}, PhD; Linda Richter^{1,3}, PhD; Mary Jane Rotheram-Borus⁷, PhD

¹Human Sciences Research Council, Pietermaritzburg, South Africa

²School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

³MRC/WITS Developmental Pathways for Health Research Unit, Department of Paediatrics, School of Clinical Medicine, University of the Witwatersrand, Johannesburg, South Africa

⁴MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

⁵Centre for Global Health Research, Umeå University, Umeå, Sweden

⁶INDEPTH Network, Accra, Ghana

⁷Global Center for Children and Families, University of California at Los Angeles, Los Angeles, CA, United States

Corresponding Author:

Alastair van Heerden, PhD

Human Sciences Research Council

PO Box 90

Msunduzi

Pietermaritzburg, 3201

South Africa

Phone: 27 333245015

Fax: 27 333241131

Email: avanheerden@hsrc.ac.za

Abstract

Background: Most of the world's women living with human immunodeficiency virus (HIV) reside in sub-Saharan Africa. Although efforts to reduce mother-to-child transmission are underway, obtaining complete and accurate data from rural clinical sites to track progress presents a major challenge.

Objective: To describe the acceptability and feasibility of mobile phones as a tool for clinic-based face-to-face data collection with pregnant women living with HIV in South Africa.

Methods: As part of a larger clinic-based trial, 16 interviewers were trained to conduct mobile phone-assisted personal interviews (MPAPI). These interviewers (participant group 1) completed the same short questionnaire based on items from the Technology Acceptance Model at 3 different time points. Questions were asked before training, after training, and 3 months after deployment to clinic facilities. In addition, before the start of the primary intervention trial in which this substudy was undertaken, 12 mothers living with HIV (MLH) took part in a focus group discussion exploring the acceptability of MPAPI (participant group 2). Finally, a sample of MLH (n=512) enrolled in the primary trial were asked to assess their experience of being interviewed by MPAPI (participant group 3).

Results: Acceptability of the method was found to be high among the 16 interviewers in group 1. Perceived usefulness was reported to be slightly higher than perceived ease of use across the 3 time points. After 3 months of field use, interviewer perceptions of both perceived ease of use and perceived usefulness were found to be higher than before training. The feasibility of conducting MPAPI interviews in this setting was found to be high. Network coverage was available in all clinics and hardware, software, cost, and secure transmission to the data center presented no significant challenges over the 21-month period. For the 12 MHL participants in group 2, anxiety about the multimedia capabilities of the phone was evident. Their concern centered on the possibility that their privacy may be invaded by interviewers using the mobile phone camera to photograph them. For participants in group 3, having the interviewer sit beside vs across from the interviewee during the MPAPI interview was received positively by 94.7% of MHL. Privacy (6.3%) and confidentiality (5.3%) concerns were low for group 3 MHL.

Conclusions: Mobile phones were found both to be acceptable and feasible in the collection of maternal and child health data from women living with HIV in South Africa.

Trial Registration: Clinicaltrials.gov NCT00972699; <http://clinicaltrials.gov/ct2/show/NCT00972699> (Archived by WebCite at <http://clinicaltrials.gov/ct2/show/NCT00972699>)

(*J Med Internet Res* 2013;15(6):e116) doi:[10.2196/jmir.2207](https://doi.org/10.2196/jmir.2207)

KEYWORDS

mobile phones; human immunodeficiency virus; mobile health

Introduction

One-half of people living with human immunodeficiency virus (HIV) globally are women and 76% of all HIV-positive women live in sub-Saharan Africa [1]. Sub-Saharan Africa accounts for almost half of the world's maternal, newborn, and child deaths with 4.7 million children and 276,000 women dying annually [2,3]. Mothers living with HIV (MLH) are particularly vulnerable and at risk of adverse maternal outcomes with at least 20% of maternal deaths being HIV-related [1]. South Africa makes up a large proportion of this disease burden with 3.2 million women living with HIV [1], of whom 200,000 annually are pregnant [4]. The Province of KwaZulu-Natal has the highest HIV prevalence in South Africa [5]. The national prevalence is approximately 11% [5], whereas 40% to 60% of pregnant women in rural KwaZulu-Natal are HIV-positive [6,7].

In response to this crisis, South Africa has implemented the Prevention of Maternal-to-Child Transmission (PMTCT) package as recommended by the World Health Organization (WHO). The PMTCT program requires newly pregnant women to complete a series of sequential steps, also known as the PMTCT cascade [8], that are aimed at first diagnosing and then treating HIV infection [9]. Under ideal circumstances, in which no barriers exist to the completion of all tasks, PMTCT has been shown to be highly effective at reducing HIV transmission to less than 2% at childbirth [10]. Under the less than ideal circumstances faced by many MLH in low- and middle-income countries, transmission can occur in approximately 1 in 4 deliveries [11]. Some of the challenges that make adherence to each of the PMTCT tasks difficult are being unable to afford transportation to the clinic, fear of stigmatization, increased household conflict, and lack of partner support [8]. To better understand these barriers and their impact on the loss of women through the PMTCT cascade requires accurate, timely, and detailed information. Yet gathering data that tracks women's progress through PMTCT to determine how well the system is performing is beset by its own challenges.

One of the major challenges faced in the collection of high-quality data is human resource constraints. Trained and qualified health staff are in short supply in many resource-constrained settings [12]. Task-shifting strategies used to address these shortages, have led to a perceived increased burden for data collection and collation [13]. Minimal support, delayed feedback, little understanding of the usefulness of the data, and no interpretation of raw scores mean that the completion and submission of multiple paper-based registers for statistical purposes is often regarded as a low priority for busy and overburdened staff [13]. Therefore, it is unsurprising

that register data has been shown to be fragmented, error prone, incomplete, and inaccessible [14,15].

The usefulness of this register data is further compromised by the fact that it is unlinked, ie, it is not possible to link individual clinic visits to the same person. This means that although aggregate data may be available, it is difficult to track the path of individuals through the PMTCT cascade. This makes it difficult to identify and understand the bottlenecks in the cascade. These constraints and challenges raise the question whether there is not a more efficient way to support staff and perform monitoring and evaluation of the PMTCT program in geographically remote primary health care facilities [16].

There is currently a groundswell of interest in the use of mobile phones and information communication technologies in the support of health [17,18]. The growing body of mHealth literature provides examples of the use of mobile phones as data collection tools in low- and middle-income countries. Although evidence is mixed, the use of mobile phones as data collection tools has been found to increase data quality, speed up the turnaround time from collection to analysis, and improve interfacility communications [19-22]. Poor follow-up rates could potentially be improved with mobile phones by providing links to participants that are cheaper and more immediate than travel to facilities. These advantages over traditional paper-based clinic registers suggest mobile phones are a potential tool with which to address some of the challenges currently experienced in collecting health information through the PMTCT cascade. The aim of this paper is to describe the feasibility and acceptability of using a mobile phone survey application to collect data from pregnant women living with HIV enrolled into the PMTCT program in KwaZulu-Natal, South Africa.

Method

Study Design

This study was nested within a larger clinic-based randomized cluster trial known as Project Masihambisane ("let us walk together"; Clinicaltrials.gov NCT00972699) [23]. The primary study aimed to improve mental and physical health outcomes of HIV-positive mothers and their babies by supplementing the PMTCT with paraprofessional peer mentors. Using a mixed-methods design, qualitative data from a single, small focus group were supplemented with 2 quantitative questionnaires collected using a mobile phone survey application. Paradata, or data on the data collection process, were gathered in order to examine the feasibility of mobile phone-assisted personal interviewing (MPAPI).

Mobile Phone-Assisted Personal Interviewing

The MPAPI survey platform was supplied by Mobenzi Researcher [24], a commercial vendor based in South Africa. The solution offered by Mobenzi includes both a mobile application and a Web portal. The Java Platform Micro Edition (Java ME) application runs on all handsets compliant with mobile information device profile (MIDP) 2.0. It provides full survey functionality, including the ability to create various question types, mark fields as mandatory, and intelligently manage survey branching (Figure 1). The software is now also available for Android handsets. The Java ME application was installed on Nokia E61 handsets. These mobile phones run on the Symbian S60 operating system, have a 2.9-inch thin film transistor screen, 64 megabyte random-access memory (RAM), Bluetooth, Wi-Fi, a QWERTY keyboard, and a 1500 milliamp hour battery.

Figure 1. Example of the Mobenzi Researcher application running on a Nokia handset.

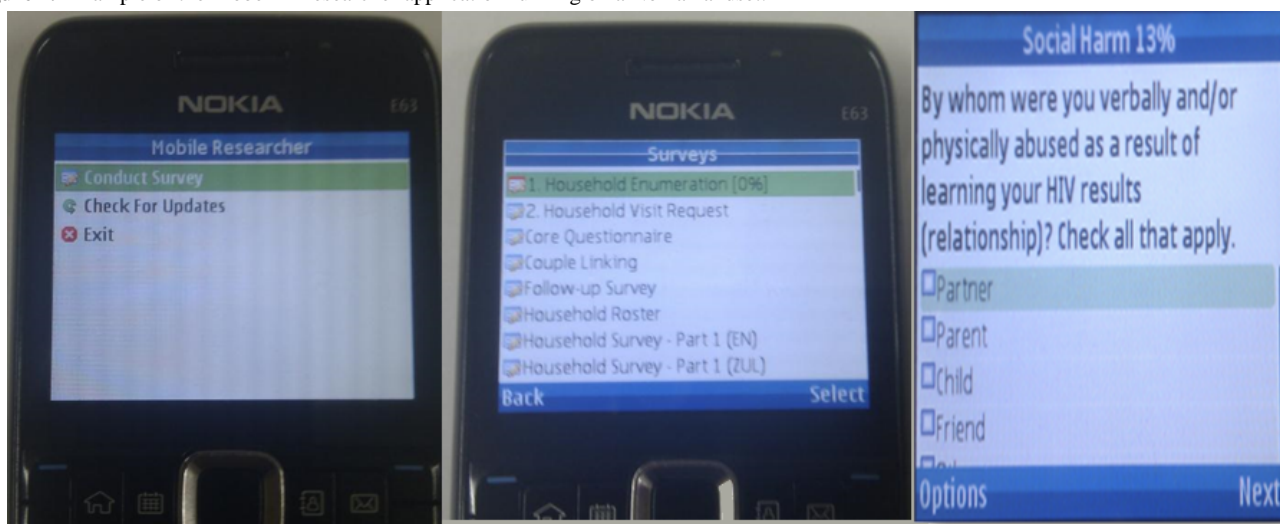


Figure 2. Example of Mobenzi Researcher Web portal: survey design.

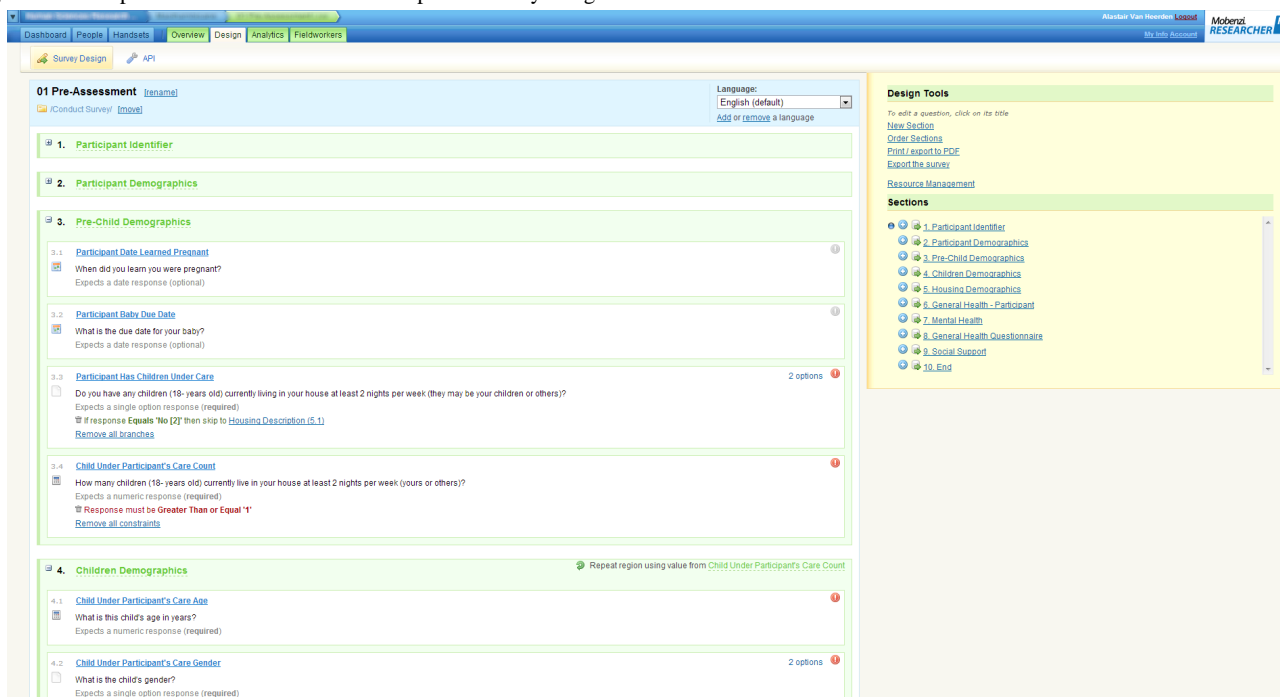


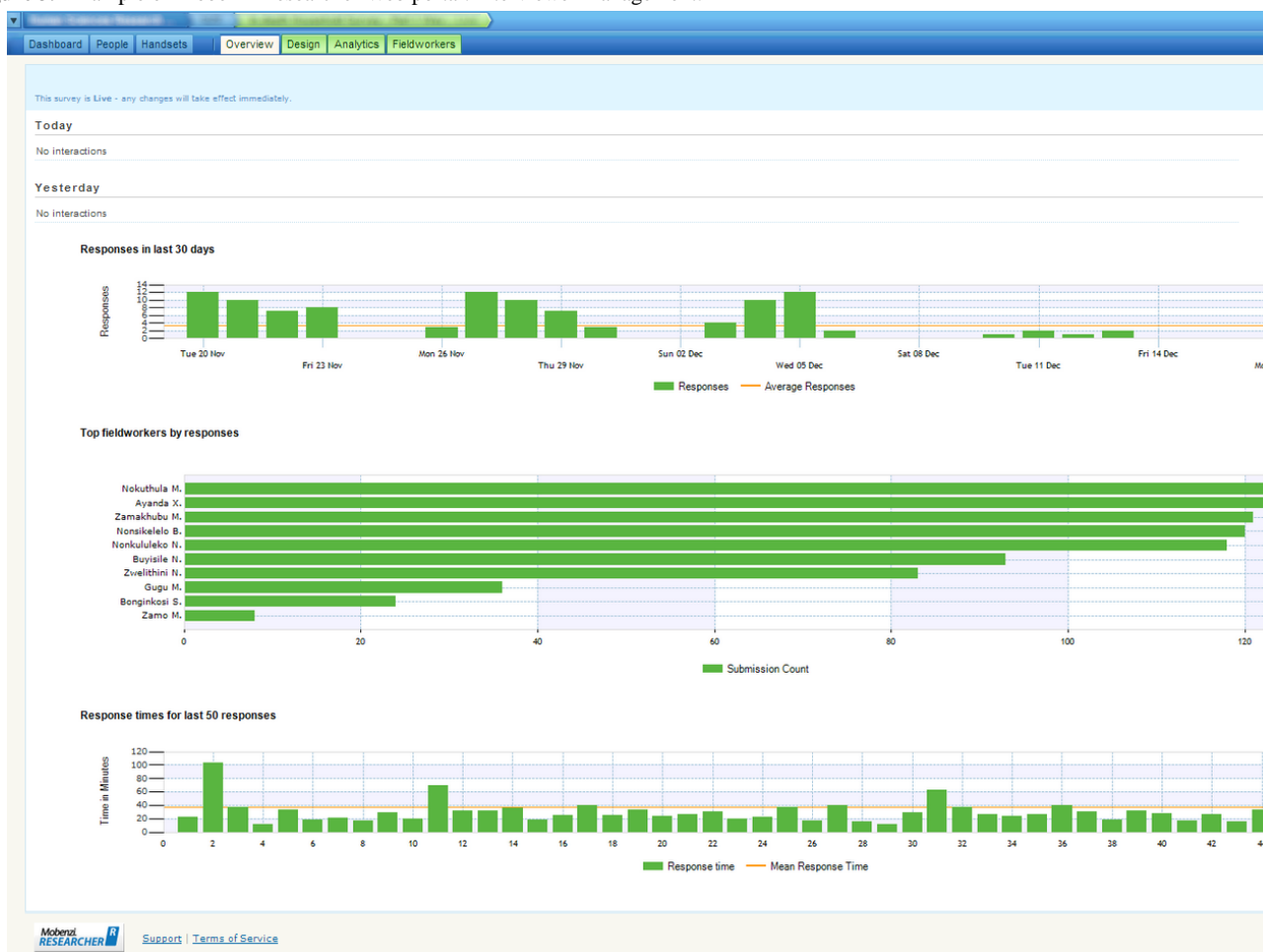
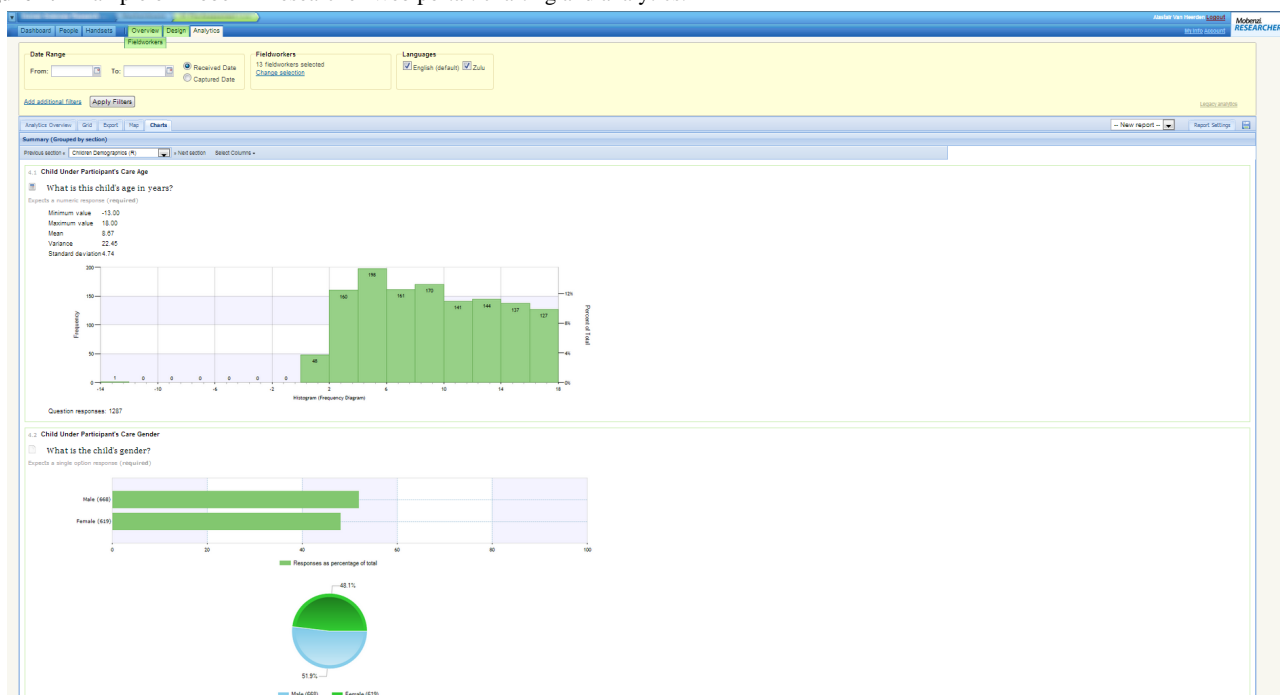
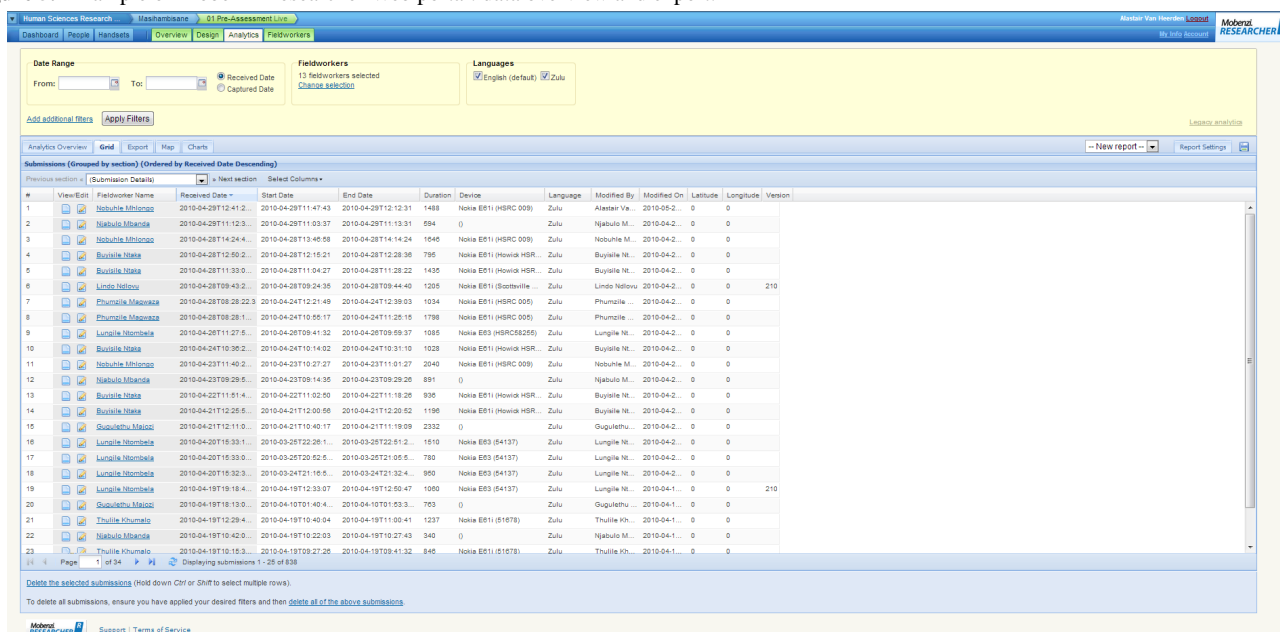
Figure 3. Example of Mobenzi Researcher Web portal: interviewer management.**Figure 4.** Example of Mobenzi Researcher Web portal: charting and analytics.

Figure 5. Example of Mobenzi Researcher Web portal: data overview and export.**Figure 6.** Example of Mobenzi Researcher in field use.

Participants and Procedures

Three groups of participants took part in this study. The first group of participants (participant group 1) included the 16 interviewers recruited and trained on MPAPI for the primary

study. Although these participants were all familiar with pen-and-paper questionnaires, none had ever used MPAPI. Before receiving any training on MPAPI, the group were asked to complete a short questionnaire with items from the Technology Acceptance Model proposed by Davis [25]. The

scale contains 12 items, 6 relating to the perceived usefulness of mobile phones for research purposes and 6 related to the perceived ease of their use (Table 1). Each item was rated on a 5-point Likert-type scale ranging from extremely likely to extremely unlikely. Following 3 days of training, a posttraining assessment was conducted using the same questionnaire. After completing training, the interviewers were dispatched to 8 primary health care clinics in a rural district typical of those found in the region where the interviewers were originally interviewing pregnant women living with HIV for the primary study. After 3 months, the questionnaire was administered for a third time to gauge if use of the phone in the field had altered perceptions and attitudes toward the tool.

The second group (participant group 2), recruited for participation in the focus group, consisted of 12 pregnant women with HIV living in this region of South Africa who were enrolled through community forums set up for the larger study. Before the focus group was held, each woman was interviewed by a research assistant using a mobile phone. The questions asked ranged from general health questions about the participants' knowledge of HIV to more sensitive questions about recent sexual activity, condom use, and disclosure of HIV status to their partners. After completing the questionnaire, the group was brought together to participate in a focus group about their experience. Five questions were designed to facilitate the discussion about general mobile phone ownership and their reactions to the interview conducted by mobile phone. The focus group lasted for an hour with extensive notes being taken about responses.

The third group of participants (participant group 3) were pregnant women living with HIV recruited through the primary study. Participants were recruited in the Umgungundlovu Health District of KwaZulu-Natal, a province in South Africa with 10 million people, over half (57%) living in rural areas. The

Umgungundlovu District includes 7 local authorities, has 48 fixed clinics, 4 community health centers, 9 tertiary hospitals, and an estimated population of some 995,000 persons according to a 2007 estimate [26]. From this district, 8 clinics were selected through a clinic audit conducted in December 2007. The selection criteria, applied to all potential clinics in the district during the audit, were presence of other research trials, clinic size, availability of antenatal and postnatal services, and uptake of antenatal and postnatal services at the clinic. The clinic audit resulted in the selection of 4 pairs of clinics for the study, matched on size (small vs large) and geography (rural vs urban). Women who met the eligibility criteria of being 18 years or older, less than 34 weeks pregnant, and planning to reside in the study area for the duration of their pregnancy were invited to participate in the primary study. If they accepted, a baseline health questionnaire was completed by an interviewer using a mobile phone survey application. Using a cross-sectional design and a continuous sampling strategy that started approximately halfway through the primary study, 512 of the 1204 women enrolled into the primary study were recruited to participate in this substudy. This cross-section of women completed a second questionnaire, again using MPAPI, in which they were asked to describe their views about the mobile phone survey.

Statistical Analysis

A thematic analysis was performed on the qualitative data generated by focus group discussions. This data was supplemented with exploratory data analysis techniques, such as frequency analysis and chi-square (χ^2) statistics, performed on the quantitative questionnaire data. Univariate analysis was used to analyze perceived ease of use and perceived usefulness scale data. Data was downloaded as comma-separated values from the online Mobile Research database. This comma-separated values file was then imported into SPSS 19 (IBM Corp, Armonk, NY, USA).

Table 1. Technology Acceptance Model scale.

Items per subscale	Question text
Perceived usefulness	
1	Using cell phones to collect data in my job would enable me to accomplish tasks more quickly
2	Using cell phones to collect data would improve my job performance
3	Using cell phones to collect data in my job would increase my productivity
4	Using cell phones to collect data would enhance my effectiveness on the job
5	Using cell phones to collect data would make it easier to do my job
6	I would find cell phones useful in my job
Perceived ease of use	
1	Learning to operate a cell phone would be easy for me
2	I would find it easy to get a cell phone to do what I want it to do
3	My interaction with cell phones would be clear and understandable
4	I would find cell phones flexible to interact with
5	It would be easy for me to become skilful at using a cell phone
6	I would find a cell phone easy to use

Fieldwork Supervision and Ethical Approval

A team of 16 data collectors administered the questionnaire and 2 coordinators employed by the primary trial supervised all aspects of the study in the field. The coordinators' role included managing informed consent, supporting field activities, and monitoring data quality through quality assurance checks and ongoing training and supervision of interviewers. Ethical approval for the study was obtained from the Committee for Research on Human Subjects (Medical) Witwatersrand Human Ethics Committee (M091035). A study information sheet was presented to all participants and the study was explained in detail before participants signed a form giving written consent to participate.

Results

Acceptability

Participant Group 1: Interviewers

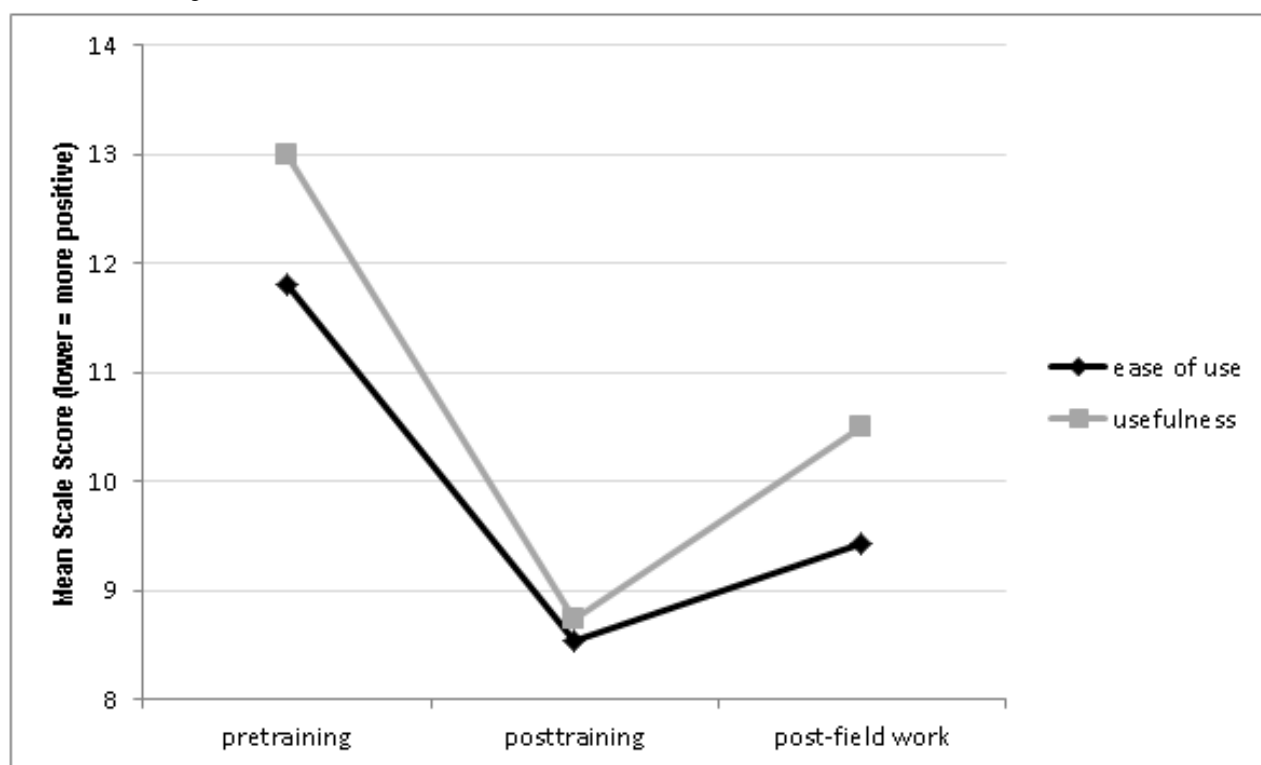
Sixteen interviewers were recruited, trained, and dispatched to 8 health facilities as part of the primary trial. All participants were female with an average age of 27 years. Phone ownership was high with 15 of 16 owning their own mobile phone. All interviewers had completed their secondary education, with 14 having completed a 3-year tertiary degree. All had previous experience with paper-and-pen questionnaires, but none reported ever having performed a mobile phone-assisted interview.

A 2-way repeated measures ANOVA was performed on the 2-factor Technology Acceptance Model scale (ease of use and

usefulness) across 3 time points (pretraining, posttraining, and postuse). There was a significant main effect for the time point at which the data was collected ($F_{1,189,16,649}=8.62$, $P=.007$). However, there was no significant main effect between perceived ease of use and usefulness ($F_{1,14}=0.924$, $P=.35$) nor any interaction effect between these factors and time ($F_{2,28}=0.621$, $P=.54$). The findings are presented in Figure 7. A maximum score of 55 implies extreme dislike of the tool; therefore, interviewer attitudes toward ease of use and usefulness of mobile phones were already positive before training.

After 3 days of training, interviewers reported perceiving the mobile phones as significantly easier to use than at the baseline. Further, the perception of mobiles as a useful tool with which to collect data underwent an even greater positive shift. The training also increased positive perceptions of the mobile as a useful way in which to collect data. The positive perceptions were linked to the insight that the mobile phones would allow more accurate data to be collected in a timely manner with less administrative burden than paper forms. After 3 months of use, perceptions of both the usefulness and ease of use of mobile phone-assisted questionnaires decreased, but remained above levels reported pretraining. This decrease in perceived usefulness and ease of use is thought to be related to a more realistic view of the device's potential being realized through practical experience. It is important to note that real world use did not significantly reduce perceptions of either usefulness or ease of use.

Figure 7. Estimated marginal means for the main effect of time.



Participant Group 2: Focus Group With 12 Mothers Living With HIV

All focus group participants were HIV positive and pregnant. Most (9/12) had learned of their HIV status through the PMTCT program during their first pregnancy. For 4 of the women, this was their first pregnancy and only 1 of the 4 already knew that she was HIV positive. The mean age of the focus group participants was 26 years. Despite an estimated household (not per capita) income, in South African rand (R), of R1600 (equivalent to US \$177) per annum for rural communities in this district [26], all 12 women in the focus group had access to a mobile phone at home, with a full two-thirds (8/12) personally owning a phone. The 4 women who did not own a cell phone had access to a sibling's phone. When asked how important access to a cell phone was for them, all but 1 stated that it was extremely important.

Reflecting on the experience of being interviewed using a mobile phone, a number of women reported initial feelings of anxiety about the multimedia capabilities of the phone. In particular, women were concerned about being filmed or photographed by the camera, which had a lens situated on the back of the phone. One woman said, "I feel it is ok [to be interviewed on the phone], but it is slightly scary. I feel as though you are recording a video of me. But it's all right. It was not nice when the interviewer held the phone too close to me because I was worried about the phone camera...whether I was being recorded."

Two important factors mediated this experience. Firstly, the model of Nokia phone used influenced how comfortable the interviewees felt. The less familiar-looking Nokia E61 business phone raised more concerns from participants than the more familiar candy bar-shaped Nokia 2630. Secondly, the interview style adopted by the interviewer impacted significantly on the experience. Although all interviews were one-on-one, some interviewers took a position directly opposite the woman whereas other interviewers adopted a side-by-side style. For those women who experienced the side-by-side interview, being filmed was reported as being less of a concern. Women who were interviewed by an interviewer sitting directly opposite them felt that side-by-side interviewing would have alleviated much of their anxiety about being photographed. The side-by-side approach had the added benefit of allowing women, who reported being curious about the questionnaire, to see what had been asked and captured on the phone screen. One participant said, "I enjoyed that the interviewer let me see the questions and shared with me what was being recorded." This information was incorporated into the primary trial by training field workers to offer MLH the opportunity to sit alongside them during the interview. Table 2 presents other significant themes that emerged from the analysis.

Participant Group 3: Questionnaire About Mobile Phone-Assisted Personal Interviews

From October 2010 to March 2011, Project Masihambisane conducted 708 interviews with 520 participants (some participants completed interviews at 2 different time points). Of these 520 participants, 512 (98.5%) agreed to take part in the mobile phone questionnaire. The mean age of participants

was 27 years with participants speaking primarily isiZulu (93.4%). A small minority had either no schooling (1.6%) or a tertiary qualification (5.8%). Of the remaining 92.6%, 36.4% completed primary education, 39.9% started but did not complete secondary schooling, and 16.3% graduated from secondary school. Approximately the same percentage of women were married (10%) as were living alone (11.9%); the others were either married and living together or apart, or single and in a coresident or non-coresident relationship. There was a 60/40 split between women living in rural vs periurban areas with 35.5% having running water on the premises, 59.2% having flush toilets, and 82.6% having electricity.

Despite 42.4% of the sample being unemployed, 84.0% stated that they owned a mobile phone. Most of these handsets were manufactured by Nokia, running on the Symbian operating system and were able to run Java ME applications (87.4%), a version of the Java programming language designed to run on resource-constrained devices such as mobile phones. From self-reports, the mean length of ownership was 3.3 years with a median amount of R30 being spent on airtime per month. Table 3 compares the demographic profile of those who owned vs those who did not own a mobile phone. Chi-square tests found significant associations between age ($\chi^2_2=6.34$, $P=.04$), education ($\chi^2_2=33.52$, $P<.001$), socioeconomic status ($\chi^2_2=27.65$, $P<.001$), occupation ($\chi^2_2=8.16$, $P=.004$), and housing characteristics (water: $\chi^2_1=9.79$, $P=.002$; electricity: $\chi^2_1=9.6$, $P=.002$; flush toilet: $\chi^2_1=12.68$, $P<.001$), and ownership of a mobile phone. Examining confidence intervals for each demographic category support these results for all but age. Pregnant women with HIV were less likely to own a phone if they were 20 years old or younger, had a primary school education, were in the lowest socioeconomic status group, and did not have household luxuries, such as running water, electricity, or flushing toilets.

On a 3-point scale, the participants rated the experience of the mobile questionnaire as positive, neutral, or negative. Of the 512 participants, 485 (94.7%) reported a positive experience (Table 4). There were 24 (4.7%) neutral responses from participants and 3 (0.6%) who disliked the mobile questionnaire. With respect to questionnaire mode preference, one-third (29.5%) had no preference, two-thirds preferred the mobile questionnaire (67.8%), and 2.7% preferred pen-and-paper questionnaires.

A 1-way ANOVA revealed a significant increase in positive ratings when comparing the number of interviews completed before taking the questionnaire. That is, people who had completed 4 MPAPI interviews before taking the questionnaire were significantly more likely to rate the questionnaire experience positively than those participants taking an MPAPI questionnaire for the first time ($F_{3,508}=6.795$, $P=.009$).

Feasibility

Participant Group 3: Questionnaire About Mobile Phone-Assisted Personal Interviews

Over the course of the primary study (July 2008 to April 2010), 13,653 mobile phone–assisted questionnaires were submitted and 13,650 were received by the data center. This number includes the questionnaires that were used to capture 3 full-length health assessments, the patient-carried maternal health cards, and the daily clinic attendance registers. All clinics had good network coverage and questionnaires were received, on average, 4.2 minutes after being completed. Participants were identified through the capturing of 2 unique numeric identifiers. Of the 3012 full-length questionnaire assessments, 57 (1.9%) were not able to be linked to a participant because of errors in the first 11-digit unique identifier. All 57 cases were resolved by referring to the second 5-digit identifier.

Of the 18 mobile phones that were bought for the study, 1 handset was stolen during a mugging and 1 became unusable 13 months into the study after a hardware failure. The remaining handsets are still functional and in use. Over the course of the study, there were very few hardware and/or software challenges.

To deal with the possibility that phones would be used for private calls and other electronic communications, mobile phones were loaded with a fixed amount of airtime (US \$25) at the beginning of every month. When this allocation was consumed, interviewers were responsible for airtime top-ups. This situation occurred only a few times during the course of the study. During these times, interviewers were able to come to the office and upload the data using a Wi-Fi connection. Automatic processing of submitted data enabled field team leaders to monitor the number of interviews, length of interviews, and quality of interview responses being submitted by interviewers on a Web-based dashboard page. An example from the remote monitoring Web dashboard is presented in [Figure 8](#). At a glance, field coordinators were able to see the minimum and maximum number of interviews per clinic, typical monthly range, current performance, and a trend indicator. These figures could be filtered by questionnaire, year, and month. This quick access to reliable data made it possible to rapidly respond to the ever-changing challenges faced by interviewers working in a rural clinic environment. The approach would also be more easily scaled up to the monitoring of many more clinics than alternative management strategies based on paper and data capturing into a locally stored database.

Table 2. Themes emerging from the analysis of the focus group session held with 12 mothers living with HIV (participant group 2).

Theme	Indicative quote ^a
Connectivity	<p><i>My phone is important to me as it allows me to communicate immediately with people who want to contact me.</i> [Gladys Bhengu, 22 years]</p> <p><i>I use it to get in contact with my family members and people who are living around me, like my neighbors.</i> [Ethyl Zuma, 36 years]</p>
Safety	<p><i>Public phones are far away from where I live. Cell phones [is] convenient and allows one to get help from others through communication.</i> [Marsha Ntuli, 22 years]</p> <p><i>...to be able to reach people when I have a problem and vice versa.</i> [Gladys Bhengu, 22 years]</p> <p><i>I use it for emergencies...to get help.</i> [Mavis Sithole, 24 years]</p>
Functionality	<p><i>I put reminders in my phone to help remind me of the times when I must take my medication. So I am on time with my medication.</i> [Ethyl Zuma, 36 years]</p>
Knowledge	<p><i>Because you first explain everything before taking any information on the cell phone it will be ok.</i> [Wendy Zulu, 28 years]</p> <p><i>As long as the interviewer will explain the study well people won't mind.</i> [Ayanda Sithole, 31 years]</p>
Privacy and confidentiality	<p><i>They will think that researchers are taking pictures of HIV-positive people to label them.</i> [Zama Ndwandwe, 33 years]</p> <p><i>Researchers are using the devices to film people and to put them on television.</i> [Ethyl Zuma, 36 years]</p> <p><i>Information will be between the two of us because the phone probably will have a code...I don't think the data will have a possibility to be lost.</i> [Wendy Zulu, 28 years]</p>

^aAll names are fictitious.

Table 3. Demographic characteristics of sample (N=512).

Respondent characteristics	Do not own phone (n=82)			Own phone (n=430)		
	n	%	95% CI	n	%	95% CI
Age (years)						
≤20	19	23.2	15.4-33.4	55	12.8	10.0-16.3
21-30	39	47.6	37.1-58.2	247	57.4	52.7-62
31-40	24	29.3	20.5-39.9	128	29.8	25.6-34.3
Education						
Primary schooling ^a	28	34.1	24.8-44.9	45	10.5	7.9-13.7
Secondary schooling ^a	53	64.6	53.8-74.1	356	82.8	78.9-86.1
Tertiary schooling	1	1.2	0.2-6.6	29	6.7	4.7-9.5
Socioeconomic status						
Lower ^a	25	30.5	21.6-41.1	48	11.2	8.5-14.5
Middle	53	64.6	53.8-74.1	293	68.1	63.6-72.4
Upper ^a	4	4.9	1.9-11.9	89	20.7	17.1-24.8
Occupation						
Full-time ^a	6	7.3	3.4-15.1	89	20.7	17.1-24.8
Part-time	27	32.9	23.7-43.7	121	28.1	24.1-32.6
Unemployed	44	53.7	42.9-64.0	174	40.5	35.9-45.2
Other	5	6.1	2.6-13.5	46	10.7	8.1-14
Region						
Local government administration	36	43.9	33.7-54.7	171	39.8	35.3-44.5
Traditional administration	46	56.1	45.3-66.3	259	60.2	55.5-64.7
Housing						
Running water of premises ^a	48	58.5	47.7-68.6	324	75.3	71.1-79.2
Electricity ^a	58	70.7	60.1-79.5	365	84.9	81.2-88.0
Flush toilets ^a	34	41.5	31.4-52.3	269	62.6	57.9-67.0

^aStatistically significant ($P<.05$).

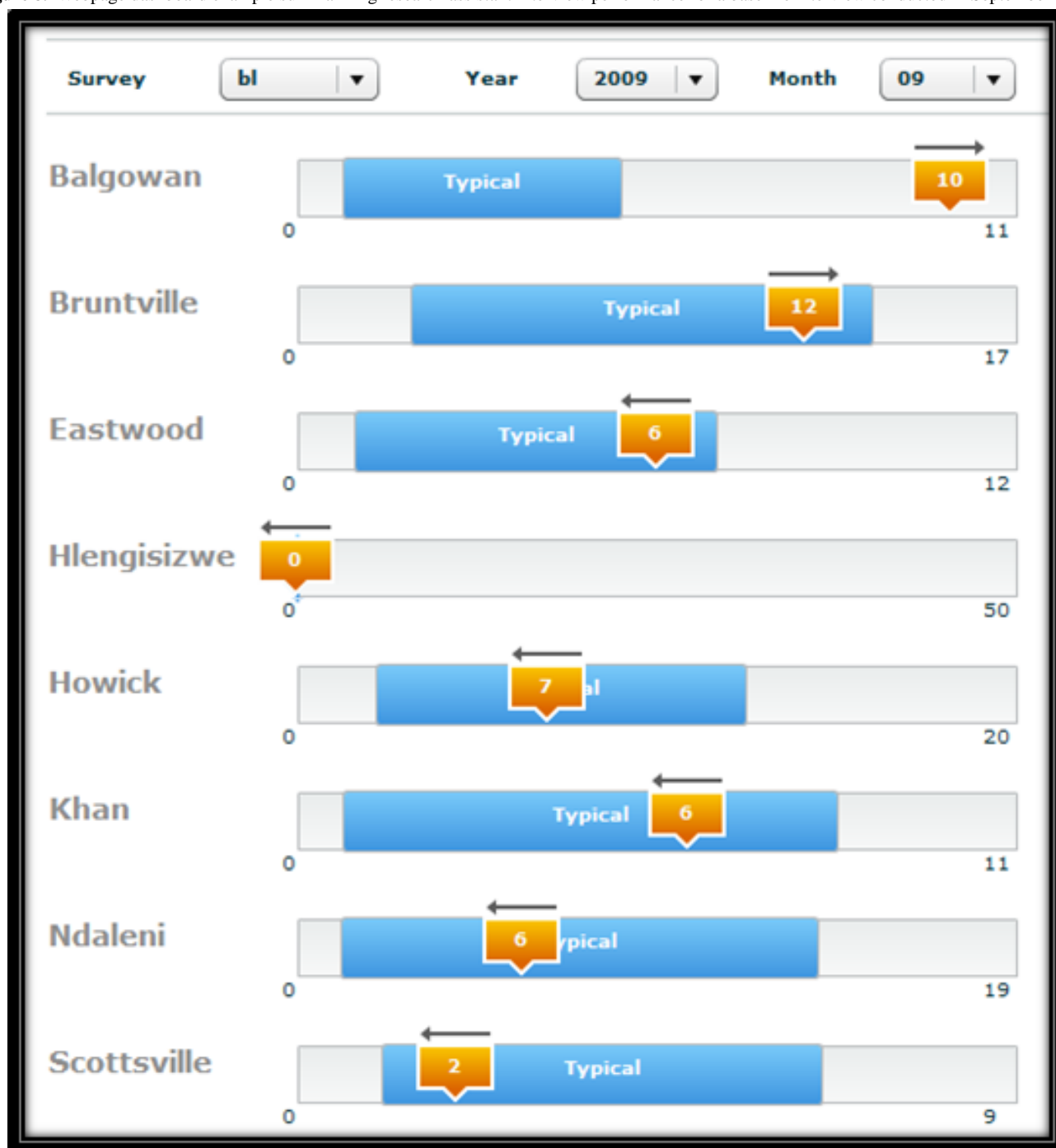
Figure 8. Webpage dashboard example summarizing research assistant interview performance for a baseline interview conducted in September 2009.

Table 4. Questionnaire experience and preference of sample (N=512).

Respondent responses	Do not own phone (n=82)			Own phone (n=430)		
	n	%	95% CI	n	%	95% CI
Questionnaire experience						
Positive	76	92.7	84.9-96.6	409	95.1	92.6-96.8
Neutral	6	7.3	3.4-15.1	18	4.2	2.7-6.5
Negative	0	0.0	0-4.5	3	0.7	0.2-2.0
Questionnaire preference						
Phone	50	61.0	50.2-70.8	297	69.1	64.5-73.3
Paper and pen	5	6.1	2.6-13.5	9	2.1	1.1-3.9
No preference	27	32.9	23.7-43.7	124	28.8	24.8-33.3
Questionnaire concerns						
Concerned about privacy	9	11.0	5.9-19.6	23	4.1	3.6-7.9
Concerned about confidentiality	6	7.3	3.4-15.1	21	4.9	3.2-7.4

Discussion

The present study evaluated the acceptability and feasibility of using mobile phones to collect data from pregnant women living with HIV who were enrolled into the PMTCT program in KwaZulu-Natal, South Africa. Results suggest acceptability of the method by both affected women and the interviewers trained to collect the mobile phone questionnaires. Feasibility of collecting data from remote primary health care facilities was found to be high in this particular context.

Acceptability

Women Living With HIV (Participant Groups 2 and 3)

To prepare for interviewing in a clinic environment pregnant women living with HIV, a focus group with 12 women was undertaken. The experience of completing a face-to-face mobile phone-assisted interview was acceptable to most women. The major concerns raised in the focus group discussion suggested 2 important recommendations. First, if mobile phones are to be used in health settings as data collection tools, it is advisable for health staff to adopt an open side-by-side interviewing style. By placing themselves alongside the women, concerns about data privacy, participant confidentiality, and interview anxiety can be minimized. Secondly, using the informed consent process to explain exactly how data will be collected, how it will be transmitted, and who will have access to this private information may reduce participant anxiety about the MPAPI process. Taking these considerations into the clinic-based questionnaire, 512 women were asked to complete a short questionnaire about their experience. Overall, acceptability of this interview method was confirmed with 94.7% of women, who viewed the experience positively. Only 3 women reported the experience as being negative for them. Given a choice as to whether future interviews should be conducted using mobile phone-assisted interviewing or paper-and-pen, only 2.7% expressed a preference for paper-and-pen.

Ownership of mobile phones among pregnant MLH, although high at 84%, was significantly associated with age, education,

socioeconomic status, and occupation. This has both positive and negative implications. Positively, the potential of mHealth to reach MLH was confirmed with 8 out of every 10 women owning a mobile phone. However, caution must be exercised to ensure that mHealth programs, which often aim to increase access to marginalized and underresourced individuals, do not inadvertently further disadvantage the most disadvantaged MLH.

Interviewers (Participant Group 1)

Perceived ease of use and perceived usefulness of MPAPI were consistently rated highly by clinic-based interviewers. It is not clear whether this rating is in response to a dislike of paper-based data collection or an affinity for mobile phones. Although perceived ease of use and perceived usefulness dropped after 3 months of field use, they remained above pretraining perceptions. The decrease is likely linked to the fact that real world use is often more limited than portrayed under the ideal conditions made possible in a controlled training environment. Nevertheless, the decrease was not significant and remained high for both perceived usefulness and ease of use. This finding suggests that after extended clinic-based use, the mobile phone-assisted interviews remained useful and easy. Training was found to significantly increase perceived ease of use and perceived usefulness and highlights the importance of providing a comprehensive training package to familiarize interviewers with both the concepts and practice of MPAPI.

Feasibility

The feasibility of using mobile phone-assisted interviews as part of a clinic-based PMTCT program was monitored for 21 consecutive months. During this time, over 13,000 data forms were submitted with only 3 reported as being sent by interviewers and not received. Very few phones were lost, broken, or stolen and network coverage was good at all 8 clinic facilities. A number of strategies were used to control airtime use, the most effective being a fixed amount of airtime loaded onto the handset at the beginning of the month and interviewers topping up if they depleted this amount. Other strategies included interviewers requesting small amounts of airtime as they needed, weekly checks by a coordinator of what numbers

had been dialed, and a usage policy that outlined what the phone may and may not be used for. Initial concerns about charging handset batteries were unfounded as all clinics had electricity.

These limited challenges were supplemented by a number of perceived benefits. The management of geographically dispersed interviewers was improved through the use of real-time data presented in visually appealing summaries. This allowed problems to be flagged much quicker than if data was captured on paper and collected from the clinics on a monthly basis. Although not used in this study, the ability to geotag questionnaire responses could also assist in the management of community health workers providing home-based care. Questionnaire management was also found to be improved with a virtual dashboard giving valuable up-to-date information on key project indicators. Whenever a new questionnaire needed to be introduced or changes made to an existing questionnaire, updates could be rolled out immediately to each clinic without the need to recall out-of-date paper forms. This problem besets current paper-based registers. Whenever a new indicator needs to be collected or a change in indicator definitions occur, old forms continue to be submitted by facilities who either do not receive the updated forms or prevent wastage by continuing to use old form stock. Finally, this study found it feasible to track and link the PMTCT data of participants over a 10- to 14-month period. Although 2% of the records could not be linked because of data entry errors in the unique identifier, the use of a second identifier resolved these cases.

Relevance to the Prevention of Maternal-to-Child Transmission Cascade

Paper-based records that inform the KwaZulu-Natal District Information System have been shown to be of poor quality and

in a format that make the tracking of women through the PMTCT cascade difficult [15]. Proposed as an alternative in other settings [19,27], little is known about the acceptability and feasibility of using mobile phones to collect data from WLH in this setting. The findings presented here suggest that MPAPI is indeed acceptable to both MLH and interviewers. The use of mobile phones for data collection also proved feasible and provided a number of advantages to the data collection process over traditional ways of collecting data in a rural clinic setting. One key benefit for the PMTCT program was the ability to link women's data over a number of visits. This made it possible to track enrollment, uptake, and loss of individual women through the PMTCT cascade.

Limitations and Future Work

The results obtained in this study suggest that MPAPI may be both feasible and acceptable to use within the South African PMTCT program. South Africa is unique among African countries in its broad mobile network coverage, well-maintained mobile infrastructure, and electrification of rural health facilities. How well this approach to the collection of PMTCT data would perform in the context of other low- and middle-income countries is, therefore, a limitation of this work.

The bidirectional communication abilities of mobile phones also holds promise as a way to provide regular and relevant feedback to clinics. With a large proportion of women owning mobile phones, future studies should explore the acceptability and feasibility of extending PMTCT data collection into the home using mobile phone-assisted self-interviewing.

Acknowledgments

This study was funded by the National Institute of Mental Health (1R01MH077553).

Conflicts of Interest

None declared.

References

1. UNAIDS Report on the Global AIDS Epidemic 2012. Geneva: Joint United Nations Programme on HIV/AIDS (UNAIDS); 2012. URL: http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2012/gr2012/20121120_UNAIDS_Global_Report_2012_en.pdf [accessed 2013-05-23] [WebCite Cache ID 6GpTkZVFs]
2. State of the World's Children 2009: Maternal and Newborn Health. New York: United Nations Children's Fund (UNICEF); 2008. URL: http://www.unicef.org/cotedivoire/SOWC_2009_.pdf [accessed 2013-05-23] [WebCite Cache ID 6GpU0nMax]
3. Kinney MV, Kerber KJ, Black RE, Cohen B, Nkrumah F, Coovadia H, Science in Action: Saving the lives of Africa's Mothers, Newborns, and Children working group, et al. Sub-Saharan Africa's mothers, newborns, and children: where and why do they die? PLoS Med 2010 Jun;7(6):e1000294 [FREE Full text] [doi: [10.1371/journal.pmed.1000294](https://doi.org/10.1371/journal.pmed.1000294)] [Medline: [20574524](https://pubmed.ncbi.nlm.nih.gov/20574524/)]
4. Children and AIDS: Fourth Stocktaking Report 2009. Geneva, Switzerland: UNAIDS; 2009. URL: http://www.unaidsrsta.org/sites/default/files/children_aids_unicef_report_2009.pdf [accessed 2013-05-23] [WebCite Cache ID 6GpUENmNw]
5. Shisana O, Rehle T, Simbayi L, Zuma K, Jooste S, Pillay-van-Wyk V. Human Sciences Research Council Press. Cape Town, South Africa: HSRC Press; 2010. South African national HIV prevalence, incidence, behaviour and communication survey 2008: A turning tide among teenagers? URL: <http://www.mrc.ac.za/pressreleases/2009/sanat.pdf> [accessed 2013-05-23] [WebCite Cache ID 6GpUfTZxF]

6. Hussain A, Moodley D, Naidoo S, Esterhuizen TM. Pregnant women's access to PMTCT and ART services in South Africa and implications for universal antiretroviral treatment. *PLoS One* 2011;6(12):e27907 [FREE Full text] [doi: [10.1371/journal.pone.0027907](https://doi.org/10.1371/journal.pone.0027907)] [Medline: [22162993](https://pubmed.ncbi.nlm.nih.gov/22162993/)]
7. Karim QA, Kharsany AB, Frohlich JA, Werner L, Mashego M, Mlotshwa M, et al. Stabilizing HIV prevalence masks high HIV incidence rates amongst rural and urban women in KwaZulu-Natal, South Africa. *Int J Epidemiol* 2011 Aug;40(4):922-930 [FREE Full text] [doi: [10.1093/ije/dyq176](https://doi.org/10.1093/ije/dyq176)] [Medline: [21047913](https://pubmed.ncbi.nlm.nih.gov/21047913/)]
8. Bwirire LD, Fitzgerald M, Zachariah R, Chikafa V, Massaquoi M, Moens M, et al. Reasons for loss to follow-up among mothers registered in a prevention-of-mother-to-child transmission program in rural Malawi. *Trans R Soc Trop Med Hyg* 2008 Dec;102(12):1195-1200. [doi: [10.1016/j.trstmh.2008.04.002](https://doi.org/10.1016/j.trstmh.2008.04.002)] [Medline: [18485431](https://pubmed.ncbi.nlm.nih.gov/18485431/)]
9. Doherty T, Sanders D, Goga A, Jackson D. Implications of the new WHO guidelines on HIV and infant feeding for child survival in South Africa. *Bull World Health Organ* 2011 Jan 1;89(1):62-67 [FREE Full text] [doi: [10.2471/BLT.10.079798](https://doi.org/10.2471/BLT.10.079798)] [Medline: [21346892](https://pubmed.ncbi.nlm.nih.gov/21346892/)]
10. Towards universal access: Scaling up priority HIV/AIDS interventions in the health sector: progress report 2009. Geneva, Switzerland: World Health Organization; 2009. URL: http://www.who.int/entity/hiv/pub/tuapr_2009_en.pdf [accessed 2013-05-23] [WebCite Cache ID 6GpV15Ebd]
11. Rollins N, Little K, Mzolo S, Horwood C, Newell ML. Surveillance of mother-to-child transmission prevention programmes at immunization clinics: the case for universal screening. *AIDS* 2007 Jun 19;21(10):1341-1347. [doi: [10.1097/QAD.0b013e32814db7d4](https://doi.org/10.1097/QAD.0b013e32814db7d4)] [Medline: [17545711](https://pubmed.ncbi.nlm.nih.gov/17545711/)]
12. Bärnighausen T, Bloom DE, Humair S. Universal antiretroviral treatment: the challenge of human resources. *Bull World Health Organ* 2010 Dec 1;88(12):951-952 [FREE Full text] [doi: [10.2471/BLT.09.073890](https://doi.org/10.2471/BLT.09.073890)] [Medline: [21124722](https://pubmed.ncbi.nlm.nih.gov/21124722/)]
13. Garrib A, Stoops N, McKenzie A, Dlamini L, Govender T, Rohde J, et al. An evaluation of the District Health Information System in rural South Africa. *S Afr Med J* 2008 Jul;98(7):549-552. [Medline: [18785397](https://pubmed.ncbi.nlm.nih.gov/18785397/)]
14. Sherman GG, Jones SA, Coovadia AH, Urban MF, Bolton KD. PMTCT from research to reality-results from a routine service. *S Afr Med J* 2004 Apr;94(4):289-292 [FREE Full text] [Medline: [15150944](https://pubmed.ncbi.nlm.nih.gov/15150944/)]
15. Mate KS, Bennett B, Mphatswe W, Barker P, Rollins N. Challenges for routine health system data management in a large public programme to prevent mother-to-child HIV transmission in South Africa. *PLoS One* 2009;4(5):e5483 [FREE Full text] [doi: [10.1371/journal.pone.0005483](https://doi.org/10.1371/journal.pone.0005483)] [Medline: [19434234](https://pubmed.ncbi.nlm.nih.gov/19434234/)]
16. Wringe A, Cataldo F, Stevenson N, Fakoya A. Delivering comprehensive home-based care programmes for HIV: a review of lessons learned and challenges ahead in the era of antiretroviral therapy. *Health Policy Plan* 2010 Sep;25(5):352-362 [FREE Full text] [doi: [10.1093/heapol/czq005](https://doi.org/10.1093/heapol/czq005)] [Medline: [20144935](https://pubmed.ncbi.nlm.nih.gov/20144935/)]
17. mHealth: new horizons for health through mobile technologies. Geneva, Switzerland: World Health Organization; 2011. URL: http://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2013-05-23] [WebCite Cache ID 6GpVKcbgO]
18. Zhang S, Wu Q, van Velthoven MH, Chen L, Car J, Rudan I, et al. Smartphone versus pen-and-paper data collection of infant feeding practices in rural China. *J Med Internet Res* 2012;14(5):e119 [FREE Full text] [doi: [10.2196/jmir.2183](https://doi.org/10.2196/jmir.2183)] [Medline: [22989894](https://pubmed.ncbi.nlm.nih.gov/22989894/)]
19. Crankshaw T, Corless IB, Giddy J, Nicholas PK, Eichbaum Q, Butler LM. Exploring the patterns of use and the feasibility of using cellular phones for clinic appointment reminders and adherence messages in an antiretroviral treatment clinic, Durban, South Africa. *AIDS Patient Care STDS* 2010 Nov;24(11):729-734. [doi: [10.1089/apc.2010.0146](https://doi.org/10.1089/apc.2010.0146)] [Medline: [21039181](https://pubmed.ncbi.nlm.nih.gov/21039181/)]
20. Tomlinson M, Solomon W, Singh Y, Doherty T, Chopra M, Ijumba P, et al. The use of mobile phones as a data collection tool: a report from a household survey in South Africa. *BMC Med Inform Decis Mak* 2009;9:51 [FREE Full text] [doi: [10.1186/1472-6947-9-51](https://doi.org/10.1186/1472-6947-9-51)] [Medline: [20030813](https://pubmed.ncbi.nlm.nih.gov/20030813/)]
21. Derenzi B, Borriello G, Jackson J, Kumar VS, Parikh TS, Virk P, et al. Mobile phone tools for field-based health care workers in low-income countries. *Mt Sinai J Med* 2011;78(3):406-418. [doi: [10.1002/msj.20256](https://doi.org/10.1002/msj.20256)] [Medline: [21598267](https://pubmed.ncbi.nlm.nih.gov/21598267/)]
22. Patnaik S, Brunskill E, Thies W. Evaluating the Accuracy of Data Collection on Mobile Phones: A Study of Forms, SMS, and Voice. 2009. URL: <http://research.microsoft.com/pubs/137974/patnaik-ictd09.pdf> [accessed 2013-05-24] [WebCite Cache ID 6Gr7Khy9N]
23. Rotheram-Borus MJ, Richter L, Van Rooyen H, van Heerden A, Tomlinson M, Stein A, et al. Project Masihambisane: a cluster randomised controlled trial with peer mentors to improve outcomes for pregnant mothers living with HIV. *Trials* 2011;12:2 [FREE Full text] [doi: [10.1186/1745-6215-12-2](https://doi.org/10.1186/1745-6215-12-2)] [Medline: [21205302](https://pubmed.ncbi.nlm.nih.gov/21205302/)]
24. Mobenzi Researcher. URL: <http://www.mobenzi.com/> [accessed 2013-05-23] [WebCite Cache ID 6GpWC2ji5]
25. Davis F. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly* 1989 Sep;3:19-239.
26. Jacobs E, Punt C, Buthelezi D. Western Cape Government of South Africa: Agriculture. Elsenburg, South Africa: Provincial Decision-Making Enabling (PROVIDE) Project; 2009. A Profile of the KwaZulu-Natal Province: Demographics, Poverty, Income, Inequality and Unemployment from 2000 till 2007 URL: http://www.elsenburg.com/provide/documents/BP2005_1_5%20Demographics%20KZ.pdf [accessed 2013-05-23] [WebCite Cache ID 6GpW1fV71]
27. Alam M, Khanam T, Khan R. Assessing the scope for use of mobile based solution to improve maternal and child health in Bangladesh: A case study. 2010 Presented at: Proceedings of the 4th International Conference on Information and

Communication Technologies and Development; Dec 14-15, 2010; London, England p. 14-15 URL: <http://www.gg.rhul.ac.uk/ict4d/ictd2010/posters/ICTD2010%20Alam%20et%20al.pdf>

Abbreviations

ANOVA: analysis of variance

API: application programming interface

HIV: human immunodeficiency virus

MLH: mothers living with HIV

MPAPI: mobile phone-assisted personal interview

PMTCT: Prevention of Maternal-to-Child Transmission

Edited by G Eysenbach; submitted 27.06.12; peer-reviewed by W Brown; comments to author 06.10.12; revised version received 21.12.12; accepted 13.01.13; published 10.06.13.

Please cite as:

van Heerden A, Norris S, Tollman S, Richter L, Rotheram-Borus MJ

Collecting Maternal Health Information From HIV-Positive Pregnant Women Using Mobile Phone-Assisted Face-to-Face Interviews in Southern Africa

J Med Internet Res 2013;15(6):e116

URL: <http://www.jmir.org/2013/6/e116/>

doi: [10.2196/jmir.2207](https://doi.org/10.2196/jmir.2207)

PMID: [23748182](https://pubmed.ncbi.nlm.nih.gov/23748182/)

©Alastair van Heerden, Shane Norris, Stephen Tollman, Linda Richter, Mary Jane Rotheram-Borus. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 10.06.2013. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Comparison of Physical Activity Measures Using Mobile Phone-Based CalFit and Actigraph

David Donaire-Gonzalez^{1,2,3,4}, MSc; Audrey de Nazelle^{1,2,3,5}, PhD; Edmund Seto⁶, PhD; Michelle Mendez^{1,2,3,7}, PhD; Mark J Nieuwenhuijsen^{1,2,3}, PhD; Michael Jerrett⁶, PhD

¹Center for Research in Environmental Epidemiology (CREAL), Barcelona, Catalonia, Spain

²Hospital del Mar Research Institute (IMIM), Barcelona, Catalonia, Spain

³Spanish Consortium for Research on Epidemiology and Public Health (CIBERESP), Barcelona, Catalonia, Spain

⁴Physical Activity and Sports Sciences Department, Fundació Blanquerna, Ramon Llull University, Barcelona, Catalonia, Spain

⁵Center for Environmental Policy, Imperial College London, London, United Kingdom

⁶Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, CA, United States

⁷Department of Nutrition, University of North Carolina, Chapel Hill, NC, United States

Corresponding Author:

Mark J Nieuwenhuijsen, PhD

Center for Research in Environmental Epidemiology (CREAL)

C Doctor Aiguader, 88.

Barcelona, Catalonia, 08003

Spain

Phone: 34 93 2147337

Fax: 34 93 2147330

Email: mnieuwenhuijsen@creal.cat

Abstract

Background: Epidemiological studies on physical activity often lack inexpensive, objective, valid, and reproducible tools for measuring physical activity levels of participants. Novel sensing technologies built into smartphones offer the potential to fill this gap.

Objective: We sought to validate estimates of physical activity and determine the usability for large population-based studies of the smartphone-based CalFit software.

Methods: A sample of 36 participants from Barcelona, Spain, wore a smartphone with CalFit software and an Actigraph GT3X accelerometer for 5 days. The ease of use (usability) and physical activity measures from both devices were compared, including vertical axis counts (VT) and duration and energy expenditure predictions for light, moderate, and vigorous intensity from Freedson's algorithm. Statistical analyses included (1) Kruskal-Wallis rank sum test for usability measures, (2) Spearman correlation and linear regression for VT counts, (3) concordance correlation coefficient (CCC), and (4) Bland-Altman plots for duration and energy expenditure measures.

Results: Approximately 64% (23/36) of participants were women. Mean age was 31 years (SD 8) and mean body mass index was 22 kg/m² (SD 2). In total, 25/36 (69%) participants recorded at least 3 days with at least 10 recorded hours of physical activity using CalFit. The linear association and correlations for VT counts were high (adjusted $R^2=0.85$; correlation coefficient .932, 95% CI 0.931-0.933). CCCs showed high agreement for duration and energy expenditure measures (from 0.83 to 0.91).

Conclusions: The CalFit system had lower usability than the Actigraph GT3X because the application lacked a means to turn itself on each time the smartphone was powered on. The CalFit system may provide valid estimates to quantify and classify physical activity. CalFit may prove to be more cost-effective and easily deployed for large-scale population health studies than other specialized instruments because cell phones are already carried by many people.

(*J Med Internet Res* 2013;15(6):e111) doi:[10.2196/jmir.2470](https://doi.org/10.2196/jmir.2470)

KEYWORDS

cellular phone; accelerometry; global positioning systems; motor activity; monitoring; physiologic

Introduction

Physical inactivity now ranks as the tenth leading cause of premature mortality worldwide [1,2]. Inactivity has increased substantially over the past 15 years [2]. Physical inactivity contributes to the development of major chronic diseases, such as coronary heart disease, stroke, hypertension, colon cancer, breast cancer, Type 2 diabetes, and osteoporosis [3].

Information on physical activity in epidemiological studies is generally obtained by questionnaires and more recently with accelerometers [4]. The latter is becoming the accepted method because of better accuracy and reliability of the physical activity measures [4]. Accelerometers use the acceleration in the subjects' movements to quantify intensity over short epochs of time. Although an improvement over questionnaires, deploying accelerometers is labor intensive and burdensome to the participant at times and may lead to potential changes in behavior, such as not wearing the accelerometer or increasing the measured behavior [5,6].

To address these problems and take advantage of the increased use and improved technology of smartphones, we developed CalFit [7-10]. CalFit is open-source software that runs on Android smartphones. The system makes use of the accelerometry and Global Positioning System (GPS) sensors that are built into smartphones to record physical activity and the time and location in which an activity occurs. It has the potential to reduce cost and allow for enrollment of more participants because smartphones are now in widespread use in the general population [11]. Smartphones equipped with CalFit could potentially make better physical activity measurements compared to a common accelerometer, particularly because of the addition of GPS measurements that can help researchers better understand the spatial context of activity [12]. Calibration and validation work of CalFit has been conducted thus far only under laboratory conditions [13].

The aim of this research is to study the usability of CalFit software and to assess the validity of its physical activity measures in real world situations by comparing its physical activity measures under free-living conditions with those obtained from a well-known and validated accelerometer, the Actigraph GT3X [14].

Methods

Sample

We enrolled volunteers to wear the CalFit phone and a conventional accelerometer for 5 days. Thirty-six participants were recruited by way of emails sent to colleagues from the Centre for Research in Environmental Epidemiology (CREAL) and to friends of colleagues as part of a larger study based on active travel behaviors. Inclusion criteria were to live and work or study in Barcelona, to live more than 10 minutes walking distance from the workplace or school, and be able to ride a bike for at least 20 minutes. Volunteers who met the eligibility requirements were enrolled in the study after an information session in which they were provided with details on study objectives and procedures. The field study took place from November 2011 to February 2012.

Our study protocol was approved by the Ethics Committee of Hospital del Mar Research Institute, and written informed consent was obtained from all the participants.

Instruments

Each participant was given an Actigraph GT3X accelerometer [15] and a smartphone fitted with the CalFit application (see Table 1). The devices were worn during 5 consecutive days on a belt attached to the waist (see Figure 1). Participants were instructed to remove devices only when performing aquatic activities or sleeping, or when necessary to charge the smartphone battery.

Table 1. Characteristics of CalFit and Actigraph GT3X.

Characteristics	Google G1 with CalFit	Actigraph GT3X
Size	11.7×5.6×1.7 cm	3.8×3.7×1.8 cm
Weight	158 g	27 g
Placement	Frontal mean point between both anterior superior iliac spines	Anterior superior iliac spine of the right hip
Sample rate	10 Hz	30 Hz
Data storage	16 GB	16 MB
Battery life	18 hours	31 days
Accelerometer sensor	AK8976A triaxial accelerometer (Asahi Kasei Microsystems, Japan)	ADXL335 triaxial accelerometer (Analog Devices, Norwood, MA)
Registered range of acceleration	±2.8g	±3g
Outcomes (measured)	Acceleration of the 3 axes	Acceleration of the 3 axes
Outcomes (estimated)	Not wearing; energy expenditure and duration of physical activity	Not wearing; standing, sitting, and lying; energy expenditure and duration of physical activity

Data Treatment

Data from both devices were summarized to 1-minute intervals. We merged data streams from both accelerometers identifying the time alignment that yielded the highest association (adjusted R^2) between the 2 vertical (VT) axis measures, within a maximum of 5-minute differences in time. To maximize the comparability [16], the intensity of physical activity measured in metabolic equivalents (METs) by both devices was calculated according to the equation of Freedson et al [17], which uses a linear function based on vertical axis counts to produce their estimates: ActiGraph GT3X METs = $1.439008 + (0.000795 * VT \text{ counts/min})$. Because VT axis measures were recorded by each instrument with different units (counts from ActiGraph GT3X versus g-force from CalFit), we first developed a linear regression between the 2 vertical measures to convert the CalFit g-force/min into counts/min, leading to the following adaptation to the Freedson equation for estimating METs from smartphone data: CalFit METs = $1.2907087 + (0.4141791 * VT \text{ g/min})$.

The accelerometer nonwear intervals were defined as episodes of at least 40 consecutive minutes of 0 counts and below 0.3g

in vertical axis for ActiGraph GT3X and CalFit, respectively. The latter threshold was established after analyzing CalFit nighttime measurements. The American College of Sports Medicine considers having at least 3 days with at least 10 hours of recorded activity as a valid assessment of physical activity [18]. We measured the usability of CalFit and of ActiGraph GT3X, understood as the ease of use to reach valid assessment of physical activity, in 4 ways: (1) number of subjects with valid assessments of physical activity (previous definition); (2) number of recorded days per participant (ie, reflecting participants' ability to keep CalFit turned on); (3) total recorded time per participant (ie, reflecting participants' ability to keep batteries charged); and (4) percent wearing time from total recorded time (ie, reflecting participants' ability to wear CalFit).

Physical activity was defined as any minute with intensity equal or greater than 1.5 METs. Physical activity was partitioned into light, moderate, and vigorous levels of physical activity following the conventional cutoffs of 3 and 6 METs. The main summary measures of physical activity were vertical axis counts, and duration and intensity of physical activity.

Figure 1. Set of devices that were worn during 5 consecutive days.



Analysis

Participants and physical activity characteristics are presented as number (percentage) for categorical variables, mean (SD) for continuous variables with normal distribution, or median

(interquartile range, IQR) for continuous variables with non-normal distribution.

The comparison between CalFit and ActiGraph GT3X was conducted using several approaches. First, to assess differences on usability as defined above, we performed a Kruskal-Wallis

rank sum test (difference of medians components of usability). Second, the correlation and association between the vertical axis measures during coinciding time periods were assessed through a Spearman correlation and linear regression, respectively. Third, the agreement in the main summary measures of physical activity, as previously defined and during coinciding time periods, was studied using Lin's concordance correlation coefficient (CCC) [19] and Bland-Altman plots. The CCC can be conceptualized as the ratio of between-subject variance to total variance [20]. In other words, it provides a measure of the percentage of differences attributable to the participants, and its complement (1-CCC) gives the percentage of differences attributable to the method (ie, CalFit vs Actigraph GT3X). The bias between instruments was evaluated using a linear regression analysis between the differences (CalFit-Actigraph GT3X) and the mean, as $0.5 \times (\text{CalFit} - \text{Actigraph GT3X})$, of the 2 physical activity measures, considering the bias to be significant when the confidence interval of the coefficient did not contain the value zero. Both regression coefficient and regression line of bias were also plotted into Bland-Altman plots indicated with red letter and line, respectively.

As a sensitivity analysis, previous comparisons were also performed during coinciding days with at least 10 hours of recorded activity, without control of the coinciding time periods, to test the influence of nonmeasured periods on physical activity

agreement. All analyses were conducted using R-2.14.1 2011 (The R Foundation for Statistical Computing).

Results

The sample consisted of 36 participants, most of which were women (23/36, 64%), with a mean age of 30.9 years (SD 7.9), and mean body mass index of 22.2 kg/m^2 (SD 2.4) (Table 2). Approximately 83% (30/36) were of Spanish nationality, 92% (33/36) had high school or greater education, and 50% (18/36) earned more than €2000 per month.

CalFit Usability

Of 180 possible days for recorded data, 19 were missing from CalFit and 8 from the Actigraph GT3X. During recorded days, there was a significant difference between the median time recorded: 22 hours for CalFit and 24 hours for Actigraph GT3X (Table 3). Also, there were differences for the percent of wear time between CalFit and Actigraph GT3X (52% vs 59%) (Table 3). The median number of days with at least 10 wearing hours was 3 and 5 for CalFit and Actigraph GT3X, respectively.

The main reasons for failed CalFit data collection among the 11 subjects who recorded less than 3 valid assessment of physical activity were: (1) 6 lost an average of 2 days of recording because CalFit was inadvertently turned off, (2) 2 had problems with phone battery life and their daily routine, and (3) 3 did not wear the phone.

Table 2. Sociodemographic and physical characteristics of all participants (N=36).

Sample characteristics	Participants
Age (years), mean (SD)	31 (8)
Gender, n (%)	
Male	13 (36)
BMI (kg/m^2), mean (SD)	22 (2)
Educational level, n (%)	
More than high school	33 (92)
Less than high school	3 (8)
Nationality, n (%)	
Spanish	30 (83)
Others	6 (17)
Monthly income (€)	
More than 2000	18 (50)
Less than 2000	18 (50)
Working status, n (%)	
Working	32 (89)
Studying	4 (11)

Table 3. Comparison of usability characteristics between Actigraph and CalFit.

Characteristic	Actigraph GT3X	CalFit	P value
Days recorded (day), median (IQR)	5 (5-5)	5 (4.8-5.0)	.03
Recorded time (min), median (IQR)	7200 (7200-7200)	6474 (4635-7068)	<.001
Wearing time (min), median (IQR)	4109 (3735-4373)	2938 (2269-3652)	<.001
Time coincident (min), median (IQR)	2825 (2110-3556)		
Recorded time within recorded days (hour/day), median (IQR)	24 (24-24)	22 (20-24)	<.001
Worn time within recorded days (hour/day), median (IQR)	14 (12.5-15)	11 (10-13)	<.001
Percent of worn time on recorded time within recorded days (%), median (IQR)	58.5 (53-63)	51.6 (46-58)	.03
Number of days with at least 10 wearing hours (day), median (IQR)	5 (4-5)	3 (2-4.2)	<.001
Participants with valid assessment of physical activity, n (%)	34 (94)	25 (69)	<.001

Validity of Physical Activity

The linear regression and correlation analysis for average vertical (VT) axis measures from both devices during coinciding wear-time periods showed a high association (adjusted $R^2=0.85$; Spearman correlation coefficient .932, 95% CI 0.931-0.933) (Figure 2, part A). During coinciding time periods (mean time/day, 2600 min), the mean difference between Actigraph GT3X and CalFit for the duration of active time (>1.5 METs) was 2.24% (95% CI 0.76-3.72) and for intensity of physical activity was 0.07 METs (95% CI 0.04-0.1) (Figures 2, parts B and C). The comparison for both the duration and intensity of physical activity showed that the variability attributable to the measurement method (which is the complementary to the ratio of between-subject variance to total variance) was less than 20% (Figures 2, parts B and C). There was no association between difference and average of both measures neither for duration ($P=.55$) nor for intensity ($P=.22$) of physical activity.

Validity of Physical Activity Through the Different Intensity Thresholds

The comparison of measures of light, moderate, and vigorous physical activity showed that less than 30% of the variability

was attributable to the method of measurement (Figure 3). In contrast to light and moderate physical activity, the CalFit measures of vigorous physical activity showed a tendency to underestimate the duration in vigorous physical activity as activity levels increased ($P=.01$) compared to the Actigraph GT3X measure (Figure 3, part C). Figure 4 shows there was a significant underestimation in the intensity recorded by CalFit when participants performed vigorous activity according to Actigraph GT3X (CalFit: mean 5.9, SD 1.0; Actigraph: mean 7.1, SD 1.1; $P<.001$).

Sensitivity to Measurement Period

Depending on the time inclusion criteria selected, the average difference between Actigraph GT3X and CalFit during light physical activity changed from a small but significant overestimation of 1.7% (95% CI 0.4-3.1) for coinciding time periods to a nonsignificant underestimation of -11.5 min (95% CI -27 to 4.3) for coinciding valid days (Figure 3, part A vs Figure 5, part A). There were no differences in duration of moderate and vigorous physical activity across time inclusion criteria (Figures 3, parts B and C vs Figures 5, parts B and C).

Figure 2. Agreement between CalFit and Actigraph GT3X in vertical axis, duration, and energy expenditure in physical activity within the coinciding measurement time periods. (A) accelerometer vertical axis measures, (B) duration in physical activity, and (C) intensity of physical activity.

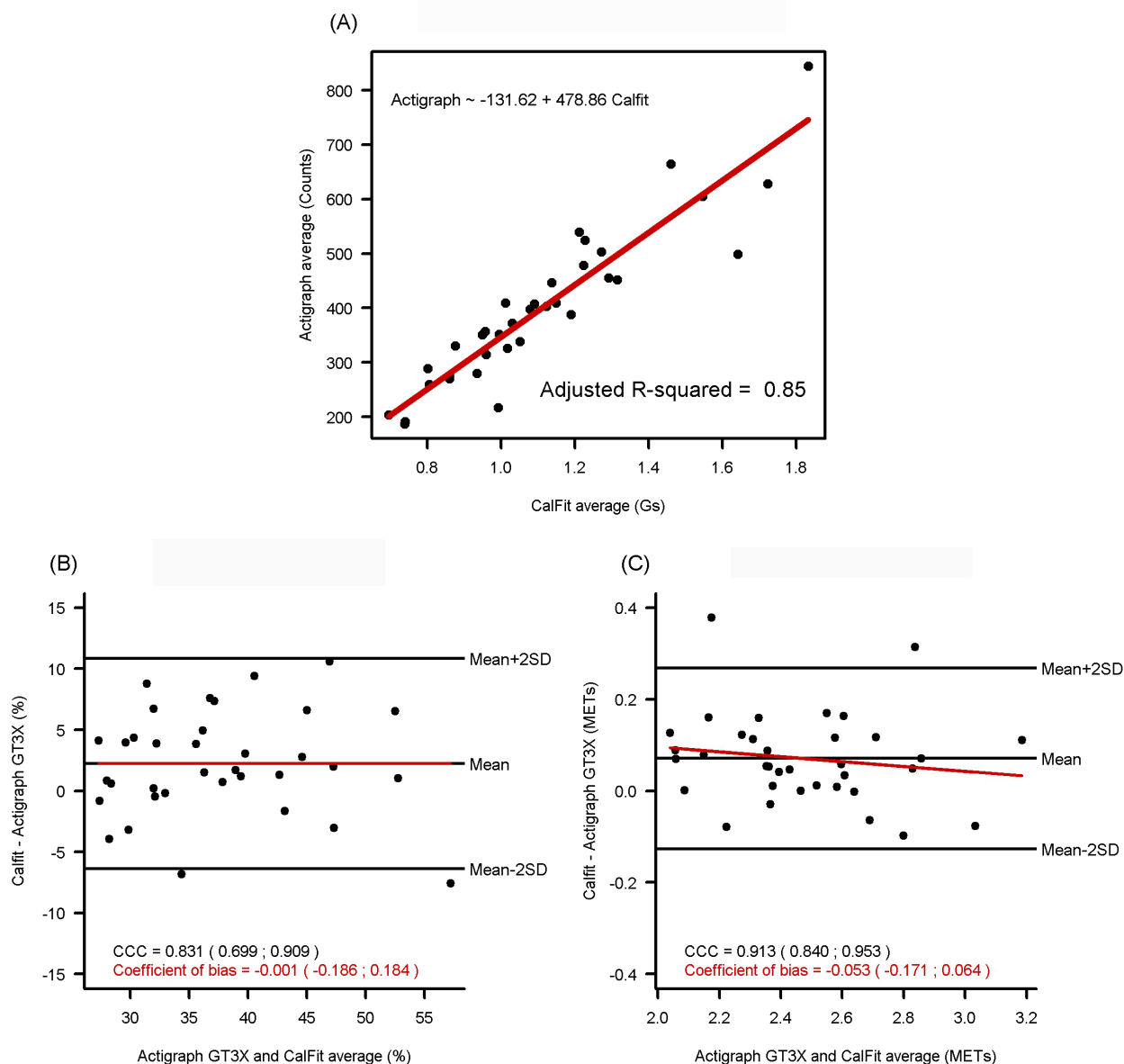


Figure 3. Agreement between CalFit and Actigraph GT3X for duration of light, moderate, and vigorous physical activity within the coinciding measurement time periods. (A) duration of light physical activity, (B) duration of moderate physical activity, and (C) duration of vigorous physical activity.

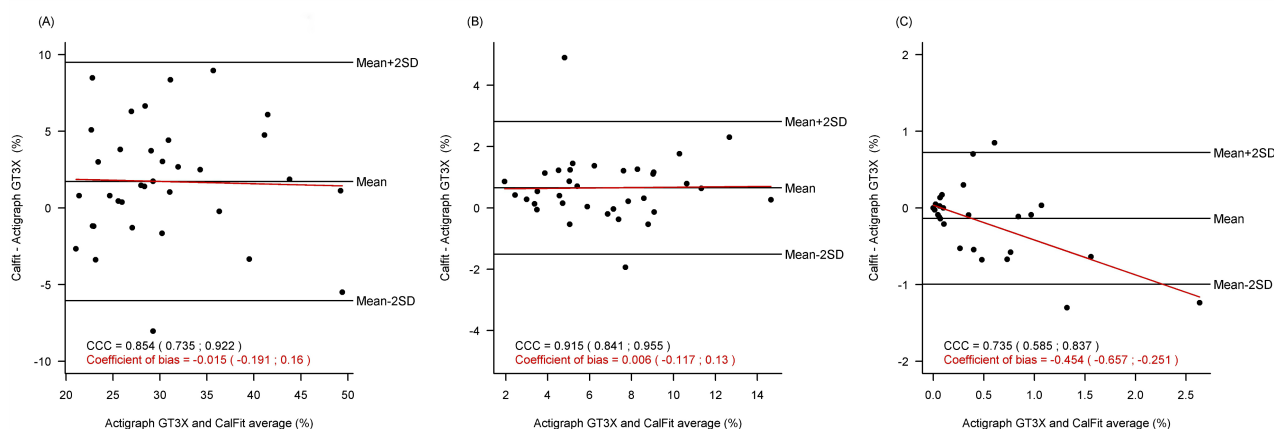


Figure 4. Comparison of average intensity recorded by CalFit and Actigraph GT3X within light, moderate, and vigorous physical activity identified by Actigraph.

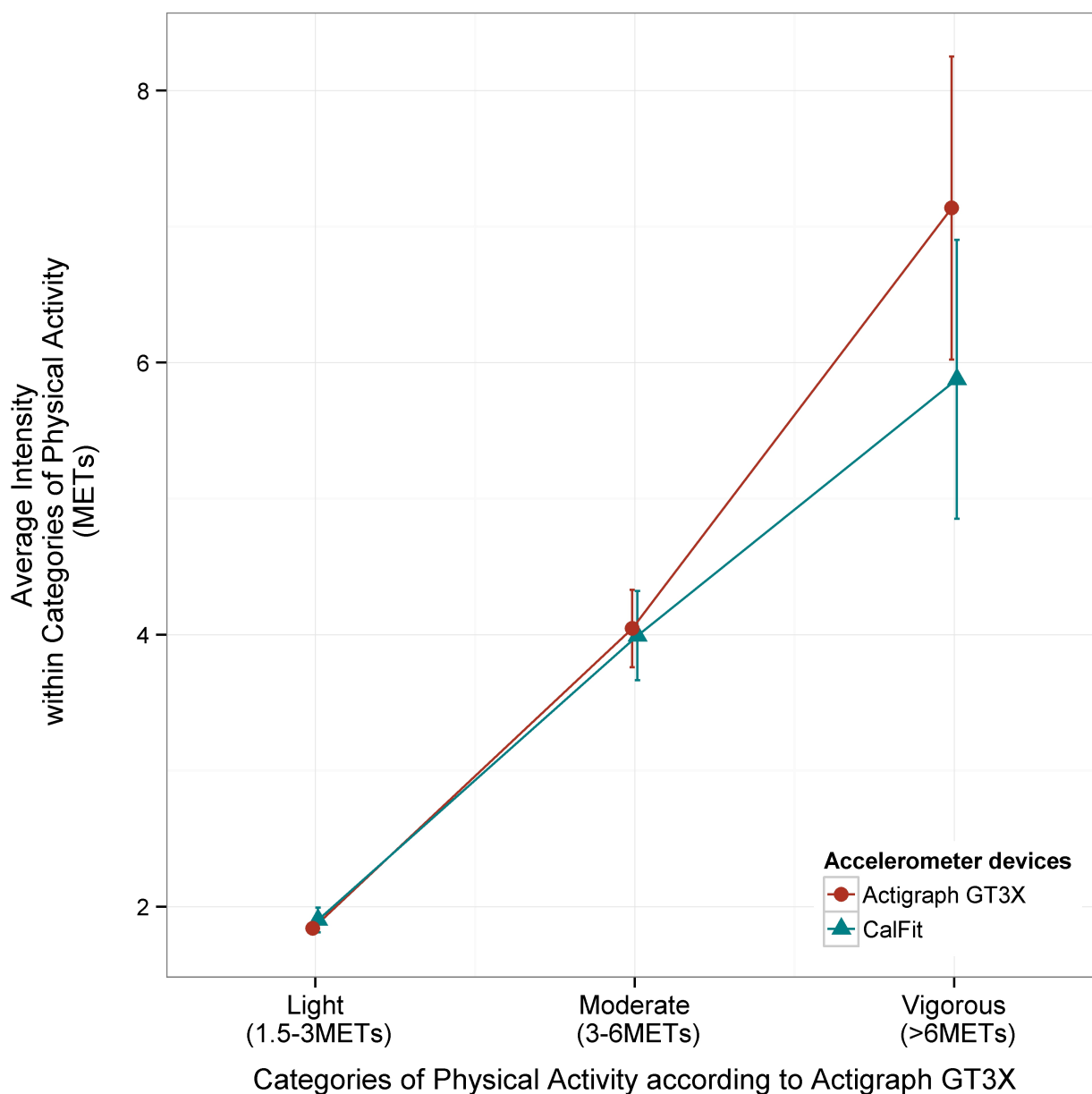
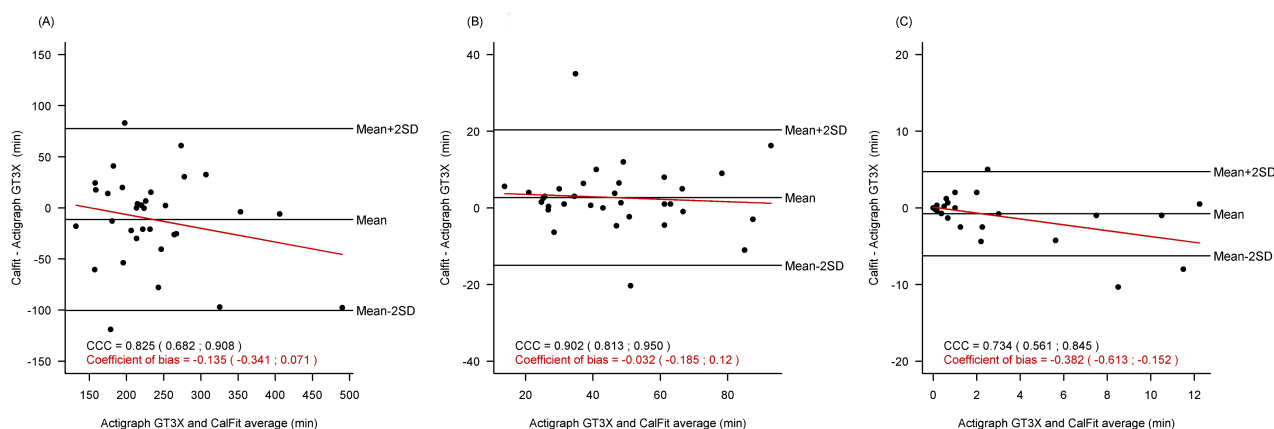


Figure 5. Agreement between CalFit and Actigraph GT3X during light, moderate, and vigorous physical activity within the coinciding days with at least 10 hours of recorded activity. (A) duration of light physical activity per day, (B) duration of moderate physical activity per day, and (C) duration of vigorous physical activity per day.



Discussion

Principal Results

This study assessed the usability and validity of CalFit software in a group of free-living volunteers. We compared CalFit to physical activity measures with those obtained from the Actigraph GT3X. The several approaches used to assess the properties of the CalFit showed that (1) there is a strong association between vertical axis measures from both devices; (2) the measures of duration and energy expenditure in overall, light, and moderate physical activity were highly concordant between devices, whereas vigorous physical activity was underestimated; (3) CalFit had lower usability compared to Actigraph GT3X resulting in a lower proportion of participants with valid assessment of physical activity; and (4) sensitivity analysis that compared the agreement within coinciding time periods to the agreement within coinciding days with at least 10 hours of recorded activity showed that the disparities in wearing-time periods between devices did not contribute to any significant bias into the measured validity.

Comparison With Prior Work

To our knowledge, this is the first study to compare accelerometer use on smartphones to measure physical activity with a currently well-validated instrument [12]. Previous research on physical activity assessment with mobile phones has shown that they are a useful tool to perform interventions [21] and are helpful for activity recognition [22-24]. In addition to these advantages, using smartphones for physical activity research opens up opportunities for reaching large numbers of participants at a relatively low cost [9]. The acceptance and usability of smartphones to measure physical activity on free-living conditions was previously unknown. Here, we showed that 25 of 36 (69%) participants who used CalFit recorded at least 3 valid days, which is the minimum recommended to assess daily physical activity [25]. The greatest weakness in CalFit usability was loss of data because of the phone turning off and not having CalFit restart when the phone was powered back on (50% of the missing data). In the current version of CalFit, this problem has been solved by automatically restarting CalFit each time the phone is turned on. The second weakness was the battery life (recorded time and wearing time), which was responsible for the other half of the missing data. The difference in wearing time with the accelerometer was partly because of participants having to charge the smartphone during waking hours (as they were instructed to do). Since conducting our study, we have found that newer generation smartphones have improved battery life, and current field tests indicate that CalFit is recording for longer durations without charging [26].

This is also the first study to compare the validity of the vertical axis measures and to use the same algorithm for estimating physical activity in 2 different instruments. The association of the vertical axis measures between the 2 tools was high (adjusted R^2 0.86; correlation coefficient .932; 95% CI 0.931-0.933), which is within the range of the literature comparing different models of Actigraph [27]. The concordance found in duration and energy expenditure in physical activity measures of CalFit

and Actigraph GT3X showed that the measures from CalFit and Actigraph GT3X are interchangeable (less than 20% of the variability is attributable to differences between instruments).

Concordance in physical activity measures across different definitions of time inclusion criteria showed that the results remained constant despite the shorter wearing time of CalFit. This suggests that the time charging the smartphone or the shorter battery life did not have a significant influence on the final measures. There was also a statistically significant bias toward underestimation in measures of vigorous physical activity estimated by CalFit compared to Actigraph GT3X. This may be partially explained by the fact that we used average VT measures instead of all measures per minute per participant and because we assumed a linear relationship between acceleration forces from smartphone and counts from Actigraph GT3X.

Strengths

One of the main strengths of this study is the use and testing under free-living conditions. Participants maintained their daily routines, which is difficult to replicate in controlled environments. Another strength was the use of the concordance measures for quantifying physical activity in addition to the commonly used correlations. A third was using the Freedson algorithm of physical activity for both instruments, which is a valid algorithm for the different Actigraph models (CSA 1764, GT1M, and GT3X) [28] that maximized the comparability between the instruments [16]. Furthermore, because we used a first-generation smartphone, our findings can be generalized and expected to be better for the latter generations of smartphones as a consequence of hardware evolution. Also, future versions on CalFit will be developed for Android and iPhone platforms.

The validation of smartphone accelerometry-based energy expenditure has implications for both epidemiologic research on physical activity as well as for the growth of the practice of medicine and public health by mobile applications (mHealth applications). Beyond the current CalFit application, which is focused on unobtrusive sensing of physical activity, may be novel mHealth smartphone applications that not only record physical activity, but attempt to intervene upon behavior [11,21]. For example, future use of smartphones may allow for recognition of patterns of physical activity to better tailor interventions to personal baselines and goals. Additionally, future interventions may employ other aspects of smartphone technology (eg, call, text messaging, and Internet communication capabilities) to combine physical activity monitoring with motivational social interactivity [21]. There are many possibilities for creative uses of smartphones, and the research presented here provides a foundation for better understanding the energy expenditure estimates from this technology.

The CalFit smartphone system has several advantages over conventional accelerometers because of geolocation information both from cell phone towers and Wi-Fi networks and from GPS satellites. This geolocation will allow us to improve the current physical activity algorithm by including information such as velocity of displacements, topographical challenges faced by participants (stairs, slopes), and the environments (home, work)

where physical activity occurred. Furthermore, this tool also allows assessments of how the built and natural environment may affect behavior or lead to other exposures. Our research group has begun to demonstrate some of these advantages with the same participants by characterizing where the physical activity was done and quantifying the amount of pollution inhaled by participants in these environments [29].

Limitations

One limitation of the present study was the use of a convenience sample of 36 participants with a high educational level to assess CalFit usability. However, this design has been efficient in detecting the problems in usability. Further work needs to be conducted in the population at large. Second, the use of the Actigraph GT3X accelerometer as a gold standard could be seen as a limitation, but it is the reference tool for assessing physical activity in real life for 5 days and has well-established validity [30]. The use of an algorithm for the METs estimation that only takes into account the vertical axis of the accelerometers is a

limitation that we could not avoid because there are currently no published Actigraph GT3X algorithms using the 3 axes of the accelerometer [31,32]. Finally, our definition of CalFit wearing time was an operational definition and should not be considered as a reference until it is tested in studies specifically designed for this purpose.

Conclusions

Compared to the current gold standard instrument for population studies, the smartphones fitted with CalFit supply useful and valid estimates for quantifying and classifying physical activity under free-living conditions. Although user compliance for CalFit was lower than with the Actigraph GT3X, this difference would likely diminish if participants were allowed to load CalFit onto their existing smartphones, which will be feasible in the future. Such deployment would provide a cost-effective approach for large epidemiological studies and mHealth applications that rely upon measured physical activity.

Acknowledgments

We thank Jaume Matamala and Meritxell Portella for the fieldwork done in this study. Funding for this project was provided by the Centre for Research in Environmental Epidemiology (CREAL) Internal Grant Program, the Coca-Cola Foundation through the Transportation Air pollution and Physical Activities (TAPAS) Program, and NIH NIEHS grant R01-ES020409. Finally, we thank Dr Ruzena Bajcsy and seed funding from the Center for Information Technology in the Interest of Society (CITRIS), which were instrumental in the development of CalFit.

Conflicts of Interest

None declared.

References

1. Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, Lancet Physical Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet* 2012 Jul 21;380(9838):219-229. [doi: [10.1016/S0140-6736\(12\)61031-9](https://doi.org/10.1016/S0140-6736(12)61031-9)] [Medline: [22818936](https://pubmed.ncbi.nlm.nih.gov/22818936/)]
2. World Health Organization. Global Health Risks: Mortality and Burden of Disease Attributable to Selected Major Risks. Switzerland: World Health Organization; 2009. URL: http://www.who.int/healthinfo/global_burden_disease/GlobalHealthRisks_report_full.pdf [accessed 2013-05-16] [WebCite Cache ID 6GfHnj9nS]
3. Department of Health, Physical Activity, Health Improvement and Prevention. 2004. At least five a week: Evidence on the impact of physical activity and its relationship to health URL: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4080994 [accessed 2012-12-02] [WebCite Cache ID 6CcIJ3NC4]
4. Prince SA, Adamo KB, Hamel ME, Hardt J, Gorber SC, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *Int J Behav Nutr Phys Act* 2008;5:56 [FREE Full text] [doi: [10.1186/1479-5868-5-56](https://doi.org/10.1186/1479-5868-5-56)] [Medline: [18990237](https://pubmed.ncbi.nlm.nih.gov/18990237/)]
5. Belton S, O'Brien W, Wickel EE, Issartel J. Patterns of non-compliance in adolescent field based accelerometer research. *J Phys Act Health* 2012 Nov 30;1-10 (forthcoming). [Medline: [23223684](https://pubmed.ncbi.nlm.nih.gov/23223684/)]
6. Clemes SA, Deans NK. Presence and duration of reactivity to pedometers in adults. *Med Sci Sports Exerc* 2012 Jun;44(6):1097-1101. [doi: [10.1249/MSS.0b013e318242a377](https://doi.org/10.1249/MSS.0b013e318242a377)] [Medline: [22595985](https://pubmed.ncbi.nlm.nih.gov/22595985/)]
7. Gravina R, Alessandro A, Salmeri A, Buondonno L, Raveendranathan N, Loseu V, et al. Enabling Multiple BSN Applications Using the SPINE Framework. In: Proceedings of the 2010 International Conference on Body Sensor Networks. USA: IEEE Computer Society; 2010 Presented at: Body Sensor Networks (BSN), 2010 International Conference on Body Sensor Networks; June 7-9, 2010; Washington, DC. [doi: [10.1109/BSN.2010.34](https://doi.org/10.1109/BSN.2010.34)]
8. Kuryloski P, Giani A, Giannantonio R, Gilani K, Gravina R, Seppa VP, et al. DexterNet: An Open Platform for Heterogeneous Body Sensor Networks and its Applications. USA: IEEE Computer Society; 2009 Presented at: Sixth International Workshop on Wearable and Implantable Body Sensor Networks, 2009. BSN 2009; June 3-5, 2009; Washington, DC. [doi: [10.1109/BSN.2009.31](https://doi.org/10.1109/BSN.2009.31)]
9. Seto E, Martin E, Yang A, Yan P, Gravina R, Lin I, et al. Opportunistic strategies for lightweight signal processing for body sensor networks. New York: ACM; 2010 Presented at: PETRA 2010 Proceedings of the 3rd International Conference

- on Pervasive Technologies Related to Assistive Environments; June 13-15, 2010; Samos, Greece p. 1-6. [doi: [10.1145/1839294.1839361](https://doi.org/10.1145/1839294.1839361)]
10. Yan P, Lin I, Roy M, Seto E, Wang C, Bajcsy R. WAVE and CalFit - Towards social interaction in mobile body sensor networks. 2010 Presented at: International Workshop on Ubiquitous Body Sensor Networks (USBN); Mar 1-3, 2010; Singapore, China. [doi: [10.4108/ICST.WICON2010.8602](https://doi.org/10.4108/ICST.WICON2010.8602)]
 11. Pratt M, Sarmiento OL, Montes F, Ogilvie D, Marcus BH, Perez LG, Lancet Physical Activity Series Working Group. The implications of megatrends in information and communication technology and transportation for changes in global physical activity. *Lancet* 2012 Jul 21;380(9838):282-293. [doi: [10.1016/S0140-6736\(12\)60736-3](https://doi.org/10.1016/S0140-6736(12)60736-3)] [Medline: [22818940](https://pubmed.ncbi.nlm.nih.gov/22818940/)]
 12. Rodríguez DA, Brown AL, Troped PJ. Portable global positioning units to complement accelerometry-based physical activity monitors. *Med Sci Sports Exerc* 2005 Nov;37(11 Suppl):S572-S581. [Medline: [16294120](https://pubmed.ncbi.nlm.nih.gov/16294120/)]
 13. Seto E, Yan P, Kuryloski P, Bajcsy R, Abresch T, Henricson E, et al. Mobile Phones as Personal Environmental Sensing Platforms: Development of the CalFit System. 2011 Presented at: The 23rd Annual Conference of the International Society of Environmental Epidemiology (ISEE); Sep 13-16, 2011; Barcelona, Spain.
 14. Van Remoortel H, Raste Y, Louvaris Z, Giavedoni S, Burtin C, Langer D, PROactive consortium. Validity of six activity monitors in chronic obstructive pulmonary disease: a comparison with indirect calorimetry. *PLoS One* 2012;7(6):e39198 [FREE Full text] [doi: [10.1371/journal.pone.0039198](https://doi.org/10.1371/journal.pone.0039198)] [Medline: [22745715](https://pubmed.ncbi.nlm.nih.gov/22745715/)]
 15. John D, Freedson P. ActiGraph and Actical physical activity monitors: a peek under the hood. *Med Sci Sports Exerc* 2012 Jan;44(1 Suppl 1):S86-S89 [FREE Full text] [doi: [10.1249/MSS.0b013e3182399f5e](https://doi.org/10.1249/MSS.0b013e3182399f5e)] [Medline: [22157779](https://pubmed.ncbi.nlm.nih.gov/22157779/)]
 16. Mâsse LC, Fuemmeler BF, Anderson CB, Matthews CE, Trost SG, Catellier DJ, et al. Accelerometer data reduction: a comparison of four reduction algorithms on select outcome variables. *Med Sci Sports Exerc* 2005 Nov;37(11 Suppl):S544-S554. [Medline: [16294117](https://pubmed.ncbi.nlm.nih.gov/16294117/)]
 17. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc* 1998 May;30(5):777-781. [Medline: [9588623](https://pubmed.ncbi.nlm.nih.gov/9588623/)]
 18. Heil DP, Brage S, Rothney MP. Modeling physical activity outcomes from wearable monitors. *Med Sci Sports Exerc* 2012 Jan;44(1 Suppl 1):S50-S60. [doi: [10.1249/MSS.0b013e3182399dce](https://doi.org/10.1249/MSS.0b013e3182399dce)] [Medline: [22157775](https://pubmed.ncbi.nlm.nih.gov/22157775/)]
 19. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989 Mar;45(1):255-268. [Medline: [2720055](https://pubmed.ncbi.nlm.nih.gov/2720055/)]
 20. Carrasco JL, Jover L. Estimating the generalized concordance correlation coefficient through variance components. *Biometrics* 2003 Dec;59(4):849-858. [Medline: [14969463](https://pubmed.ncbi.nlm.nih.gov/14969463/)]
 21. Fanning J, Mullen SP, McAuley E. Increasing physical activity with mobile devices: a meta-analysis. *J Med Internet Res* 2012;14(6):e161 [FREE Full text] [doi: [10.2196/jmir.2171](https://doi.org/10.2196/jmir.2171)] [Medline: [23171838](https://pubmed.ncbi.nlm.nih.gov/23171838/)]
 22. Lee MH, Kim J, Jee SH, Yoo SK. Integrated solution for physical activity monitoring based on mobile phone and PC. *Healthc Inform Res* 2011 Mar;17(1):76-86 [FREE Full text] [doi: [10.4258/hir.2011.17.1.76](https://doi.org/10.4258/hir.2011.17.1.76)] [Medline: [21818460](https://pubmed.ncbi.nlm.nih.gov/21818460/)]
 23. Xia Y, Cheung V, Garcia E, Ding H, Karunaithi M. Development of an automated physical activity classification application for mobile phones. *Stud Health Technol Inform* 2011;168:188-194. [Medline: [21893928](https://pubmed.ncbi.nlm.nih.gov/21893928/)]
 24. Wu W, Dasgupta S, Ramirez EE, Peterson C, Norman GJ. Classification accuracies of physical activities using smartphone motion sensors. *J Med Internet Res* 2012;14(5):e130 [FREE Full text] [doi: [10.2196/jmir.2208](https://doi.org/10.2196/jmir.2208)] [Medline: [23041431](https://pubmed.ncbi.nlm.nih.gov/23041431/)]
 25. Trost SG, McIver KL, Pate RR. Conducting accelerometer-based activity assessments in field-based research. *Med Sci Sports Exerc* 2005 Nov;37(11 Suppl):S531-S543. [Medline: [16294116](https://pubmed.ncbi.nlm.nih.gov/16294116/)]
 26. Eason E. Smartphone Battery Inadequacy. 2010 Dec 17. URL: <http://large.stanford.edu/courses/2010/ph240/eason1/> [accessed 2013-02-24] [WebCite Cache ID 6Efraplk]
 27. Bassett DR, Ainsworth BE, Swartz AM, Strath SJ, O'Brien WL, King GA. Validity of four motion sensors in measuring moderate intensity physical activity. *Med Sci Sports Exerc* 2000 Sep;32(9 Suppl):S471-S480. [Medline: [10993417](https://pubmed.ncbi.nlm.nih.gov/10993417/)]
 28. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. *J Sci Med Sport* 2011 Sep;14(5):411-416. [doi: [10.1016/j.jsams.2011.04.003](https://doi.org/10.1016/j.jsams.2011.04.003)] [Medline: [21616714](https://pubmed.ncbi.nlm.nih.gov/21616714/)]
 29. de Nazelle A, Seto E, Donaire-Gonzalez D, Mendez M, Matamala J, Nieuwenhuijsen MJ, et al. Improving estimates of air pollution exposure through ubiquitous sensing technologies. *Environ Pollut* 2013 May;176:92-99. [doi: [10.1016/j.envpol.2012.12.032](https://doi.org/10.1016/j.envpol.2012.12.032)] [Medline: [23416743](https://pubmed.ncbi.nlm.nih.gov/23416743/)]
 30. Lyden K, Kozey SL, Staudenmeyer JW, Freedson PS. A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. *Eur J Appl Physiol* 2011 Feb;111(2):187-201 [FREE Full text] [doi: [10.1007/s00421-010-1639-8](https://doi.org/10.1007/s00421-010-1639-8)] [Medline: [20842375](https://pubmed.ncbi.nlm.nih.gov/20842375/)]
 31. Swartz AM, Strath SJ, Bassett DR, O'Brien WL, King GA, Ainsworth BE. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Med Sci Sports Exerc* 2000 Sep;32(9 Suppl):S450-S456. [Medline: [10993414](https://pubmed.ncbi.nlm.nih.gov/10993414/)]
 32. Crouter SE, Kuffel E, Haas JD, Frongillo EA, Bassett DR. Refined two-regression model for the ActiGraph accelerometer. *Med Sci Sports Exerc* 2010 May;42(5):1029-1037 [FREE Full text] [doi: [10.1249/MSS.0b013e3181c37458](https://doi.org/10.1249/MSS.0b013e3181c37458)] [Medline: [20400882](https://pubmed.ncbi.nlm.nih.gov/20400882/)]

Abbreviations

CCC: concordance correlation coefficient

GPS: Global Positioning System

MET: Metabolic Equivalent of Task

VT: vertical

Edited by G Eysenbach; submitted 07.12.12; peer-reviewed by W Wu, H Van Remoortel; comments to author 22.12.12; revised version received 29.01.13; accepted 12.04.13; published 13.06.13.

Please cite as:

*Donaire-Gonzalez D, de Nazelle A, Seto E, Mendez M, Nieuwenhuijsen MJ, Jerrett M
Comparison of Physical Activity Measures Using Mobile Phone-Based CalFit and Actigraph
J Med Internet Res 2013;15(6):e111*

URL: <http://www.jmir.org/2013/6/e111/>

doi: [10.2196/jmir.2470](https://doi.org/10.2196/jmir.2470)

PMID: [23896156](https://pubmed.ncbi.nlm.nih.gov/23896156/)

©David Donaire-Gonzalez, Audrey de Nazelle, Edmund Seto, Michelle Mendez, Mark J Nieuwenhuijsen, Michael Jerrett. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 13.06.2013. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Review

Mobile Health Applications for the Most Prevalent Conditions by the World Health Organization: Review and Analysis

Borja Martínez-Pérez^{1*}, MSc; Isabel de la Torre-Díez^{1*}, PhD; Miguel López-Coronado^{1*}, PhD

University of Valladolid, Department of Signal Theory and Communications, and Telematics Engineering., University of Valladolid, Valladolid, Spain

* all authors contributed equally

Corresponding Author:

Borja Martínez-Pérez, MSc

University of Valladolid

Department of Signal Theory and Communications, and Telematics Engineering

University of Valladolid

Paseo de Belén, 15

Valladolid, 47011

Spain

Phone: 34 983423000 ext 3703

Fax: 34 983423667

Email: borja.martinez@uva.es

Abstract

Background: New possibilities for mHealth have arisen by means of the latest advances in mobile communications and technologies. With more than 1 billion smartphones and 100 million tablets around the world, these devices can be a valuable tool in health care management. Every aid for health care is welcome and necessary as shown by the more than 50 million estimated deaths caused by illnesses or health conditions in 2008. Some of these conditions have additional importance depending on their prevalence.

Objective: To study the existing applications for mobile devices exclusively dedicated to the eight most prevalent health conditions by the latest update (2004) of the Global Burden of Disease (GBD) of the World Health Organization (WHO): iron-deficiency anemia, hearing loss, migraine, low vision, asthma, diabetes mellitus, osteoarthritis (OA), and unipolar depressive disorders.

Methods: Two reviews have been carried out. The first one is a review of mobile applications in published articles retrieved from the following systems: IEEE Xplore, Scopus, ScienceDirect, Web of Knowledge, and PubMed. The second review is carried out by searching the most important commercial app stores: Google play, iTunes, BlackBerry World, Windows Phone Apps+Games, and Nokia's Ovi store. Finally, two applications for each condition, one for each review, were selected for an in-depth analysis.

Results: Search queries up to April 2013 located 247 papers and more than 3673 apps related to the most prevalent conditions. The conditions in descending order by the number of applications found in literature are diabetes, asthma, depression, hearing loss, low vision, OA, anemia, and migraine. However when ordered by the number of commercial apps found, the list is diabetes, depression, migraine, asthma, low vision, hearing loss, OA, and anemia. Excluding OA from the former list, the four most prevalent conditions have fewer apps and research than the final four. Several results are extracted from the in-depth analysis: most of the apps are designed for monitoring, assisting, or informing about the condition. Typically an Internet connection is not required, and most of the apps are aimed for the general public and for nonclinical use. The preferred type of data visualization is text followed by charts and pictures. Assistive and monitoring apps are shown to be frequently used, whereas informative and educational apps are only occasionally used.

Conclusions: Distribution of work on mobile applications is not equal for the eight most prevalent conditions. Whereas some conditions such as diabetes and depression have an overwhelming number of apps and research, there is a lack of apps related to other conditions, such as anemia, hearing loss, or low vision, which must be filled.

(*J Med Internet Res* 2013;15(6):e120) doi:[10.2196/jmir.2600](https://doi.org/10.2196/jmir.2600)

KEYWORDS

apps; mHealth; mobile applications; prevalent conditions; World Health Organization (WHO)

Introduction

Since the creation of the Internet, its massive use, especially in developed countries, has generated new forms of technology in almost every aspect of life [1]. One of these aspects is health care; Internet technologies have initiated major advances in telemedicine and telehealth, now present in every modern health care organization [2]. In the field of telehealth, eHealth has arisen as a paradigm involving the concepts of health, technology, and commerce, with commerce and technology as tools in the service of health [3]. Chang Liu et al (2011) perceive eHealth applications as the software applications that provide tools, processes, and communications in order to support electronic health care practice [4]. In addition to this, with the advent of wireless communications, there are no longer barriers of space and time between health care providers and patients [5]. The use of new wireless communications technology, such as mobile telecommunications networks (2.5G, 3G, 4G, HSPA+), Wireless Local Area Networks (WLAN), Wireless Personal Area Networks (WPAN) including Bluetooth and ZigBee, Wireless Body Area Network (WBAN), Wireless Sensor Networks (WSN), Radio-frequency Identification (RFID), and Worldwide Interoperability for Microwave Access (WiMAX), has greatly boosted telemedicine and eHealth [5-12].

In this context and thanks to these advances in communications, a new term arises: mHealth, a component of eHealth. The Global Observatory for eHealth (GOe) of the World Health Organization (WHO) defines mHealth or mobile health as “medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices” [13]. While new wireless technologies were being developed, new mobile devices were being created. In this way, PDAs, tablets, and smartphones appeared on the market. Although PDAs experienced a boom in the 1990s and early 2000s, they have been replaced by smartphones and tablets with new functions and utilities, which are common now in developed countries [4]. There are already more than 1.08 billion smartphones of a total of 5 billion mobile phones around the world, with 80% of the population having a mobile phone [14]. Regarding tablets, International Data Corporation (IDC) conducted research on their shipments showing 70.9 million shipments of tablets worldwide in 2011 and an estimated 117.1 million and 165.9 million in 2012 and 2013 respectively [15]. Thus, there is great opportunity for mHealth in using these mobile devices and, in fact, a significant number of mHealth applications have been already developed for these platforms.

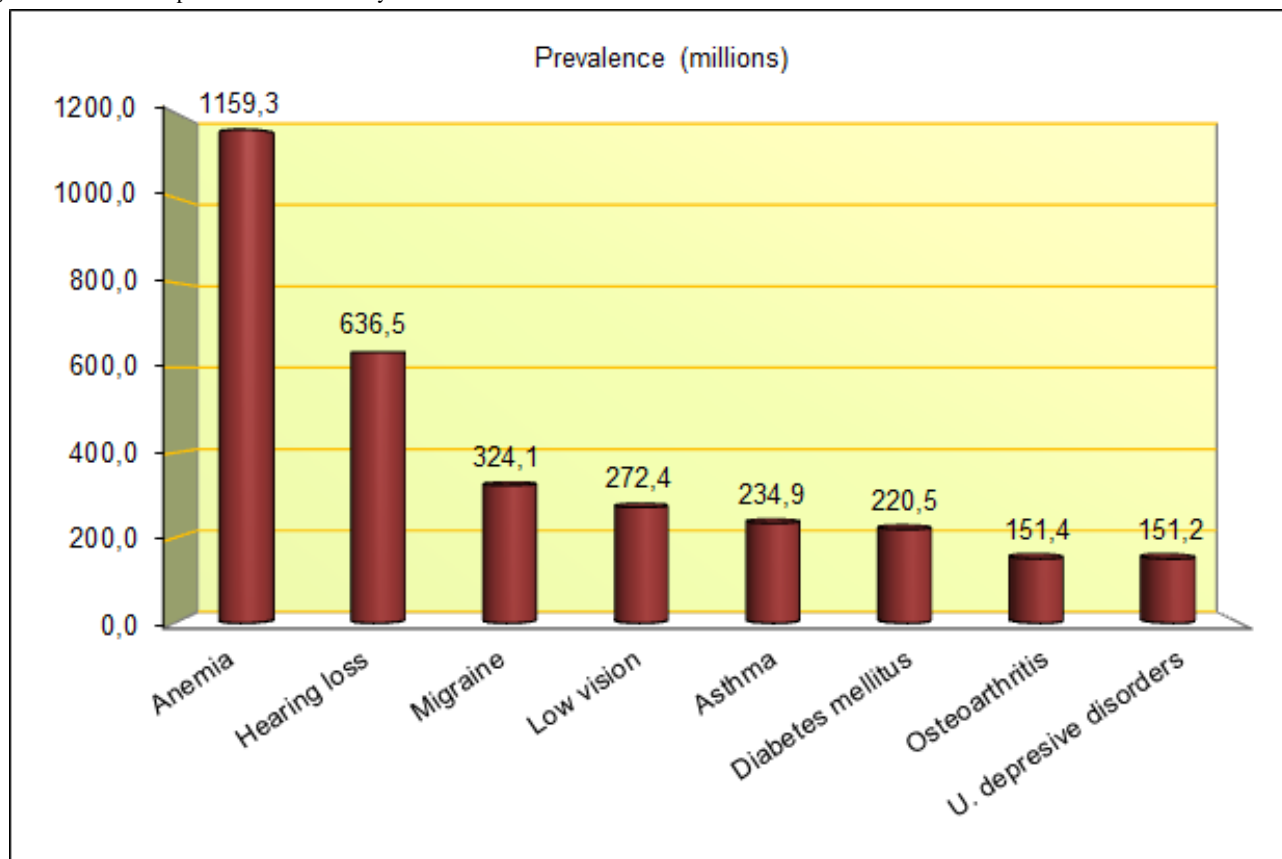
Telecommunications technology aside, it is clear that there is still a long way to go in defeating illness. In 2008, WHO estimated a total of 56.8 million deaths and only 5.1 million of them were caused by injuries. The rest were caused by communicable disease, maternal and perinatal conditions and nutritional deficiencies (15.6 million deaths), and noncommunicable conditions (36.1 million deaths) [16].

Nevertheless, attention should be focused not only on the diseases that cause death, but also the diseases or conditions that can cause a disability or loss of health. In 2004, 2.9% of the world's population were severely disabled, and 12.4% were moderately long-term disabled. In this context, it is essential to know the prevalence of an illness or condition, ie, the number of people who have the condition at any moment [17].

According to WHO's latest update (2004) of the Global Burden of Disease (GBD) [17], the most prevalent conditions are iron-deficiency anemia (IDA), hearing loss, migraine, low vision, asthma, diabetes mellitus, osteoarthritis (OA), and unipolar depressive disorder. The prevalence of each condition is shown in Figure 1. IDA represents 50% of the total cases of anemia (even though both terms are usually used interchangeably, they are not the same). The biggest percentage of affected can be found in underdeveloped and developing zones, in Africa, South East Asia, and Western Pacific, most of them women of reproductive age and children [18-20]. There are two types of hearing loss: moderate or greater hearing loss, which affect 275.7 million individuals, and mild hearing loss, with 360.8 million individuals [21-24]. Migraines are the most prevalent chronic neurological disorder in adults [25], with 11% of affected in Western countries [26-30].

The Global Data on visual impairment 2010 [31] indicates 246 million people with low vision and 39 million blind, equaling a total of 285 million people with any type of visual impairment [32-34]. Asthma is the most common chronic disease in childhood, and most asthma-related deaths take place in poor and developing countries [35-38]. It is estimated that 347 million people have diabetes mellitus [39], commonly named diabetes, but different from diabetes insipidus [40-55]. OA is the most prevalent musculoskeletal disease, and it is thought that 9.6% of men and 18% of women over 60 years have this condition [56-59]. Finally, there are more than 350 million individuals with any unipolar or bipolar depressive disorder [60-63].

To date, there are many published articles about types of wireless connections for mobile devices [8,9,64,65], articles about evaluations of apps for specific objectives [66-68], and reviews of apps of a determined device, software, or field [4,69,70], but there are not articles about the deadliest or the most prevalent conditions and diseases. Hence, the main aim of this paper is to study the existing applications for mobile devices exclusively dedicated to the eight most prevalent conditions [17] and to analyze a sample of the apps for each condition. The goal was to find the number of apps related to each condition, their common features, comparing the commercial ones with those used in research, and finding possible gaps in the development of these types of applications and whatever else might arise. For these purposes, a review has been done: (1) research of published articles containing specific target strings, obtained by search queries in a number of databases, and (2) research of applications related to these conditions in mobile phone application stores.

Figure 1. The 8 most prevalent conditions by the WHO.

Methods

Two different reviews were developed. The first was a literature review and the second, a commercial applications review. Both were current as of April 2013. Finally, one app for each review and each condition was chosen for an in-depth analysis and comparison.

Literature Review

The literature review was developed on the following systems and databases: IEEE Xplore, Scopus, ScienceDirect, Web of Knowledge, and PubMed. After trying several combinations of words and expressions, the following combinations of terms were sought in the metadata field on each of these databases: “condition name” AND mobile AND (applications OR application OR apps OR app); m-health AND “condition name”; “mobile phone” AND “condition name”; smartphone AND “condition name”, where “condition name” is the name of each of the most prevalent conditions as they are issued by [17]. The results are limited to the last 10 years, from 2003 forward. The eligibility criteria were the following: publications not centered on applications using mobile phones or devices were dismissed, as were Web applications not optimized for mobile phone displays. Only papers published in English were studied. Papers on applications for several different diseases or conditions were rejected except for those with an important part of the application dedicated to the related condition. Some aspects were added due to the lack of sufficient results when executing some searches: instead of the search strings with “iron-deficiency anemia”, just “anemia” was used and the British

term “anaemia”. Similarly, on IEEE Xplore, the search string with “anemia” was changed for the string anemia AND applications. Finally, the strings with “unipolar depressive disorders” were replaced by strings with “depressive disorder”, “depressive disorders”, “depressive episodes”, “depressive episode”, and “depression”.

Commercial Apps Review

The second review, the study of the commercial applications, was carried out on the application stores of the most popular smartphones brands [14,71] which are, in descending order of market share, Google play of Google Android [72], Apple iTunes [73], BlackBerry World of BlackBerry (previously RIM or Research In Motion Limited) [74], Windows Phone Apps+Games Store of Microsoft [75], and Nokia’s Ovi Store [76].

For this research, the name of the condition was searched (eg, “hearing loss”) and the eligibility criteria used were the following: applications not centered on the specific condition, not in English, or those included in the category of games, entertainment, or music were dismissed, as well as applications designed for pets or animals, and applications that are actually journals or magazines about the condition. During the search, the following issues were faced:

- On iTunes, apps for iPod and iPhone were separate from the ones for iPad, hence only apps for the first ones were searched, excluding the apps exclusively designed for the Apple tablet.
- Instead of “iron-deficiency anemia”, the word searched was “anemia” due to the lack of results.

- The Ovi search engine does not handle search phrases nor logical operators such as “AND” or “OR” correctly; therefore, the results for “hearing loss” and “low vision” were totally distorted and these results were not considered. In the case of “diabetes mellitus”, no problems were presented because both words are very specific about the condition.
- In the case of diabetes mellitus, the strings “diabetes mellitus” and “diabetes” were used.
- For “unipolar depressive disorders”, instead of this string, “depressive disorders” and “depression” were used because of the absence of results.
- There were some issues with Google play: when searching for “depression” and “diabetes”, the store indicated at least 1000 results but showed only 480. Google was asked about this conflict and the issue is still under investigation, but it is assumed that the first number is the correct one (>1000). In other cases such as the search for anemia, a certain number of results had been found but while exploring the results pages, the last pages (usually between one and three) were blank. Consequently, the number indicated by the store was different from the number of apps shown. In these cases, the number of apps shown was the one used.
- In the case of depression and diabetes, only the first 20 applications were analyzed due to the number of results; therefore, there is potential for completely new investigative work.

Eligibility Criteria for the In-Depth Analysis and Procedure

For the in-depth analysis, two applications were chosen for each condition: one obtained from the literature research and the other from the commercial apps review.

For the mobile applications found in publications, we decided to study the most recent paper on each condition taking into account only the year. In the case of two or more articles published in the same year, the final article was selected considering the impact factor of the journal where it is published (a conference article is considered at a lower level than journal articles), and its number of citations. If, again, two or more articles were published in the same journal and had the same number of citations, these papers were read and the most interesting according to the authors’ opinion was chosen. Articles about reviews of several apps or with insufficient information about the app were dismissed. If the app studied was available in stores, it was downloaded and personally tested on an iPhone 4 in the case of an iOS app or a Samsung Galaxy S SCL GT-I9003 in case of an Android app.

To evaluate the papers, after reading them individually, the authors convened to discuss opinions and fill in a table of features. For apps available on the market, one of the authors downloaded them for a joint evaluation at the meeting.

In the case of commercial apps, we preferred to choose apps from the same store for every condition and to use the store with the most extended software for smartphones. Therefore, Google play was selected because Android fulfilled this prerequisite [71]. For each condition, it was determined to opt for the first relevant free app with a rating, by users, of 3 or

more stars, which Google play shows when searching by the condition sorted by popularity. In addition, another prerequisite was that the app had to be designed for patients, not for health care providers. This way, the most popular free app related to the condition with an evaluation over the mean and designed for the general public was analyzed. However, in the case of anemia, every result with a rating of 3 stars or more was aimed at caregivers; hence, we made an exception where the selected app was not intended for patients and did not fulfil all the previous requisites. The apps were tested on a Samsung Galaxy S SCL GT-I9003.

For the analysis of the commercial apps, the procedure followed was similar to the one developed with the research papers. The authors downloaded them on the mentioned mobile phone Samsung Galaxy S before meeting to study the apps together and complete the previously initiated table of features.

Results

Mobile Applications in Literature

The results of relevant papers on each database and each condition are shown in Table 1. The last row contains the number of different papers found for each condition.

Diabetes was the most investigated condition followed by asthma and depression. There is a significant gap between these three conditions and the rest since the next most investigated conditions were hearing loss and low vision with 9 different publications; this contrasts significantly with the 32 papers on depression. The order of the remaining conditions is: OA sixth, anemia seventh, and migraines last.

Mobile Applications in Stores

The findings on the commercial apps review are shown in Table 2. It shows the number of relevant applications out of the total number of applications found in each store. In the case of diabetes mellitus and depression, the results obtained were separated by “diabetes mellitus” and “diabetes” in the first case, and “depressive disorders” and “depression” in the second case. The last row shows the addition of all the applications located in all the stores for each condition, but it is important to note that some of the applications designed for a specific system were also designed for other systems. Thus, an application developed for Apple iOS can also be developed for Android or Windows Phone, for example.

These results illustrate that the store with most apps is Google play, followed closely by iTunes. The rest of the stores have fewer apps for the conditions searched. Windows Phone Apps+Games Store is the third in number of applications, and it seems that Ovi Store has more applications than BlackBerry World, although that is not clear because of the malfunction of its searching engine mentioned in the Methods section.

Comparing the number of conditions, it is obvious that the conditions with more applications are diabetes and depression. After these two, migraine and asthma are equal with 112 apps, followed by low vision, hearing loss, OA, and finally anemia.

Analysis of a Sample of Reference Apps

The papers and the commercial apps studied are summarized in Table 3 [77-97]. Figures 2 to 9 show snapshots from the

commercial apps.

Tables 4 and 5 show the results from the analysis of some characteristics of the selected applications.

Table 1. Results of the literature review.

	Anemia	Hearing loss	Migraine	Low vision	Asthma	Diabetes mellitus	OA	Depression
IEEE	2	4	2	0	8	16	0	6
Scopus	3	8	2	7	29	112	6	22
ScienceDirect	0	0	0	1	1	5	1	5
WoK	1	5	1	3	25	79	5	18
PubMed	1	2	1	1	16	53	3	13
Total	5	9	3	9	36	140	6	32

Table 2. Results of the commercial apps review.

	Anemia	Hearing loss	Migraine	Low vision	Asthma	Diabetes	Diabetes mellitus	OA	Depression	Depressive disorders
Google Play	7/74	17/42	57/201	33/43	44/226	>1000	19/67	16/46	>1000	1/5
iTunes	7/21	32/37	46/102	30/46	57/124	605	17/21	5/16	419	0/0
BlackBerry	0/0	0/0	5/6	0/0	6/7	33	0/0	0/0	13	0/0
Windows	0/0	3/5	4/8	1/1	4/14	81	2/3	2/2	69	0/0
Ovi Store	0/0	-	0/0	-	1/2	40	15/40	1/1	35	-
Total	14	52	112	64	112	>1759	53	24	>1536	1

Table 3. Summary of the papers and commercial apps.

Health condition	Name of the paper/app	Description
Anemia	Activity and school attendance monitoring system for adolescents with Sickle cell disease [77]	Paper about the app SickLeSAM, designed for Android, whose purpose is monitoring the school attendance of children affected by sickle cell disease, which usually causes anemia
	MD Series: Anemia - Free [78]	App for caregivers that provides some educational tools for the diagnosis and management of different types of anemia in adult patients
Hearing loss	Mobile software Apps support personalized-SRO and serial monitoring with results indicating early detection of hearing loss [79]	Article presenting a software application called OtoID used in a PDA connected to an audiometric unit for monitoring hearing change because of ototoxic medication and others factors, using a testing protocol called Sensitive Range for Ototoxicity (SRO)
	Hearing Tests [80]	App with a hearing test that uses sounds of different frequencies in order to check the user's hearing
Migraine	From a traditional behavioral management program to an mHealth app: Lessons learned in developing mHealth apps for existing health care programs [81]	Application for behavioral migraine management for iPad called iBMM, which can be used for learning relaxing and pain management techniques, tracking migraine attacks and contacting a counselor
	My Headache Log Pro [82]	App for tracking headache attacks by creating a diary of them, with its triggers, symptoms and medications used, and it also allows emailing these notes to the doctor
Low vision	Crowdsourcing subjective fashion advice using VizWiz: Challenges and opportunities [83]	The authors use VizWiz [84] for assisting people with vision impairments in matching garments and being dressed in a fashion way, with the advice of some volunteers
	A.I.type EZReader Theme Pack [85]	Keyboard design for Android smartphones with big keys, high contrast, helpful colors, and audio aid, specially developed for people with visual problems
Asthma	Control of Allergic Rhinitis and Asthma Test (CARAT): dissemination and applications in primary care [86]	The authors talks about an app called m.Carat [87] developed for Android and iOS, consisting of several modules in which asthma and allergic rhinitis (ARA) patients can read news about ARA, record daily events and medications, quantify the level of control of their ARA, and note tasks or reminders
	SIGN Asthma Patient Guide [88]	Guidance for asthmatic patients and relatives in order to know and take control over their condition. It has a section dedicated to patients and other dedicated to parents or carers of asthmatic children
Diabetes	The development of an innovative mobile phone App for Type 1 diabetes alcohol education [89]	The authors develop the app for iOS and Android Type 1 diabetes friend: alcohol guide [90,91], which tries to educate young people with type 1 diabetes about alcohol, meeting clinical guidelines
	OnTrack Diabetes [92]	App for managing diabetes by tracking several data, such as blood glucose and pulse, medication, exercise, and weight
OA	PAGAS Portable and accurate gait analysis system [93]	Paper about a system called PAGAS (Portable and Accurate Gait Analysis System), which includes an Android app in a smartphone connected via Bluetooth to a sensor positioned on the foot and whose purpose is monitoring the gait of patients with altered gait because of different health conditons
	Osteoarthritis of knee [94]	App with information and animated exercises specially designed for osteoarthritis of knee
Depression	CBT for depression: a pilot RCT comparing mobile phone vs computer [95]	The authors use in their study the application VirtualClinic - The Get Happy Program [96] for iOS, which is a cognitive therapy intervention for the management of depression through a comic book story in which users will learn how to manage their depression
	Positive Thinking [97]	App that contains many quotes for helping depressed people and even allows users to write and share their own thoughts

Table 4. Analysis of features (Part 1) of the selected apps.

Condition	Name	Rating	Class	Internet requirement	Clinical/Non-clinical	Data visualization
Anemia	SickleSAM (not commercial)	-	Tracking	No	Clinical	Graphs
	MD Series: Anemia – Free	4.9	Educational	No	Both	Text
Hearing loss	OtoID (not commercial)	-	Diagnosis	No	Clinical	Text, graphs
	Hearing Tests	3	Diagnosis	No	Nonclinical	Text
Migraine	iBMM (not commercial)	-	Educational, guidelines, monitoring	Some functions	Both	Video, graphs, text
	My Headache Log Pro	4.1	Monitoring	Only for sending mails	Both	Text, graphs
Low vision	VizWiz	4.5	Assistive	Yes	Nonclinical	Photos, text, audio
	EZReader Theme Pack	4.4	Assistive	No	Nonclinical	Text, audio
Asthma	m.Carat	-	Monitoring, assistive	Some functions	Nonclinical	Graphs, text, pictures
	SIGN Asthma Patient Guide	4.9	Informative, guidelines	Some functions	Nonclinical	Text, pictures
Diabetes mellitus	Type 1 diabetes friend: alcohol guide	-	Educational, informative	Some sections	Nonclinical	Text, photos
	OnTrack Diabetes	4.5	Monitoring	Only for sending mails	Both	Text, graphs
Osteoarthritis	PAGAS (not commercial)	-	Medical results	No	Both	Text, graphs
	Osteoarthritis of knee	5	Treatment	No	Nonclinical	Text, video
Depression	VirtualClinic – The Get Happy Program	-	Educational, guidelines	Unknown	Nonclinical	Pictures, comic, text
	Positive thinking	4.3	Kind of treatment	Some functions	Nonclinical	Text

Table 5. Analysis of features (Part 2) of the selected apps.

Name	Context awareness	Therapist intervention	Interaction with users	Frequency of use	Interface	Public
SickleSAM (not commercial)	User, location	Yes	No	Continuous	Simple	School children
MD Series: Anemia – Free	No	-	No	Occasional	Not intuitive	Anemia specialists
OtoID (not commercial)	Ambient noise	Yes	No	Regular	Basic	Ear specialists
Hearing Tests	No	No	No	Occasional	Basic	General
iBMM (not commercial)	No	Yes	No	Frequency of migraine attacks	Not intuitive	General
My Headache Log Pro	Preferences, location	Possible	No	Frequency of migraine attacks	Complex, several functions	General
VizWiz	Location	No	Yes, several ways	Frequent	Simple	General
EZReader Theme Pack	Language	No	No	Every time keyboard is used	Basic	General
m.Carat	User, preferences, language	No	No	Constant	Complex	General
SIGN Asthma Patient Guide	No	No	No	Occasional	Basic	General
Type 1 diabetes friend: alcohol guide	No	No	No	Occasional	Simple	Young people (aged 18-21)
OnTrack Diabetes	Preferences	Possible	No	Several times per day	Normal, several functions	General
PAGAS app (not commercial)	Sensor	Possible	No	Regular	Basic	General
Osteoarthritis of knee	No	No	No	Frequent	Simple	General
VirtualClinic – The Get Happy Program	Unknown	No	No	Regular	Simple	General
Positive thinking	Preferences	No	No	Regular	Basic	General

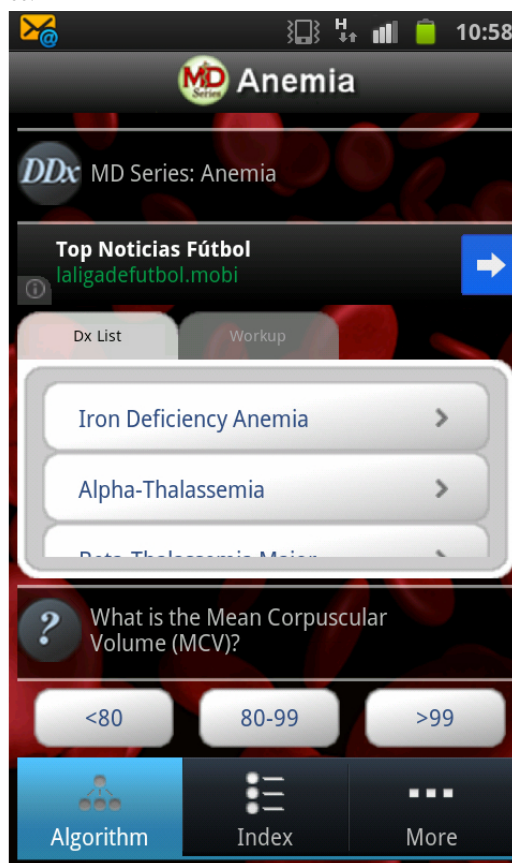
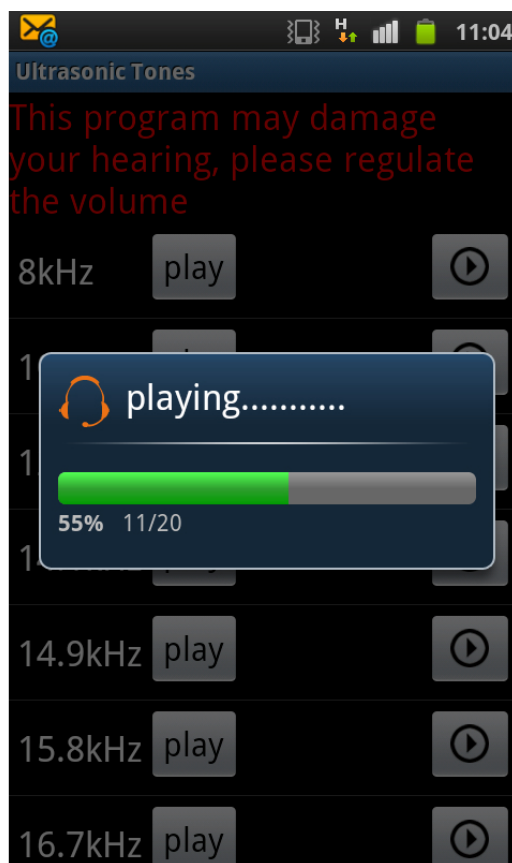
Figure 2. Snapshot of MD Series: Anemia–Free.**Figure 3.** Snapshot of Hearing Tests.

Figure 4. Snapshot of My Headache Log Pro.



Figure 5. Snapshot of AI type EZReader Theme Pack.



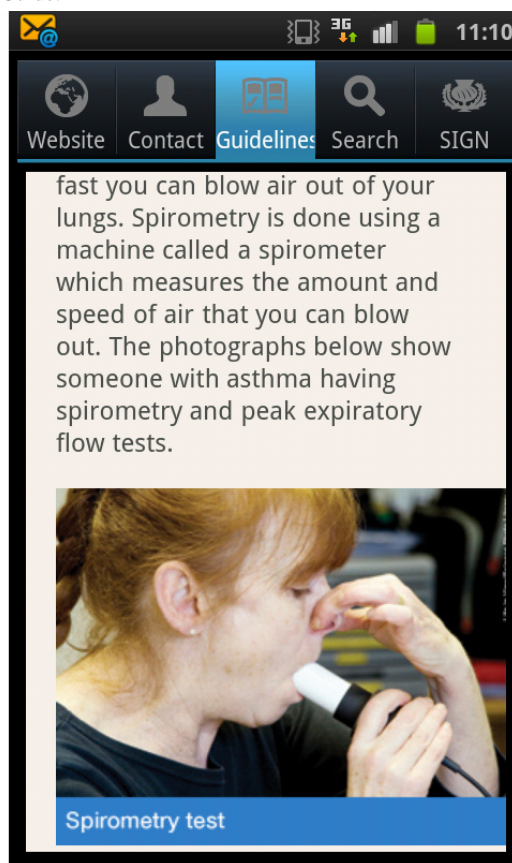
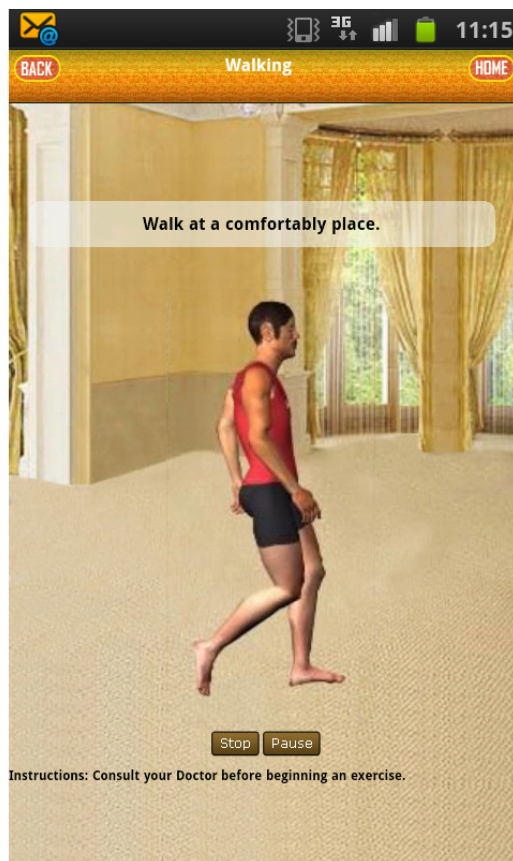
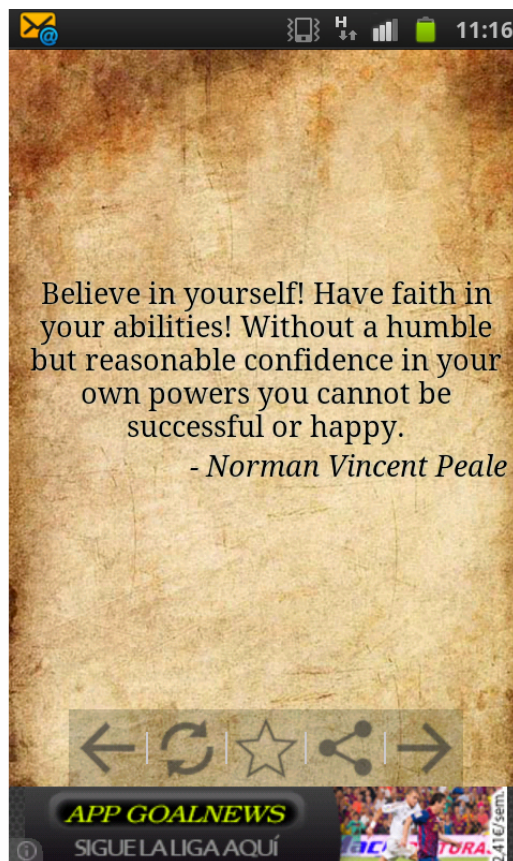
Figure 6. Snapshot of SIGN Asthma Patient Guide.**Figure 7.** Snapshot of OnTrack Diabetes.

Figure 8. Snapshot of Osteoarthritis of knee.**Figure 9.** Snapshot of Positive Thinking.

Discussion

The results of the literature and the commercial reviews show several interesting conclusions. First, the literature review indicates that there are few results compared to the commercial review. This means that the development of mHealth apps has a commercial and economic motivation more than a research motivation. Maybe it would be better to merge both, ie, first developing an app while investigating and then using it for commercial purposes.

Another important finding is the difference in the number of mobile applications for different conditions. There are conditions with more than 1000 apps, such as diabetes or depression while there are others with a range between 14 and 112 apps. The most prevalent condition (anemia) is surprisingly the second with less research and the first with less commercial apps, while two of the less prevalent (diabetes and depression) are the object of a huge number of applications and research. This may be because the majority of IDA cases are located in underdeveloped or developing countries [17,18] where smartphones or tablets are not as widely spread as in the developed countries, and therefore, it is not worthwhile developing apps for IDA. Meanwhile, diabetes or unipolar depression are common conditions associated with modern lifestyles typical of developed zones where there is a strong social conscience of these conditions. However, this is not borne out when comparing the numbers of those affected by diabetes or depression and by IDA in developed countries, because IDA is more extended in these zones than the others. Hence, the probable explanation for this lack of apps for IDA is that the social conscience of it is much less than that of the others and, therefore, it is underinvestigated. For this reason, in light of the numbers, it could be profitable and worthwhile creating apps for IDA.

Following the previous finding, the four most prevalent conditions have fewer apps and research than the remaining four, excluding OA. In addition, there is an important social factor for some of the most prevalent conditions, such as hearing loss or low vision. Therefore, it seems that there is a lack in the research on apps for these conditions and an opportunity for developers.

Contrasting the literature review with the commercial review, it is worth highlighting the cases of asthma and migraines. In the literature, asthma has more results than depression, but it has far fewer results in the commercial review. The opposite occurs with migraines, which move from the last position in number of literature results to the third position in the commercial review. Therefore, there is more work done on commercial apps for migraines than in research applications, as opposed to the situation for asthma.

Finally, comparing the number of apps available on each store, it is clear that application developers prefer Android and iOS for their projects, followed distantly by Windows Phone. Ovi Store and BlackBerry World suffer from a significant lack of apps, which means that developers are not interested in these markets. This is somewhat surprising in the case of BlackBerry since it is the third smartphone platform in market share [71];

hence, there may be an opportunity for developers to fill this empty space in BlackBerry's app store.

Determined criteria were followed in order to select the latest apps published in literature and the most popular and best valued by users' apps in commercial stores. The goal was to choose a sample of reference apps to analyze. These findings are given below.

The majority of apps are for monitoring, assistance, or informing about the condition. In general, apps for diabetes, migraines, and asthma are designed for monitoring the condition, and many have informative sections (or stand-alone informative apps). Apps for low vision are principally assistive, and the aim of most apps for depression is to raise the mood of the affected in several ways. But there are also informative and educational apps, which cover the other conditions, except for anemia, for almost every app is aimed for professional caregivers. For OA, there are an equal number of educational tools and apps with remedies or exercises for managing the pain, and for hearing loss the apps are divided between apps for hearing checks and informative apps.

Table 4 shows that for most apps, the Internet is not required or only required for some functions such as sending emails, which is useful for circumstances or situations where an Internet connection is not available. However, there are apps where an Internet connection is required, eg, for VizWiz. Typically the apps are not designed only for clinical purposes; proof of this is that only some apps developed for research are intended for clinical use. Others can have nonclinical and clinical use (always with the consent of the professional caregiver), but the majority have nonclinical purposes. In addition, there is a link between clinical apps and therapist intervention: if the app is clinical, the therapist intervention is needed. If the app is applicable for both clinical and nonclinical use, the therapist intervention is possible, but not obligatory. With respect to users' interaction, most of the apps do not have this functionality and work as stand-alone apps. Only those conditions that can be relieved in some way with the collaboration of other affected individuals or a determined community of users, such as depression or diabetes, have apps with modules used for these communications.

According to the results, the preferred method of data visualization is text, followed by graphs and pictures or photos. Whereas text visualization is used in almost every case, the use of graphs is common in monitoring or tracking applications in order to show data in a more comfortable and visual way. Something similar occurs with pictures and videos, which are normally used in apps with educational or informational purposes, while audio is a typical aid of low vision apps and is also used for hearing tests in hearing loss apps. Apps with only text visualization or with text and pictures have generally a basic or simple interface, whereas those with graphs or more than two visualization types have a more complex interface, not necessarily intuitive at first use. It is not shocking for apps with several types of data visualization to have a complex interface with several functions, but it is surprising that some of them have such unintuitive ones. Developers need to be careful when designing the interface and its use.

Generally there is no relation between the type of apps and the types of context awareness. Only location awareness can be linked to tracking and monitoring apps where it is important to state the place where an event occurs. For frequency of use, monitoring and assistive apps have a continuous or very frequent use, depending in some cases, such as migraine or asthma, on the frequency of the attacks. The same happens with apps designed to provide some kind of treatment. On the other hand, educational or informative apps are used more infrequently. It might be a good idea to merge monitoring and educational tools in the same app in order to improve frequency of use, number of users, and economic profit, but always with special attention to the interface design. Finally, the majority of the apps analyzed are aimed for the general public who are affected by each condition, which is logical because usually developers do not want to reduce their user circle.

For future work there are various paths to take. It is necessary to fill the lack of anemia apps by creating one aimed for patients, first educational and informative and then exploring other possible opportunities. Another field to populate is related to low vision and hearing loss apps because, as said before, there are few compared with other less prevalent conditions. Nevertheless, in this case it is important to carefully select the intended user because most people with severe hearing loss or low vision problems are over 60 years old in developed countries and do not usually use a smartphone [98]. Hence, designing an assistive app for a 16-50 year-old deaf/blind (or with a severe disability) target group that typically has smartphones can be very useful for the user and even profitable for the developers. In addition to this, the possibilities for creating assistive apps in these fields are enormous.

Acknowledgments

This research has been partially supported by Ministerio de Economía y Competitividad, Spain.

Conflicts of Interest

None declared.

References

1. Van De Belt TH, Engelen LJ, Berben SA, Schoonhoven L. Definition of Health 2.0 and Medicine 2.0: a systematic review. *J Med Internet Res* 2010;12(2):e18 [FREE Full text] [doi: [10.2196/jmir.1350](https://doi.org/10.2196/jmir.1350)] [Medline: [20542857](https://pubmed.ncbi.nlm.nih.gov/20542857/)]
2. Mariani AW, Pêgo-Fernandes PM. Telemedicine: a technological revolution. *Sao Paulo Med J* 2012;130(5):277-278 [FREE Full text] [Medline: [23174864](https://pubmed.ncbi.nlm.nih.gov/23174864/)]
3. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. *J Med Internet Res* 2005;7(1):e1 [FREE Full text] [doi: [10.2196/jmir.7.1.e1](https://doi.org/10.2196/jmir.7.1.e1)] [Medline: [15829471](https://pubmed.ncbi.nlm.nih.gov/15829471/)]
4. Liu C, Zhu Q, Holroyd KA, Seng EK. Status and trends of mobile-health applications for iOS devices: A developer's perspective. *J. Syst. Software* 2011;84(11):2022-2033. [doi: [10.1016/j.jss.2011.06.049](https://doi.org/10.1016/j.jss.2011.06.049)]
5. El Khaddar MA, Harroud H, Boulmalf M, Elkoutbi M, Habbani A. Emerging wireless technologies in e-health trends, challenges, and framework design issues. In: *Proceedings of International Conference on Multimedia Computing and Systems*. 2012 Presented at: International Conference on Multimedia Computing and Systems; Oct. 10-12, 2012; Tangiers, Morocco. [doi: [10.1109/ICMCS.2012.6320276](https://doi.org/10.1109/ICMCS.2012.6320276)]
6. Istepanian R, Jovanov E, Zhang YT. Introduction to the special section on M-Health: beyond seamless mobility and global wireless health-care connectivity. *IEEE Trans Inf Technol Biomed* 2004 Dec;8(4):405-414. [Medline: [15615031](https://pubmed.ncbi.nlm.nih.gov/15615031/)]
7. Yan H, Huo H, Xu Y, Gidlund M. Wireless sensor network based E-health system - implementation and experimental results. *IEEE Trans. Consumer Electron* 2010 Nov;56(4):2288-2295. [doi: [10.1109/TCE.2010.5681102](https://doi.org/10.1109/TCE.2010.5681102)]
8. Alinejad A, Philip N, Istepanian RS. Mapping of multiple parameter m-health scenarios to mobile WiMAX QoS variables. *Conf Proc IEEE Eng Med Biol Soc* 2011;2011:1532-1535. [doi: [10.1109/IEMBS.2011.6090447](https://doi.org/10.1109/IEMBS.2011.6090447)] [Medline: [22254612](https://pubmed.ncbi.nlm.nih.gov/22254612/)]
9. Alinejad A, Istepanian RS, Philip N. Dynamic subframe allocation for mobile broadband m-health using IEEE 802.16j mobile multihop relay networks. *Conf Proc IEEE Eng Med Biol Soc* 2012 Aug;2012:284-287. [doi: [10.1109/EMBC.2012.6345925](https://doi.org/10.1109/EMBC.2012.6345925)] [Medline: [23365886](https://pubmed.ncbi.nlm.nih.gov/23365886/)]
10. Yang SC. Mobile applications and 4G wireless networks: a framework for analysis. *Campus-Wide Information Systems* 2012;29(5):344-357. [doi: [10.1108/10650741211275107](https://doi.org/10.1108/10650741211275107)]
11. Kumar B, Singh SP, Mohan A. Emerging mobile communication technologies for health. In: *Proceedings of the International Conference on Computer and Communication Technology*. 2010 Presented at: International Conference on Computer and Communication Technology; September 17-19, 2010; Allahabad. [doi: [10.1109/ICCCT.2010.5640393](https://doi.org/10.1109/ICCCT.2010.5640393)]
12. Marinkovic S, Popovici E. Ultra low power signal oriented approach for wireless health monitoring. *Sensors (Basel)* 2012;12(6):7917-7937 [FREE Full text] [doi: [10.3390/s120607917](https://doi.org/10.3390/s120607917)] [Medline: [22969379](https://pubmed.ncbi.nlm.nih.gov/22969379/)]
13. World Health Organization. mHealth: New Horizons for Health through Mobile Technologies: Based on the Findings of the Second Global Survey on eHealth (Global Observatory for eHealth Series, Volume 3). 2011. URL: http://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2013-05-23] [WebCite Cache ID 6GpVZoY07]
14. GO-Gulf.com Web Design Company. Smartphone Users Around the World – Statistics and Facts Infographic. URL: <http://www.go-gulf.com/blog/smartphone/> [accessed 2013-02-28] [WebCite Cache ID 6EltOglJx]

15. IDC. IDC Raises Its Worldwide Tablet Forecast on Continued Strong Demand and Forthcoming New Product Launches URL: <http://www.idc.com/getdoc.jsp?containerId=prUS23696912> [accessed 2013-02-28] [WebCite Cache ID 6Eltr9kAy]
16. World Health Organization. Disease and injury regional estimates, cause-specific mortality: regional estimates for 2008. URL: http://www.who.int/gho/mortality_burden_disease/global_burden_disease_DTH6_2008.xls [accessed 2013-02-28] [WebCite Cache ID 6Elu44Zrc]
17. World Health Organization. The global burden of disease: 2004 update. 2008. URL: http://www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_full.pdf [accessed 2013-05-23] [WebCite Cache ID 6GpVpSXeB]
18. Shaw JG, Friedman JF. Iron deficiency anemia: focus on infectious diseases in lesser developed countries. *Anemia* 2011;2011:260380 [FREE Full text] [doi: [10.1155/2011/260380](https://doi.org/10.1155/2011/260380)] [Medline: [21738863](https://pubmed.ncbi.nlm.nih.gov/21738863/)]
19. World Health Organization. Worldwide prevalence of anaemia. 1993–2005: WHO Global Database on Anaemia. URL: http://whqlibdoc.who.int/publications/2008/9789241596657_eng.pdf [accessed 2013-05-23] [WebCite Cache ID 6GpVzjl9v]
20. Clark SF. Iron deficiency anemia. *Nutr Clin Pract* 2008;23(2):128-141. [doi: [10.1177/0884533608314536](https://doi.org/10.1177/0884533608314536)] [Medline: [18390780](https://pubmed.ncbi.nlm.nih.gov/18390780/)]
21. World Health Organization. Health topics: Deafness and hearing impairment. URL: <http://www.who.int/topics/deafness/en/> [accessed 2013-02-28] [WebCite Cache ID 6ElvCgWpv]
22. Hearing Loss Association of America. Basic Facts About Hearing Loss URL: <http://www.hearingloss.org/content/basic-facts-about-hearing-loss> [accessed 2013-02-28] [WebCite Cache ID 6ElwA29kp]
23. World Health Organization. Deafness and hearing loss - Fact Sheet No. 300. URL: <http://www.who.int/mediacentre/factsheets/fs300/en/index.html> [accessed 2013-02-28] [WebCite Cache ID 6ElwJD5bC]
24. American Speech-Language-Hearing Association. Hearing Loss. URL: <http://www.asha.org/public/hearing/Hearing-Loss/> [accessed 2013-02-28] [WebCite Cache ID 6ElwPibAq]
25. Donnet A, Becker H, Allaf B, Lantéri-Minet M. Migraine and migraines of specialists: perceptions and management. *Headache* 2010 Jul;50(7):1115-1125. [doi: [10.1111/j.1526-4610.2010.01660.x](https://doi.org/10.1111/j.1526-4610.2010.01660.x)] [Medline: [20408886](https://pubmed.ncbi.nlm.nih.gov/20408886/)]
26. Leonardi M, Mathers C. Global burden of migraine in the Year 2000: summary of methods and data sources. URL: http://www.who.int/healthinfo/statistics/bod_migraine.pdf [accessed 2013-02-28] [WebCite Cache ID 6ElwjvTWr]
27. World Health Organization. Atlas of headache disorders and resources in the world. 2011. URL: http://www.who.int/mental_health/management/who_atlas_headache_disorders.pdf [accessed 2013-05-23] [WebCite Cache ID 6GpWOJdZl]
28. Sauro KM, Becker WJ. The stress and migraine interaction. *Headache* 2009 Oct;49(9):1378-1386. [doi: [10.1111/j.1526-4610.2009.01486.x](https://doi.org/10.1111/j.1526-4610.2009.01486.x)] [Medline: [19619238](https://pubmed.ncbi.nlm.nih.gov/19619238/)]
29. Steiner TJ, Stovner LJ, Birbeck GL. Migraine: the seventh disabler. *Headache* 2013 Feb;53(2):227-229. [doi: [10.1111/head.12034](https://doi.org/10.1111/head.12034)] [Medline: [23317094](https://pubmed.ncbi.nlm.nih.gov/23317094/)]
30. World Health Organization. Headache disorders - Fact sheet No 277. URL: <http://www.who.int/mediacentre/factsheets/fs277/en/> [accessed 2013-02-28] [WebCite Cache ID 6ElxBfoXW]
31. World Health Organization. Global data on visual impairments 2010. URL: <http://www.who.int/blindness/GLOBALDATAFINALforweb.pdf> [accessed 2013-02-28] [WebCite Cache ID 6ElxO2PA3]
32. Chung ST, Bailey IL, Dagnelie G, Jackson JA, Legge GE, Rubin GS, et al. New challenges in low-vision research. *Optom Vis Sci* 2012 Sep;89(9):1244-1245. [doi: [10.1097/OPX.0b013e31826ba359](https://doi.org/10.1097/OPX.0b013e31826ba359)] [Medline: [22926111](https://pubmed.ncbi.nlm.nih.gov/22926111/)]
33. World Health Organization. Visual impairment and blindness - Fact Sheet No 282. URL: <http://www.who.int/mediacentre/factsheets/fs282/en/index.html> [accessed 2013-02-28] [WebCite Cache ID 6ElxFdKy]
34. Pascolini D, Mariotti SP. Global estimates of visual impairment: 2010. *Br J Ophthalmol* 2012 May;96(5):614-618. [doi: [10.1136/bjophthalmol-2011-300539](https://doi.org/10.1136/bjophthalmol-2011-300539)] [Medline: [22133988](https://pubmed.ncbi.nlm.nih.gov/22133988/)]
35. World Health Organization. Asthma Fact sheet No 307. URL: <http://www.who.int/mediacentre/factsheets/fs307/en/> [accessed 2013-02-28] [WebCite Cache ID 6Elxt2CIV]
36. Madore AM, Laprise C. Immunological and genetic aspects of asthma and allergy. *J Asthma Allergy* 2010;3:107-121 [FREE Full text] [doi: [10.2147/JAA.S8970](https://doi.org/10.2147/JAA.S8970)] [Medline: [21437045](https://pubmed.ncbi.nlm.nih.gov/21437045/)]
37. Shahidi N, Fitzgerald JM. Current recommendations for the treatment of mild asthma. *J Asthma Allergy* 2010;3:169-176 [FREE Full text] [doi: [10.2147/JAA.S14420](https://doi.org/10.2147/JAA.S14420)] [Medline: [21437051](https://pubmed.ncbi.nlm.nih.gov/21437051/)]
38. Clark NM, Griffiths C, Keteyian SR, Partridge MR. Educational and behavioral interventions for asthma: who achieves which outcomes? A systematic review. *J Asthma Allergy* 2010;3:187-197 [FREE Full text] [doi: [10.2147/JAA.S14772](https://doi.org/10.2147/JAA.S14772)] [Medline: [21437053](https://pubmed.ncbi.nlm.nih.gov/21437053/)]
39. World Health Organization. Diabetes Fact Sheet No 312. URL: <http://www.who.int/mediacentre/factsheets/fs312/en/index.html> [accessed 2013-03-01] [WebCite Cache ID 6En0J7Nql]
40. MedicineNet. Diabetes Insipidus. URL: http://www.medicinenet.com/diabetes_inspidus/article.htm [accessed 2013-03-01] [WebCite Cache ID 6En7Jzxe1]
41. World Health Organization. Diabetes Programme. URL: http://www.who.int/diabetes/action_online/basics/en/index.html [accessed 2013-03-01] [WebCite Cache ID 6En0XakxW]
42. Nelson SM, Freeman DJ, Sattar N, Johnstone FD, Lindsay RS. IGF-1 and leptin associate with fetal HDL cholesterol at birth: examination in offspring of mothers with type 1 diabetes. *Diabetes* 2007 Nov;56(11):2705-2709 [FREE Full text] [doi: [10.2337/db07-0585](https://doi.org/10.2337/db07-0585)] [Medline: [17666470](https://pubmed.ncbi.nlm.nih.gov/17666470/)]

43. Chhoun JM, Voltzke KJ, Firpo MT. From cell culture to a cure: pancreatic β -cell replacement strategies for diabetes mellitus. *Regen Med* 2012 Sep;7(5):685-695. [doi: [10.2217/rme.12.39](https://doi.org/10.2217/rme.12.39)] [Medline: [22954439](https://pubmed.ncbi.nlm.nih.gov/22954439/)]
44. Smyth DJ, Cooper JD, Howson JM, Walker NM, Plagnol V, Stevens H, et al. PTPN22 Trp620 explains the association of chromosome 1p13 with type 1 diabetes and shows a statistical interaction with HLA class II genotypes. *Diabetes* 2008 Jun;57(6):1730-1737 [FREE Full text] [doi: [10.2337/db07-1131](https://doi.org/10.2337/db07-1131)] [Medline: [18305142](https://pubmed.ncbi.nlm.nih.gov/18305142/)]
45. Long AE, Gillespie KM, Aitken RJ, Goode JC, Bingley PJ, Williams AJ. Humoral responses to islet antigen-2 and zinc transporter 8 are attenuated in patients carrying HLA-A*24 alleles at the onset of type 1 diabetes. *Diabetes* 2013 Feb 8;---. [doi: [10.2337/db12-1468](https://doi.org/10.2337/db12-1468)] [Medline: [23396399](https://pubmed.ncbi.nlm.nih.gov/23396399/)]
46. Molven A, Ringdal M, Nordbø AM, Raeder H, Støy J, Lipkind GM, Norwegian Childhood Diabetes Study Group, et al. Mutations in the insulin gene can cause MODY and autoantibody-negative type 1 diabetes. *Diabetes* 2008 Apr;57(4):1131-1135 [FREE Full text] [doi: [10.2337/db07-1467](https://doi.org/10.2337/db07-1467)] [Medline: [18192540](https://pubmed.ncbi.nlm.nih.gov/18192540/)]
47. Skyler JS, Ricordi C. Stopping type 1 diabetes: attempts to prevent or cure type 1 diabetes in man. *Diabetes* 2011 Jan;60(1):1-8 [FREE Full text] [doi: [10.2337/db10-1114](https://doi.org/10.2337/db10-1114)] [Medline: [21193733](https://pubmed.ncbi.nlm.nih.gov/21193733/)]
48. Barreda-Pérez M, de la Torre I, López-Coronado M, Rodrigues JJ, García de la Iglesia T. Development and evaluation of a Web-based tool to estimate type 2 diabetes risk: Diab_Alert. *Telemed J E Health* 2013 Feb;19(2):81-87. [doi: [10.1089/tmj.2012.0110](https://doi.org/10.1089/tmj.2012.0110)] [Medline: [23336735](https://pubmed.ncbi.nlm.nih.gov/23336735/)]
49. Galgani M, Nugnes R, Bruzzese D, Perna F, De Rosa V, Procaccini C, et al. Meta-immunological profiling of children with type 1 diabetes identifies new biomarkers to monitor disease progression. *Diabetes* 2013 Feb 8;---. [doi: [10.2337/db12-1273](https://doi.org/10.2337/db12-1273)] [Medline: [23396400](https://pubmed.ncbi.nlm.nih.gov/23396400/)]
50. Anguela XM, Tafuro S, Roca C, Callejas D, Agudo J, Obach M, et al. Nonviral-mediated hepatic expression of IGF-I increases Treg levels and suppresses autoimmune diabetes in mice. *Diabetes* 2013 Feb;62(2):551-560 [FREE Full text] [doi: [10.2337/db11-1776](https://doi.org/10.2337/db11-1776)] [Medline: [23099863](https://pubmed.ncbi.nlm.nih.gov/23099863/)]
51. Seicean S, Strohl KP, Seicean A, Gibby C, Marwick TH. Sleep disordered breathing as a risk of cardiac events in subjects with diabetes mellitus and normal exercise echocardiographic findings. *Am J Cardiol* 2013 Apr 15;111(8):1214-1220. [doi: [10.1016/j.amjcard.2012.12.053](https://doi.org/10.1016/j.amjcard.2012.12.053)] [Medline: [23415514](https://pubmed.ncbi.nlm.nih.gov/23415514/)]
52. Bakker SF, Tushuizen ME, Stokvis-Brantsma WH, Aanstoot HJ, Winterdijk P, van Setten PA, et al. Frequent delay of coeliac disease diagnosis in symptomatic patients with type 1 diabetes mellitus: Clinical and genetic characteristics. *Eur J Intern Med* 2013 Feb 12;---. [doi: [10.1016/j.ejim.2013.01.016](https://doi.org/10.1016/j.ejim.2013.01.016)] [Medline: [23414771](https://pubmed.ncbi.nlm.nih.gov/23414771/)]
53. Lawrence YR, Morag O, Benderly M, Boyko V, Novikov I, Dicker AP, et al. Association between metabolic syndrome, diabetes mellitus and prostate cancer risk. *Prostate Cancer Prostatic Dis* 2013 Jun;16(2):181-186. [doi: [10.1038/pcan.2012.54](https://doi.org/10.1038/pcan.2012.54)] [Medline: [23399780](https://pubmed.ncbi.nlm.nih.gov/23399780/)]
54. Lai CQ, Tucker KL, Parnell LD, Adiconis X, García-Bailo B, Griffith J, et al. PPARGC1A variation associated with DNA damage, diabetes, and cardiovascular diseases: the Boston Puerto Rican Health Study. *Diabetes* 2008 Apr;57(4):809-816 [FREE Full text] [doi: [10.2337/db07-1238](https://doi.org/10.2337/db07-1238)] [Medline: [18162502](https://pubmed.ncbi.nlm.nih.gov/18162502/)]
55. World Health Organization. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation. 2006. URL: http://whqlibdoc.who.int/publications/2006/9241594934_eng.pdf [accessed 2013-05-23] [WebCite Cache ID 6GpXaioVD]
56. Woolf AD, Pfleger B. Burden of major musculoskeletal conditions. *Bull World Health Organ* 2003;81(9):646-656 [FREE Full text] [Medline: [14710506](https://pubmed.ncbi.nlm.nih.gov/14710506/)]
57. World Health Organization. Chronic diseases and health promotion: Chronic rheumatic conditions. URL: <http://www.who.int/chp/topics/rheumatic/en/> [accessed 2013-03-01] [WebCite Cache ID 6En7PenG3]
58. Pereira D, Peleteiro B, Araújo J, Branco J, Santos RA, Ramos E. The effect of osteoarthritis definition on prevalence and incidence estimates: a systematic review. *Osteoarthritis Cartilage* 2011 Nov;19(11):1270-1285. [doi: [10.1016/j.joca.2011.08.009](https://doi.org/10.1016/j.joca.2011.08.009)] [Medline: [21907813](https://pubmed.ncbi.nlm.nih.gov/21907813/)]
59. Brooks P. Inflammation as an important feature of osteoarthritis. *Bull World Health Organ* 2003;81(9):689-690 [FREE Full text] [Medline: [14710513](https://pubmed.ncbi.nlm.nih.gov/14710513/)]
60. World Health Organization. Depression Fact sheet No 369. 2012 Oct. URL: <http://www.who.int/mediacentre/factsheets/fs369/en/> [accessed 2013-03-01] [WebCite Cache ID 6En7rUygT]
61. Ayuso-Mateos JL. Global burden of unipolar depressive disorders in the year 2000. URL: http://www.who.int/healthinfo/statistics/bod_depression.pdf [accessed 2013-03-01] [WebCite Cache ID 6En80Cngb]
62. Munafò MR. The serotonin transporter gene and depression. *Depress Anxiety* 2012 Nov;29(11):915-917. [doi: [10.1002/da.22009](https://doi.org/10.1002/da.22009)] [Medline: [23108923](https://pubmed.ncbi.nlm.nih.gov/23108923/)]
63. Marcus SC, Olfson M. National trends in the treatment for depression from 1998 to 2007. *Arch Gen Psychiatry* 2010 Dec;67(12):1265-1273. [doi: [10.1001/archgenpsychiatry.2010.151](https://doi.org/10.1001/archgenpsychiatry.2010.151)] [Medline: [21135326](https://pubmed.ncbi.nlm.nih.gov/21135326/)]
64. Adibi S. Link technologies and BlackBerry mobile health (mHealth) solutions: a review. *IEEE Trans Inf Technol Biomed* 2012 Jul;16(4):586-597. [doi: [10.1109/TITB.2012.2191295](https://doi.org/10.1109/TITB.2012.2191295)] [Medline: [22453643](https://pubmed.ncbi.nlm.nih.gov/22453643/)]
65. 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](https://doi.org/10.1109/EMBC.2012.6345922)] [Medline: [23365883](https://pubmed.ncbi.nlm.nih.gov/23365883/)]

66. Rajput ZA, Mbugua S, Amadi D, Chepngeno V, Saleem JJ, Anokwa Y, et al. Evaluation of an Android-based mHealth system for population surveillance in developing countries. *J Am Med Inform Assoc* 2012 Aug;19(4):655-659. [doi: [10.1136/amiainl-2011-000476](https://doi.org/10.1136/amiainl-2011-000476)] [Medline: [22366295](https://pubmed.ncbi.nlm.nih.gov/22366295/)]
67. Ferriero G, Vercelli S, Sartorio F, Muñoz Lasa S, Ilieva E, Brigatti E, et al. Reliability of a smartphone-based goniometer for knee joint goniometry. *Int J Rehabil Res* 2013 Jun;36(2):146-151. [doi: [10.1097/MRR.0b013e32835b8269](https://doi.org/10.1097/MRR.0b013e32835b8269)] [Medline: [23196790](https://pubmed.ncbi.nlm.nih.gov/23196790/)]
68. Baggott C, Gibson F, Coll B, Kletter R, Zeltzer P, Miaskowski C. Initial evaluation of an electronic symptom diary for adolescents with cancer. *JMIR Res Protoc* 2012;1(2):e23. [doi: [10.2196/resprot.2175](https://doi.org/10.2196/resprot.2175)] [Medline: [23612521](https://pubmed.ncbi.nlm.nih.gov/23612521/)]
69. 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](https://doi.org/10.2196/jmir.1874)] [Medline: [21979293](https://pubmed.ncbi.nlm.nih.gov/21979293/)]
70. West JH, Hall PC, Hanson CL, Barnes MD, Giraud-Carrier C, Barrett J. There's an app for that: content analysis of paid health and fitness apps. *J Med Internet Res* 2012;14(3):e72 [FREE Full text] [doi: [10.2196/jmir.1977](https://doi.org/10.2196/jmir.1977)] [Medline: [22584372](https://pubmed.ncbi.nlm.nih.gov/22584372/)]
71. Jones C. Apple and Google Continue to Gain US Smartphone Market Share. URL: <http://www.forbes.com/sites/chuckjones/2013/01/04/apple-and-google-continue-to-gain-us-smartphone-market-share/> [accessed 2013-03-01] [WebCite Cache ID [6En85BHQ0](https://www.webcitation.org/6En85BHQ0)]
72. Google. Google play. URL: <https://play.google.com/store> [accessed 2013-05-23] [WebCite Cache ID [6GpY3LmjG](https://www.webcitation.org/6GpY3LmjG)]
73. Apple. iTunes. URL: <http://www.apple.com/itunes/> [accessed 2013-05-23] [WebCite Cache ID [6GpY6FTSn](https://www.webcitation.org/6GpY6FTSn)]
74. BlackBerry. BlackBerry World. URL: <http://appworld.blackberry.com/webstore/product/1/> [accessed 2013-05-23] [WebCite Cache ID [6GpY8akUB](https://www.webcitation.org/6GpY8akUB)]
75. Microsoft. Windows Phone Apps+Games. URL: <http://www.windowsphone.com/en-us/store> [accessed 2013-05-23] [WebCite Cache ID [6GpYBRPWM](https://www.webcitation.org/6GpYBRPWM)]
76. Nokia. Ovi store. URL: <http://store.ovi.com/> [accessed 2013-05-23] [WebCite Cache ID [6GpYEDofL](https://www.webcitation.org/6GpYEDofL)]
77. Venugopalan J, Brown C, Cheng C, Stokes TH, Wang MD. Activity and school attendance monitoring system for adolescents with sickle cell disease. *Conf Proc IEEE Eng Med Biol Soc* 2012 Aug;2012:2456-2459. [doi: [10.1109/EMBC.2012.6346461](https://doi.org/10.1109/EMBC.2012.6346461)] [Medline: [23366422](https://pubmed.ncbi.nlm.nih.gov/23366422/)]
78. Beach-Rak Medicine LLC (Google play). MD Series: Anemia - Free. URL: https://play.google.com/store/apps/details?id=com.enifyservices.MDS.Anemia.Free&feature=search_result#?t=W251bGwsMSwxLDEsImNvbS5lbmlmeXNlcjZpY2VzLk1EUy5BbmVtaWEuRnJlZSJd [WebCite Cache ID [6En5IZ81Z](https://www.webcitation.org/6En5IZ81Z)]
79. Ellingson RM, Helt WJ, Helt PV, Wilmington DJ, Gordon JS, Fausti SA. Mobile software Apps support personalized-SRO and serial monitoring with results indicating early detection of hearing loss. In: Proceedings of the IEEE Instrumentation and Measurement Technology Conference (I2MTC). 2011 Presented at: IEEE Instrumentation and Measurement Technology Conference (I2MTC); May 10-12, 2011; Hangzhou, China. [doi: [10.1109/I2MTC.2011.5944090](https://doi.org/10.1109/I2MTC.2011.5944090)]
80. mikecaroline2008 (Google play). Hearing Tests. URL: https://play.google.com/store/apps/details?id=com.kk.tones&feature=search_result&hl=en [WebCite Cache ID [6En5pexZE](https://www.webcitation.org/6En5pexZE)]
81. Zhu Q, Liu C, Holroyd KA. From a traditional behavioral management program to an m-health app: Lessons learned in developing m-health apps for existing health care programs. In: Proceedings of the 4th International Workshop on Software Engineering in Health Care (SEHC): Lessons learned in developing m-health apps for existing health care programs. 2012 4th International Workshop on Software Engineering in Health Care (SEHC); 2012 Presented at: 4th International Workshop on Software Engineering in Health Care (SEHC); June 4-5, 2012; Zurich.
82. Solar Embedded (Google play). My Headache Log Pro. URL: https://play.google.com/store/apps/details?id=com.dontek.myheadachelog&feature=search_result [WebCite Cache ID [6En5z6FkN](https://www.webcitation.org/6En5z6FkN)]
83. Burton MA, Brady E, Brewer R, Neylan C, Bigham JP, Hurst A. Crowdsourcing subjective fashion advice using VizWiz: challenges and opportunities. 2012 Presented at: 14th international ACM SIGACCESS conference on Computers and accessibility ASSETS '12; Oct. 22-24, 2012; Boulder, Colorado. [doi: [10.1145/2384916.2384941](https://doi.org/10.1145/2384916.2384941)]
84. ROC HCI Group. VizWiz URL: <http://vizwiz.org/> [accessed 2013-03-01] [WebCite Cache ID [6En6RlzMK](https://www.webcitation.org/6En6RlzMK)]
85. AI type (Google play). A.I.type EZReader Theme Pack. URL: https://play.google.com/store/apps/details?id=com.aitype.android.theme.ezreader&feature=search_result [WebCite Cache ID [6En63Eoip](https://www.webcitation.org/6En63Eoip)]
86. Azevedo P, Correia de Sousa J, Bousquet J, Bugalho-Almeida A, Del Giacco SR, Demoly P, WHO Collaborative Center for Asthma and Rhinitis, Montpellier. Control of Allergic Rhinitis and Asthma Test (CARAT): dissemination and applications in primary care. *Prim Care Respir J* 2013 Mar;22(1):112-116 [FREE Full text] [doi: [10.4104/pcrj.2013.00012](https://doi.org/10.4104/pcrj.2013.00012)] [Medline: [23412110](https://pubmed.ncbi.nlm.nih.gov/23412110/)]
87. Control of Allergic Rhinitis and Asthma Test. m.Carat. URL: http://www.caratnetwork.org/index.php?option=com_content&view=article&id=67&Itemid=98&lang=en [WebCite Cache ID [6En6dA2WH](https://www.webcitation.org/6En6dA2WH)]
88. SIGN Executive (Google play). SIGN Asthma Patient Guide URL: https://play.google.com/store/apps/details?id=com.rootcreative.asthma&feature=search_result [WebCite Cache ID [6En675mtd](https://www.webcitation.org/6En675mtd)]
89. Hill J, Masding MG. The development of an innovative mobile phone App for Type 1 diabetes alcohol education. *Diabet Med* 2013 Mar;30(1):112-112.

90. APApps (Google play). T1D Friend: Alcohol Guide URL: https://play.google.com/store/apps/details?id=com.apapps.t1dfriendalcoholguide&feature=search_result#?t=W251bGwsMSwyLDEsImNvbS5hcGFwcHMudDFkZnJpZW5kYWxjb2hvbGd1aWRll0 [WebCite Cache ID 6GAcvRwqK]
91. AP Apps (iTunes). Type 1 diabetes friend: alcohol guide URL: <https://itunes.apple.com/es/app/type-1-diabetes-friend-alcohol/id582346932?mt=8> [accessed 2013-04-26] [WebCite Cache ID 6GAd95mVa]
92. GExperts Inc (Google play). OnTrack Diabetes URL: https://play.google.com/store/apps/details?id=com.gexperts.ontrack&feature=search_result [WebCite Cache ID 6En6C5PLa]
93. Wagner R, Ganz A. PAGAS: Portable and Accurate Gait Analysis System. Conf Proc IEEE Eng Med Biol Soc 2012;2012:280-283. [doi: [10.1109/EMBC.2012.6345924](https://doi.org/10.1109/EMBC.2012.6345924)] [Medline: [23365885](https://pubmed.ncbi.nlm.nih.gov/23365885/)]
94. Canny Technologies (Google play). Osteoarthritis of knee URL: https://play.google.com/store/apps/details?id=com.canny.ossteoarthritis&feature=search_result [WebCite Cache ID 6En6GYoZI]
95. Watts S, Mackenzie A, Thomas C, Griskaitis A, Mewton L, Williams A, et al. CBT for depression: a pilot RCT comparing mobile phone vs. computer. BMC Psychiatry 2013;13:49 [FREE Full text] [doi: [10.1186/1471-244X-13-49](https://doi.org/10.1186/1471-244X-13-49)] [Medline: [23391304](https://pubmed.ncbi.nlm.nih.gov/23391304/)]
96. Global Media Empire (iTunes). VirtualClinic - The Get Happy Program URL: <https://itunes.apple.com/us/app/virtualclinic-get-happy-program/id493461510?mt=8> [accessed 2013-03-01] [WebCite Cache ID 6En73HeML]
97. Juniper Islet (Google play). Positive Thinking URL: https://play.google.com/store/apps/details?id=quotes.positivethinking&feature=search_result [WebCite Cache ID 6En6Lk1dI]
98. Ofcom. Adults media use and attitudes report URL: <http://stakeholders.ofcom.org.uk/binaries/research/media-literacy/media-use-attitudes/adults-media-use-2012.pdf> [accessed 2013-04-26] [WebCite Cache ID 6GAN2twlG]

Abbreviations

ARA: asthma and allergic rhinitis
CARAT: Control of Allergic Rhinitis and Asthma Test
CBT: cognitive behavioral therapy
GBD: global burden of disease
GOe: Global Observatory for eHealth
HSPA: High-Speed Packet Access
IDA: iron-deficiency anemia
IDC: International Data Corporation
OA: osteoarthritis
PAGAS: Portable and Accurate Gait Analysis System
PDA: personal digital assistant
RCT: randomized controlled trial
RFID: Radiofrequency Identification
RIM: Research In Motion Limited
SIGN: Scottish Intercollegiate Guidelines Network
SRO: sensitive range for ototoxicity
WBAN: Wireless Body Area Network
WHO: World Health Organization
WiMAX: Worldwide Interoperability for Microwave Access
WLAN: Wireless Local Area Network
WPAN: Wireless Personal Area Network
WSN: Wireless Sensor Networks

Edited by G Eysenbach; submitted 13.03.13; peer-reviewed by K Lin, X Wang Gao, J Herreros, T Chomutare, J Gao; comments to author 23.04.13; revised version received 30.04.13; accepted 08.05.13; published 14.06.13.

Please cite as:

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
 URL: <http://www.jmir.org/2013/6/e120/>
 doi: [10.2196/jmir.2600](https://doi.org/10.2196/jmir.2600)
 PMID: [23770578](https://pubmed.ncbi.nlm.nih.gov/23770578/)

©Borja Martínez-Pérez, Isabel de la Torre-Díez, Miguel López-Coronado. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 14.06.2013. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Crowdsourcing Participatory Evaluation of Medical Pictograms Using Amazon Mechanical Turk

Bei Yu¹, PhD; Matt Willis¹, MA; Peiyuan Sun¹, BS; Jun Wang¹, PhD

School of Information Studies, Syracuse University, Syracuse, NY, United States

Corresponding Author:

Bei Yu, PhD

School of Information Studies

Syracuse University

Hinds Hall

Syracuse University

Syracuse, NY, 13244

United States

Phone: 1 3154433614

Fax: 1 3154436886

Email: byu@syr.edu

Abstract

Background: Consumer and patient participation proved to be an effective approach for medical pictogram design, but it can be costly and time-consuming. We proposed and evaluated an inexpensive approach that crowdsourced the pictogram evaluation task to Amazon Mechanical Turk (MTurk) workers, who are usually referred to as the “turkers”.

Objective: To answer two research questions: (1) Is the turkers’ collective effort effective for identifying design problems in medical pictograms? and (2) Do the turkers’ demographic characteristics affect their performance in medical pictogram comprehension?

Methods: We designed a Web-based survey (open-ended tests) to ask 100 US turkers to type in their guesses of the meaning of 20 US pharmacopeial pictograms. Two judges independently coded the turkers’ guesses into four categories: correct, partially correct, wrong, and completely wrong. The comprehensibility of a pictogram was measured by the percentage of correct guesses, with each partially correct guess counted as 0.5 correct. We then conducted a content analysis on the turkers’ interpretations to identify misunderstandings and assess whether the misunderstandings were common. We also conducted a statistical analysis to examine the relationship between turkers’ demographic characteristics and their pictogram comprehension performance.

Results: The survey was completed within 3 days of our posting the task to the MTurk, and the collected data are publicly available in the multimedia appendix for download. The comprehensibility for the 20 tested pictograms ranged from 45% to 98%, with an average of 72.5%. The comprehensibility scores of 10 pictograms were strongly correlated to the scores of the same pictograms reported in another study that used oral response-based open-ended testing with local people. The turkers’ misinterpretations shared common errors that exposed design problems in the pictograms. Participant performance was positively correlated with their educational level.

Conclusions: The results confirmed that crowdsourcing can be used as an effective and inexpensive approach for participatory evaluation of medical pictograms. Through Web-based open-ended testing, the crowd can effectively identify problems in pictogram designs. The results also confirmed that education has a significant effect on the comprehension of medical pictograms. Since low-literate people are underrepresented in the turker population, further investigation is needed to examine to what extent turkers’ misunderstandings overlap with those elicited from low-literate people.

(*J Med Internet Res* 2013;15(6):e108) doi:[10.2196/jmir.2513](https://doi.org/10.2196/jmir.2513)

KEYWORDS

crowdsourcing; Amazon Mechanical Turk; participatory design; medical instruction; pictogram; patient communication; readability; health literacy

Introduction

The Department of Health and Human Services defines health literacy as “the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions” [1]. This concept of health literacy is prevalent in the written materials a patient may receive at a hospital, in pharmaceutical instructions, verbal instructions, and any health information encountered online.

Lengthy, purely text-based medical instructions have been reported to result in poor patient attention, comprehension, recall, and adherence. This challenge is particularly acute for patients with low literacy levels, since medical instructions are commonly written at a level exceeding the average American’s reading level, and the average reading level is even lower in certain regions, like inner cities and impoverished areas [2,3].

Many interventions have been designed to improve patients’ understanding of medication. One promising approach is to add pictorial aids or pictograms to patient information materials. Many studies have shown that pictograms can enhance text-based instructions by increasing patients’ attention to the instructions and their comprehension and recall of the content details [2-10].

Studies have also shown that for pictograms to effectively communicate medical instructions, consumers, patients, and health professionals should be involved in the process of iterative design and testing [10-12]. However, the cost for participatory design can be high, considering the variety of medical instructions and the time expenditure for patients, health professionals, and designers. Therefore, to date, participatory design studies have been conducted only on a small scale [2,7,9].

Crowdsourcing, with its low cost of recruiting participants and almost immediate access to a large number of Internet users, provides an attractive option for participatory design and evaluation of medical pictograms [13-18]. We envisioned building a crowdsourcing tool in which Internet users could create a variety of pictograms for any medical instruction, and then the best pictograms would be selected by the crowd and be evaluated for their comprehensibility.

Our concept has two critical components: crowdsourced pictogram design and crowdsourced evaluation. In this study, we focused on the crowdsourced evaluation. Specifically, we aimed to assess the comprehensibility of standard US Pharmacopeial Convention pictograms using Amazon

Mechanical Turk (MTurk). With hundreds of thousands of turkers from over 100 countries, MTurk can help recruit a large number of diversified turkers to work on microtasks in a very short time period at a very low cost, such as a few cents per user response.

Our research questions are: (1) Is the turkers’ collective effort effective for identifying design problems in medical pictograms? and (2) Do the turkers’ demographic attributes affect their performance in medical pictogram comprehension? We hypothesized that turkers would be able to identify common design problems in medical pictograms. We also expected that turkers with higher educational level and caregivers would perform better in this task.

Methods

Our study consisted of three steps: first, searching for samples of medical pictograms; second, programming and deploying the Web-based survey; and third, setting up our survey on MTurk. This section describes the details of each step of the survey set-up.





















Selecting the Medical Pictograms

The ideal pictogram candidates for this evaluation study would be pictograms that were standardized, freely available for others to use, and found on US pharmacological products. Based on these criteria, we chose to use the US Pharmacopeial Convention’s Pictogram Library as the set of pictograms for evaluation. The pictogram library contains 81 pictograms and can be downloaded for free from their website. Many of the pictograms include identical or similar elements. For example, the only difference in the two pictograms in Figure 1 is the order of actions. If a person can understand the first pictogram, it is reasonable to expect he or she could understand the second one as well. Therefore, we selected a set of representative pictograms to minimize redundancy and maximize the inclusion of unique elements. To avoid discomfort, we excluded the pictograms illustrating private parts of human body. At the end, 20 pictograms remained in the test sample set. Figure 2 displays the pictogram images and their official textual interpretations.

Figure 1. Pictogram redundancy.



Figure 2. The 20 pictograms and their comprehensibility scores.

Pictogram	Official description	Correct guesses	Partially correct	Incorrect guesses	Score
	Take 1 hour before meals	75.5%	13.3%	11.2%	82.1
	Wash hands, place drops in nose, wash hands again	82.3%	14.1%	3.5%	89.4
	Store in refrigerator	87.6%	6.9%	5.4%	91.1
	Do not take if pregnant	97.0%	2.5%	0.5%	98.3
	Call your doctor	79.5%	17.0%	3.5%	88.0
	Dissolve in water	5.5%	93.5%	1.0%	52.2
	Injection	41.0%	54.0%	5.0%	68.0
	Shake well	51.0%	49.0%	0.0%	75.5
	If this medicine makes you dizzy, do not drive	18.8%	75.7%	5.4%	56.7
	Do not smoke	94.6%	5.0%	0.5%	97.0
	Check your pulse	95.0%	5.0%	0.0%	97.5
	Do not take other medicines with this medicine	23.5%	43.0%	33.5%	45.0
	Take in the morning	68.3%	14.4%	17.3%	75.5
	Wear medical alert	38.1%	46.0%	15.8%	61.1
	Dilute with water	11.9%	88.1%	0.0%	55.9
	Use this medicine as a gargle	9.0%	80.6%	10.4%	49.3
	Do not use additional salt	5.0%	90.6%	4.5%	50.2
	Do not chew	61.9%	12.4%	25.7%	68.1
	Take with glass of water	95.0%	3.0%	2.0%	96.5
	Drink additional water	6.9%	90.1%	3.0%	52.0
Average		52.4%	40.2%	7.4%	72.5

Designing the Web-Based Survey

We designed and implemented a Web-based survey in the form of open-ended tests. We chose the free Web service provided by the Google App Engine to develop and host the survey app.

The survey included 20 guesses, one for each test pictogram, followed by a short demographic questionnaire at the end. For each guess, the turkers viewed a medical pictogram and answered a question: “What does this medical picture tell you to do?” by typing in their responses in the textbox (see [Figure](#)

3). The demographic questionnaire asked for the participants' gender, age, educational level, number of children or senior members in household, frequency of computer use, and frequency of reading medical labels (see Figure 4).

In addition to the textual responses, we recorded the turkers' IP addresses for the purpose of removing redundant responses because some turkers may have created multiple accounts to be able to perform the same task multiple times in order to earn more. Of course, more than one turker may share a computer,

resulting in identical IP addresses in multiple records. However, we should be able to distinguish these turkers by checking the differences in their pictogram interpretations and their answers to the demographic questions.

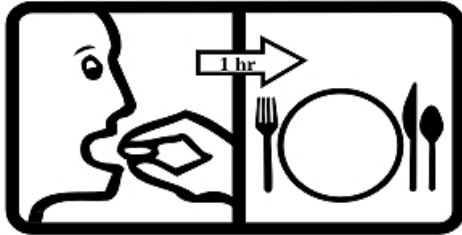
Once a participant finishes the entire survey, the survey app generates a random eight-digit code. The participants should submit this code to MTurk upon completion to verify that they went through the whole survey procedure and to receive payment.

Figure 3. Interface of survey part I.

MEDICAL PICTOGRAM

Pictogram #1 of 20

© 1997 USPC



What does this medical picture tell you to do?

Figure 4. Interface of survey part II.

MEDICAL PICTOGRAM

Thank you for describing the pictures.
Please fill out this brief survey.

Gender: ☐ Male ☐ Female Age:

Education (please select your level of completed education):

Are there any children or seniors living in your household? ☐ Yes ☐ No

How many of them are living there?

Computer use (in hours/week):

How often do you read prescription labels?

Comments:

Setting Up the Task on MTurk

We created an MTurk requestor account to deploy the task, titled “Guess what the image tells you.” Turkers would also see a brief description: “Please view each image and write an instruction of what the image is telling you to do and answer the short survey at the end.” A turker was paid US\$0.30 for interpreting the 20 pictograms and completing the short demographic questionnaire. The requested number of turkers for this task was set to 100.

To avoid spammers, we screened the turkers by setting the minimum prior approval rate to 95%. We also restricted the participants’ location to the United States. Considering that previous studies have shown that cultural backgrounds and ethnicity can have significant effects on people’s comprehension of pictograms [4,10], it is our future work to extend this survey to turkers outside the United States to study the effect of cultural backgrounds.

Results

Data Quality Control

We collected the required number of responses within 3 days. We received 104 responses in total, which means 4 respondents were not paid because they did not submit the confirmation code. We ran the following investigation to ensure the quality of the data. First, we checked for duplicate records. After sorting the data by participants’ IP addresses, we found three pairs of responses with the same IP address. In two pairs, the pictogram interpretations and the demographic survey answers were nearly identical, but the participation dates were different. We counted them as duplicate records and kept only the first record of each on file. In the third pair, the answers were different but valid; the second participation record was not paid. It is most likely that someone else “on-site” with that participating turker took the survey voluntarily but did not submit the code. In this case, we kept both records. Second, we identified the unpaid “volunteers”. We checked to see which random codes assigned by our app were not submitted to MTurk—these were unsolicited volunteers who might have happened to find our website through the participating turkers and who did the survey out of curiosity. We found 4 such “volunteers”, one of whom used the same IP address as 1 participating turker. An examination of the volunteers’ answers shows that they were not spammers; therefore, we kept their answers in the dataset.

Finally, we manually checked the quality of all responses. Only 1 participant was identified as a spammer; this record was deleted from the dataset.

At the end of the process, the data that remained included responses from 101 valid participants. The data are publicly available for download (see [Multimedia Appendix 1](#)).

Coding Open-Ended Interpretations

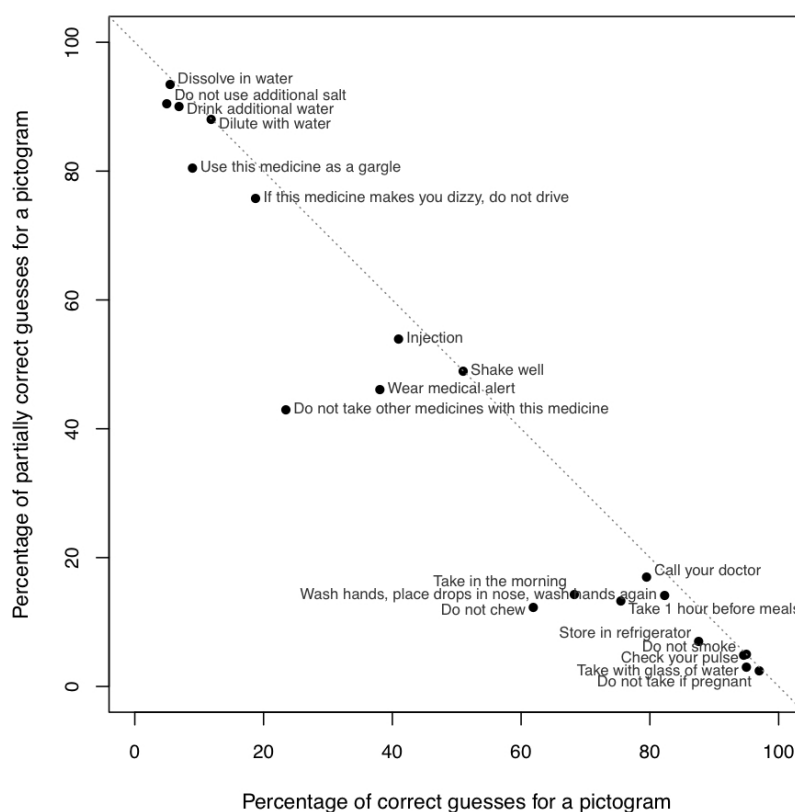
The comprehension test method we used in this study is open-ended testing, which is easier to construct and more accurate than multiple-choice testing, and is considered as a gold-standard in measuring symbol comprehension [19,20]. However, since the answers given by participants are usually short and ambiguous, it is difficult for judges to score them as either correct or incorrect [21]. To facilitate judges’ scoring the correctness of an interpretation and improving the reliability of their scoring, we used the 4-point rating scale (see [Table 1](#)). For example, for pictogram “take 1 hour before meals”, the interpretation “take the pill one hour before eating” would be rated as 1; “take before eating” as 2; “take medicine with food” as 3; and “take one hour after eating” as 4.

Two coders independently rated all the interpretations. The intercoder agreement was 0.83, based on the Krippendorff’s alpha measurement, demonstrating a high concordance between the 2 coders. A review of the disagreements showed that a large portion of the discrepancies were caused by the coders’ judgments on whether to make inferences about the implicit meaning in the responses. For example, for pictogram “take in the morning”, a number of participants gave interpretations like “take upon waking up”. In this case, one coder rated it 3 and the other, 2.

For each pictogram, we calculated the percentages of correct, partially correct, and incorrect (wrong or completely wrong) guesses by each coder’s assessment and then averaged the percentages over the 2 coders (see [Figure 2](#)). On average, 52.4% of the interpretations were correct, 40.2% were partially correct, and 7.4% were incorrect. To help readers better understand the distribution of correct, partially correct, and incorrect guesses, [Figure 5](#) presents a visualization of the distribution for each of the 20 pictograms. If a pictogram falls on the dotted diagonal line, it means all of its interpretations are either correct or partially correct; all pictograms under the diagonal line received at least one incorrect guess.

Table 1. Criteria for judging open-ended interpretations.

Rating	Category	Criteria
1	Correct	The interpretation is the same as, or very close to, the official description. The description maintains the important meaning and semantics of the official description.
2	Partially correct	The interpretation misses some information, or adds information not included in the official description. However, the discrepancies are minor.
3	Wrong	The interpretation is very different from the official description; it is difficult to understand or is confusing.
4	Completely wrong	The interpretation has no resemblance to the official description. It is completely wrong.

Figure 5. Distribution of correct and partially correct interpretations for the 20 pictograms.

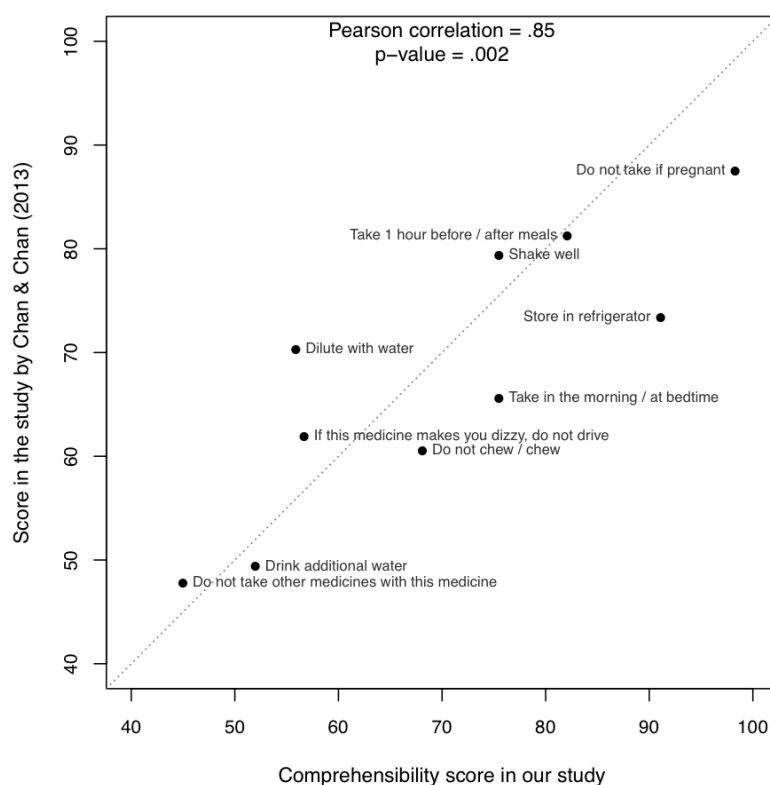
Estimating Comprehensibility of Pictograms

Comprehensibility is usually estimated as the percentage of correct answers given by participants. However, as shown in Figures 2 and 5, 40% of guesses were partially correct in our data. To differentiate correct and partially correct guesses, we adopted ISO's symbol testing procedures, which count partially correct guesses as a fraction in the total correct [20]. For the sake of simplicity, we counted each partially correct answer as 0.5 correct. With this treatment, the comprehensibility scores for the 20 pictograms ranged from 45% to 98%, with an average of 72.5% (Figure 2).

There have been several studies on the comprehensibility of the US pharmacopeial pictograms conducted with local people in South Africa [4], Finland [22], Portugal [23], and Hong Kong [24]. Among the four studies, the Portuguese one used multiple-choice test method—a method that could lead to an inflation of 30% in the comprehension scores when distractor alternatives were less plausible [25]. For the other three studies that used open-ended testing, 10 pictograms in the Hong Kong study, 7 pictograms in the Finland study, and 5 pictograms in

the South Africa study, were the same or very similar to the ones that we used. Thus, we can conduct a comparison with the Hong Kong study (the study with the closest pictograms to ours). The education background of the participants in the Hong Kong study is also the closest to ours: 81% postsecondary education in the Hong Kong study and 92% in our study. In contrast, the participants in the Finland study were children, and the participants in the South Africa study were low-literate. Note that the Hong Kong study used a different scoring mechanism: 3 judges marked each response as either correct or incorrect, and the final decision would be correct or incorrect in case of perfect agreement and 0.5 correct otherwise.

Figure 6 shows the comparison between our study and the Hong Kong study on 10 pictograms. In the figure, the dotted diagonal line represents that any pictogram falling on the line receives the same score from the two studies. The Pearson correlation between the two studies was .85 ($P=.002$). The strong correlation suggests that the pictogram evaluation result, which was obtained through recruiting online turkers to type in responses, is comparable to the result from recruiting local people to provide oral responses to open-ended tests.

Figure 6. Comparison on 10 pictograms between our study and the Hong Kong study.

Content Analysis of Common Misinterpretations

Figure 7 lists common misinterpretations (similar misinterpretations given by at least 2 turkers) for 9 pictograms. The number in the parentheses shows how many turkers described the pictogram in a similar way. For instance, in pictogram “take 1 hour before meals”, 7 participants described it as “take with food”, and 4 interpreted as “take 1 hour after food”. Such critical misinterpretations may well explain why Mansoor and Dowse added clocks in their redesign of the pictogram to prevent people from making the time order error [4,11].

An observation from these common misinterpretations is that some concepts are difficult to represent graphically. For example, it is hard to represent the modifier “additional” in the text “do not use additional water” or “drink additional water”, the verb “chew” in the text “do not chew”, and “morning” in “take in the morning”. To solve this problem, pictogram designers would use alternative strategies such as semantic associations [10,26]. In pictogram “drink additional water”, the US Pharmacopeial Convention used two (extra) glasses of water as an example to represent “additional water”. However, the concept “additional water” was still misunderstood as literally two glasses, three glasses (including the one in hand), or even four glasses (illusion). Hence, semantic associations may not be reliable because their interpretation depends on whether the underlying association or analogy can be identified by users.

Overall, the content analysis result showed that turkers’ misinterpretations shared common errors that exposed design problems in the tested pictograms. This finding validates the

utility of the crowdsourcing approach for the participatory evaluation of medical pictograms.

Participant Demographics and Pictogram Comprehension Score

Table 2 shows the demographic characteristics of the turkers in our study and the relationships between the turkers’ demographics and their comprehension performance. The gender distribution in our study, in which females accounted for 63%, was similar to the general US turker population that Ipeirotis reported in 2010 [27], in which females accounted for 65%. The age distribution was slightly different, with turkers aged 35 years or more accounting for 51% in our study and 45% in their study. The education distribution was also slightly different, with turkers who had college degrees or above accounting for 59% in our study and 54% in their study.

We used a two-tailed *t* test to compare male and female performance, and Spearman rank correlation to measure the correlations between the other ordinal demographic factors and participant performance. Females performed slightly better than males, but the difference was not significant ($P=.078$). Turkers with higher levels of education had better comprehension scores (Spearman $\rho=.25$, $P=.013$), which is consistent with previous studies [28,29]. Educational level is the only factor that affected participant performance.

We also used Gamma test to measure the correlations among the ordinal demographic factors. No correlation was found except an interesting but not surprising one, which was that the number of children or seniors living in a household was negatively correlated with the frequency of computer use (gamma=-.33, $P=.001$).

Figure 7. Common misinterpretations.

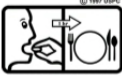








Pictogram	Description	Common misinterpretations
	Take 1 hour before meals	Take medication with food (7) Take one hour after eating (4) Take pill before you eat (3)
	Dissolve in water	Put the pill in water and drink it (7)
	Injection	Needle or syringe (12) Get vaccinated (3)
	Do not take other medicines with this medicine	Do not take this medication (9) Do not take any pills (6)
	Take in the morning	Take just after waking up (27) Take this pill for a good sleep (2)
	Use this medicine as a gargle	The medicine may cause gas and vomiting (8)
	Do not use additional salt	Do not take this medicine with salt (91)
	Do not chew	Do not allow to dissolve in the mouth (6) Let dissolve on your tongue, do not swallow (5) Do not shake your head when taking this (3)
	Drink additional water	Drink two glasses of water (78) Drink four glasses of water (3) Drink three glasses of water (3) Mix together then drink (2)

Table 2. Participant demographics and pictogram comprehension score.

	Proportion	Comprehension
Gender (n=97)		<i>P</i> =.078
Male	.37	70.7
Female	.63	73.1
Educational level (n=99)		<i>P</i> =.013
High school	.08	69.2
Some college	.32	71.4
College degree	.37	72.3
Above college degree	.22	75.0
Age (n=97)		<i>P</i> =.54
≤25	.13	69.5
26-35	.36	73.4
36-45	.14	69.9
46-55	.21	74.5
≥56	.16	72.1
Hours of computer use per week (n=99)		<i>P</i> =.69
≤10	.11	71.1
11-20	.25	72.2
21-30	.16	73.8
31-40	.19	72.4
≥41	.28	72.1
Prescription reading frequency (n=99)		<i>P</i> =.81
Daily	.15	73.0
Weekly	.19	71.5
Monthly	.23	74.1
Every several months	.25	70.9
Never	.17	72.7
Number of children and seniors (n=83)		<i>P</i> =.70
0	.41	71.8
1	.24	73.2
2	.22	71.0
≥3	.13	73.4

Discussion

Principal Findings

Our study aimed to assess whether MTurk, a popular crowdsourcing platform, can be used for participatory evaluation of medical pictograms. We recruited 100 US turkers to guess the meaning of 20 US Pharmacopeial Convention pictograms. The comprehensibility score for the 20 tested pictograms ranged from 45% to 98%, with an average of 72.5%. The scores of 10 pictograms were strongly correlated to the scores of the same pictograms reported in another study that used oral response-based open-ended testing with local people [24]. The turkers' misinterpretations shared common errors that exposed

design problems in the tested pictograms. These results demonstrate that MTurk can be an effective and inexpensive tool for evaluating pictograms and identifying problems in the design of medical pictograms.

We also investigated whether demographic factors (gender, age, educational level, etc) affect participant performance. We found that turkers with higher levels of education had better comprehension performance—a result consistent with previous studies on the effect of education [28,29].

Limitations

A limitation of our study, which was also discussed by Turner et al regarding the use of MTurk in health communication [14],

is that turkers, with relatively higher levels of education, may not be a representative sample of the general population. However, large crowds like MTurk are certainly more representative and cost-effective than the convenience samples in traditional participatory studies, which may consist of as few as 10 to 20 participants due to time and resource constraints [10,12]. To better understand the crowd's representativeness, further investigation is needed to examine to what extent turkers' misunderstandings overlap with those elicited from low-literate people. In addition, the problem of lack of less-educated participants may be greatly reduced when MTurk is available on smart phones, since ethnic minorities and less-educated people, according to the Pew Internet and American Life Project, primarily use their phone for Web access [30].

Another limitation is that our current study recruited only US turkers, and thus we could not conduct any analysis of the effects of cultural backgrounds, an important factor in pictogram comprehension [4,10]. One advantage of recruiting participants from MTurk is that one can recruit turkers from different countries with a variety of languages and cultural backgrounds [14]. It will be our future work to use this advantage to study the effects of cultural backgrounds by recruiting turkers from different countries.

Future Work

We envision building a crowdsourcing tool that allows a large number of Internet users to design and evaluate medical pictograms. In this paper, we focused only on crowdsourced evaluation; in the future, we plan to recruit online users to participate in the design of medical pictograms. Existing work on crowd design, which asked turkers to iteratively sketch, evaluate, and combine the designs of chairs for children, has shown that a crowd-based design process can also be effective [31]. It will be interesting to study how the crowd can be effectively organized and motivated to design high-quality medical pictograms and how the crowdsourcing approach could complement automated illustration of patient instructions [32].

Another interesting direction is to investigate the potential of asking the crowd to evaluate volumes of open-ended interpretations. Open-ended testing is the method recommended by ANSI [19], but it is time-consuming and tedious for judges to score a large number of interpretations—in our case each judge needed to assess about 2000 interpretations. Clearly, turkers can also be recruited for evaluating the interpretations entered by their peers. Existing studies have shown that the crowd can perform well on various annotation tasks, and actually they may perform even better than experts as a result of collective wisdom [18,33,34].

Acknowledgments

We thank all the turkers and anonymous volunteers who participated in this study and the US Pharmacopeial Convention for providing the pictograms. Thanks to the reviewers for very valuable comments. Publication of this article was co-sponsored by the Journal of Medical Internet Research. All authors declared that they failed to identify any funding source for the project described in the paper to enable coverage of the publication costs. All department/division heads/chairs of the author institutions have declared that no institutional funding mechanisms for knowledge dissemination activities and/or coverage of open-access publishing costs at their department or university level exist.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Collected survey data.

[[CSV File, 121KB](#) - [jmir_v15i6e108_app1.csv](#)]

References

1. U.S. Department of Health and Human Services. Healthy People. Washington, DC: U.S. Government Printing Office; 2010.
2. Katz MG, Kripalani S, Weiss BD. Use of pictorial aids in medication instructions: a review of the literature. *Am J Health Syst Pharm* 2006 Dec 1;63(23):2391-2397. [doi: [10.2146/ajhp060162](#)] [Medline: [17106013](#)]
3. Choi J. Literature review: using pictographs in discharge instructions for older adults with low-literacy skills. *J Clin Nurs* 2011 Nov;20(21-22):2984-2996. [doi: [10.1111/j.1365-2702.2011.03814.x](#)] [Medline: [21851434](#)]
4. Dowse R, Ehlers MS. The evaluation of pharmaceutical pictograms in a low-literate South African population. *Patient Educ Couns* 2001 Nov;45(2):87-99. [Medline: [11687321](#)]
5. Dowse R, Ehlers M. Medicine labels incorporating pictograms: do they influence understanding and adherence? *Patient Educ Couns* 2005 Jul;58(1):63-70. [doi: [10.1016/j.pec.2004.06.012](#)] [Medline: [15950838](#)]
6. Houts PS, Doak CC, Doak LG, Loscalzo MJ. The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. *Patient Educ Couns* 2006 May;61(2):173-190. [doi: [10.1016/j.pec.2005.05.004](#)] [Medline: [16122896](#)]
7. Morrow DG, Leirer VO, Andrassy JM. Using icons to convey medication schedule information. *Appl Ergon* 1996 Aug;27(4):267-275. [Medline: [15677067](#)]

8. Ngoh LN, Shepherd MD. Design, development, and evaluation of visual aids for communicating prescription drug instructions to nonliterate patients in rural Cameroon. *Patient Educ Couns* 1997 Jul;31(3):245-261. [Medline: [9277247](#)]
9. Zeng-Treitler Q, Kim H, Hunter M. Improving patient comprehension and recall of discharge instructions by supplementing free texts with pictographs. In: *AMIA Annual Symposium Proceedings*. 2008 Presented at: AMIA Annual Symposium; Nov. 8-12, 2008; Washington, DC p. 849-853 URL: <http://europepmc.org/abstract/MED/18999109>
10. Kim H, Nakamura C, Zeng-Treitler Q. Assessment of pictographs developed through a participatory design process using an online survey tool. *J Med Internet Res* 2009;11(1):e5 [FREE Full text] [doi: [10.2196/jmir.1129](#)] [Medline: [19275981](#)]
11. Mansoor LE, Dowse R. Design and evaluation of a new pharmaceutical pictogram sequence to convey medicine usage. *Ergonomics SA* 2004;16(2):29-41.
12. Ruland CM, Starren J, Vatne TM. Participatory design with children in the development of a support system for patient-centered care in pediatric oncology. *J Biomed Inform* 2008 Aug;41(4):624-635. [doi: [10.1016/j.jbi.2007.10.004](#)] [Medline: [18082468](#)]
13. Swan M. Crowdsourced health research studies: an important emerging complement to clinical trials in the public health research ecosystem. *J Med Internet Res* 2012;14(2):e46 [FREE Full text] [doi: [10.2196/jmir.1988](#)] [Medline: [22397809](#)]
14. Turner AM, Kirchhoff K, Capurro D. Using crowdsourcing technology for testing multilingual public health promotion materials. *J Med Internet Res* 2012;14(3):e79 [FREE Full text] [doi: [10.2196/jmir.2063](#)] [Medline: [22664384](#)]
15. Luengo-Oroz MA, Arranz A, Frean J. Crowdsourcing malaria parasite quantification: an online game for analyzing images of infected thick blood smears. *J Med Internet Res* 2012;14(6):e167 [FREE Full text] [doi: [10.2196/jmir.2338](#)] [Medline: [23196001](#)]
16. Heer J, Bostock M. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. In: *Proceedings of the 28th Annual ACM SIGCHI Conference on Human Factors in Computing Systems*. 2010 Presented at: 28th Annual ACM SIGCHI Conference on Human Factors in Computing Systems; Apr. 10-15, 2010; Atlanta, Georgia p. 203-212. [doi: [10.1145/1753326.1753357](#)]
17. Paolacci G, Chandler J, Ipeirotis PG. Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making* 2010;5(5):411-419. [doi: [10.2139/ssrn.1626226](#)]
18. Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science* 2011 Feb 2011;6(1):3-5. [doi: [10.1177/1745691610393980](#)]
19. American National Standards Institute. American National Standard Criteria for Safety Symbols. In: *ANSI Z535.3*. Rosslyn, VA: ANSI; 2007:3.
20. Wogalter M, Silver N, Leonard S, Zaikina H. Warning symbols. In: Wogalter M, editor. *Handbook of warnings*. Mahwah, N.J.: Lawrence Erlbaum Associates; 2006:159-176.
21. Lesch MF, McDevitt JR. Methodological Issues in Testing Comprehension of Safety Symbols. In: *Proceedings of the 46th Annual Human Factors and Ergonomics Society Annual Meeting*. 2002 Presented at: 46th Annual Human Factors and Ergonomics Society Annual Meeting; Sept. 30-Oct. 4, 2002; Baltimore, MD. [doi: [10.1177/154193120204601912](#)]
22. Hämeen-Anttila K, Kempainen K, Enlund H, Bush Patricia J, Marja A. Do pictograms improve children's understanding of medicine leaflet information? *Patient Educ Couns* 2004 Dec;55(3):371-378. [doi: [10.1016/j.pec.2003.04.006](#)] [Medline: [15582343](#)]
23. Soares MA. Legibility of USP pictograms by clients of community pharmacies in Portugal. *Int J Clin Pharm* 2013 Feb;35(1):22-29. [doi: [10.1007/s11096-012-9698-z](#)] [Medline: [23065039](#)]
24. Chan AH, Chan KW. Effects of prospective-user factors and sign design features on guessability of pharmaceutical pictograms. *Patient Educ Couns* 2013 Feb;90(2):268-275. [doi: [10.1016/j.pec.2012.10.009](#)] [Medline: [23146416](#)]
25. Wolff JS, Wogalter MS. Comprehension of Pictorial Symbols: Effects of Context and Test Method. *hum factors* 1998 Jun 1998;40(2):173-186. [doi: [10.1518/001872098779480433](#)]
26. Nakamura C, Zeng-Treitler Q. A Taxonomy of Representation Strategies in Iconic Communication. *Int J Hum Comput Stud* 2012 Aug 1;70(8):535-551. [doi: [10.1016/j.ijhcs.2012.02.009](#)] [Medline: [22754274](#)]
27. Ipeirotis P. Demographics of Mechanical Turk. New York University Tech Report 2010:1-14 [FREE Full text]
28. Dowse R, Ehlers MS. The influence of education on the interpretation of pharmaceutical pictograms for communicating medicine instructions. *International Journal of Pharmacy Practice* 2003 Mar 2003;11(1):11-18. [doi: [10.1211/002235702810](#)]
29. Richler M, Vaillancourt R, Celetti SJ, Besançon L, Arun K, Sebastien F. The use of pictograms to convey health information regarding side effects and/or indications of medications. *Journal of Communication In Healthcare* 2012;5(4):220-226. [doi: [10.1179/1753807612Y.0000000012](#)]
30. Fox S, Duggan M. Pew Research Center. 2012. Mobile Health 2012 URL: <http://www.pewinternet.org/Reports/2012/Mobile-Health.aspx> [accessed 2013-04-30] [WebCite Cache ID 6GGSXYGnc]
31. Yu L, Nickerson JV. Cooks or Cobblers? Crowd Creativity through Combination. In: *Proceedings of the 29th Annual ACM SIGCHI Conference on Human Factors in Computing Systems*. 2011 Presented at: 29th Annual ACM SIGCHI Conference on Human Factors in Computing Systems; May 7-12, 2011; Vancouver, BC p. 1393-1402. [doi: [10.1145/1978942.1979147](#)]
32. Bui D, Nakamura C, Bray BE, Zeng-Treitler Q. Automated illustration of patients instructions. In: *AMIA Annual Symposium Proceedings*. 2012 Presented at: AMIA Annual Symposium; Nov. 3-7, 2012; Chicago, IL p. 1158-1167 URL: <http://europepmc.org/abstract/MED/23304392>

33. Snow R, O'Connor B, Jurafsky D, Ng AY. Cheap and fast--but is it good?: evaluating non-expert annotations for natural language tasks. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2008 Presented at: Conference on Empirical Methods in Natural Language Processing; Oct. 25-27, 2008; Honolulu, Hawaii p. 254-263.
34. Galton F. Vox Populi. Nature 1907 Mar 07;75:450-451. [doi: [10.1038/075450a0](https://doi.org/10.1038/075450a0)]

Edited by G Eysenbach; submitted 27.12.12; peer-reviewed by Q Zeng, M Fiordelli; comments to author 18.01.13; revised version received 25.01.13; accepted 10.03.13; published 03.06.13.

Please cite as:

Yu B, Willis M, Sun P, Wang J

Crowdsourcing Participatory Evaluation of Medical Pictograms Using Amazon Mechanical Turk

J Med Internet Res 2013;15(6):e108

URL: <http://www.jmir.org/2013/6/e108/>

doi: [10.2196/jmir.2513](https://doi.org/10.2196/jmir.2513)

PMID: [23732572](https://pubmed.ncbi.nlm.nih.gov/23732572/)

©Bei Yu, Matt Willis, Peiyuan Sun, Jun Wang. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 03.06.2013. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Patients' Experiences of Helpfulness in Guided Internet-Based Treatment for Depression: Qualitative Study of Integrated Therapeutic Dimensions

Kjersti R Lillevoll¹, PsyD; Maja Wilhelmsen², MD; Nils Kolstrup², MD, PhD; Ragnhild Sørensen Høifødt³, PsyD; Knut Waterloo³, PhD; Martin Eisemann³, PhD; Mette Bech Risør², PhD

¹Department of Psychology, Faculty of Health Sciences, Tromsø, Norway

²Department of Community Medicine, General Practice Research Unit, Faculty of Health Sciences, University of Tromsø, Tromsø, Norway

³Department of Psychology, Faculty of Health Sciences, University of Tromsø, Tromsø, Norway

Corresponding Author:

Kjersti R Lillevoll, PsyD
Department of Psychology
Faculty of Health Sciences
University of Tromsø
Tromsø, 9037
Norway
Phone: 47 77646774
Fax: 47 77645291
Email: kjersti.lillevoll@uit.no

Abstract

Background: Quantitative research on Internet-based cognitive behavioral therapy (ICBT) has collected substantial evidence for the effectiveness of this treatment approach on health outcomes. Less is known about how patients find ICBT to be generally meaningful and helpful for treating depression.

Objective: To explore patients' experiences of being in ICBT treatment with a focus on the treatment dimensions that they considered helpful.

Methods: Choosing a phenomenological-hermeneutical approach, 14 patients were interviewed with semistructured qualitative interviews to elicit their understanding of using ICBT. The patients took part in a clinical trial using ICBT with MoodGYM, which also featured brief consultations with a clinical psychologist. The interviews were transcribed and analyzed according to the chosen methodology and organized into significant themes.

Results: The phenomenological-hermeneutical analysis identified 5 themes relating overall to the meaning of this mode of treatment in terms of helpfulness. Two related to treatment in general: (1) taking action to address one's problems and (2) the value of talking to a professional. The next two themes specifically addressed guided self-help using the MoodGYM program: (3) acquiring relevant knowledge, and (4) restructuring the new knowledge acquired through ICBT. A fifth theme concerned (5) actual changes in patients' perceptions and interactions, related to either the self-help material or the face-to-face consultations with the therapist.

Conclusions: Three important dimensions were made explicit: the active engagement of the patient, the guidance of the therapist, and the content of the treatment program. The findings pointed to (1) the role of MoodGYM as a source of new knowledge providing patients with a structured approach to work with their depression, (2) the patient's role as the primary agent of change through adapting relevant knowledge from MoodGYM to their situation, and (3) the dialogue with the therapist as a trusting relationship in which to share thoughts and feelings, receive feedback and advice, and to assist the patient in making use of the MoodGYM content.

(*J Med Internet Res* 2013;15(6):e126) doi:[10.2196/jmir.2531](https://doi.org/10.2196/jmir.2531)

KEYWORDS

Internet-based cognitive behavioral therapy; ICBT; guided self-help; depression; qualitative

Introduction

Research has yielded promising results on the effects of Internet-based cognitive behavioral therapy (ICBT) on a range of mental health problems, including depression [1-3]. This form of therapy has the advantage of increased availability and at the same time puts less strain on therapeutic resources [4]. ICBT can be unguided, meaning that the patient works alone with the self-help material; or it can be guided, meaning that the patient enjoys some support and guidance from a therapist. The research on ICBT and other computerized treatments to date indicates that guided self-help and traditional face-to-face-therapy may offer roughly the same success rates for health outcomes [5], but it points to the importance of providing users with support during self-help programs [2,3,6]. Still, the role of support is not well understood in terms of the amount necessary or what it should offer [7,8].

The mechanisms through which ICBT is effective in reducing depression remain unclear. Both specific and common factors of treatment may serve as active ingredients [9]. Cognitive behavioral therapy (CBT) aims to alter the maladaptive structures and processes fundamental to depression [10], by making use of both cognitive and behavioral strategies. A key assumption in CBT is that the depressive patient can use CBT to modify or deactivate his/her depressogenic schema or develop compensatory skills. A qualitative study has provided support for the notion of compensatory skills in face-to-face CBT, where patients utilize the extensive self-therapeutic activities in CBT to manage their depression [11]. Self-therapeutic activity may involve the use of specific CBT techniques or personalized adaptations. Further, the patient might understand and cope with his/her depressive symptoms in light of new psychological knowledge. Similarly, self-therapeutic activity aimed at reducing negative cognitions may be one active ingredient in ICBT. Another line of psychotherapy research has shed light on contextual factors contributing to treatment outcome. The contextual model of psychotherapy [12] recognizes the contributions and interdependencies of other elements beyond the “bare-bones” treatment models and techniques (eg, mechanisms proposed by cognitive theory). These other factors include the actual patient and his/her expectations, factors outside the therapy situation (extra-therapeutic factors), and the working alliance between therapist and patient. The term working alliance refers to the partnership emerging between the therapist and patient in order to achieve the patient’s goal [13]. A robust relationship has been found between the quality of the working alliance—which depends on both therapist and patient factors—and outcome of treatment [14].

Although the effectiveness of computerized and Internet-based CBT is established, there is little agreement concerning the core content [15] and therapeutic process in self-help treatments [16]. Research on possible mechanisms of change has been emerging [9,16], and there is a need for studies aiming at furthering our understanding of active ingredients and processes at work in ICBT. Change in treatment specific factors for ICBT for depression, such as dysfunctional attitudes, worrying, and perceived control, has been found to mediate outcome [17]. Studies of the working alliance in Internet treatments report

overall high ratings that are within the range of alliance ratings in face-to-face therapy [18]. Qualitative research on unguided ICBT has identified both CBT-related and patient aspects as influencing depression. Issues such as computer and Internet skills and the patient’s need for emotional support were reported as important influences during ICBT [6]. Qualitative research into guided ICBT yielded similar findings. It seems that the way patients work with CBT relates to the success of their outcome and their opinion about the therapy: an active, hands-on, self-reliant approach correlates with successful outcomes and favorable opinions. The opposite case is a passive style of working that does not put new knowledge into practice, skips parts of the course material, and is in need of more support [19]. It is not clear, however, what lies behind such differences in approach or whether the cause might be low expectations to the treatment, or because the patient feels a lack of helpfulness during the ICBT process. Purves and Dutton [16] explored patients’ experiences of the therapeutic process in an unguided computerized CBT program and identified four themes in interviews with patients. These included a meaningful relationship with the self-help material, using the self-help material to create structure to their psychological state, being stimulated by the self-help material creating engagement, and an increased sense of personal agency. In sum, the current literature implies that specific and common factors contribute to the outcome of ICBT. In order to improve effectiveness and acceptability of such treatment packages, further examination of these issues are warranted [9].

This study explores patients’ experiences of helpfulness in guided Internet-based cognitive behavioral therapy for depression. We intend to combine a pragmatic intention with a phenomenological-hermeneutical approach on patient experiences rather than on isolated patient factors. Such an approach, inspired by Husserl, focuses on human experience in everyday life, explores a natural attitude, and understands the whole of an experience [20]. Further, human beings have intentional relationships with their surroundings and things in their everyday lives, that is, relationships that are experienced as meaningful. In this case, we mean to elucidate how patients experience and give meaning to the phenomenon of ICBT in an everyday context. Such experience can be narrated and presented as text, which again calls for a hermeneutical interpretation [21]. To grasp the essential meaning of ICBT, this study aims to explore the experiences of patients in an ICBT intervention with therapist support. Knowledge about what patients experience as most helpful, and how patients understand and implement the principles of CBT, is essential if we want to improve the quality of ICBT programs and patient support. The current study aims to explore patients’ experiences with ICBT, focusing especially on those aspects of the therapy that they consider most helpful.

Methods

Design

This qualitative study was conducted in parallel with a randomized controlled trial (RCT; ACTRN12610000257066). The aim of the RCT was to test a treatment approach with ICBT

that could be feasible in general practice. It compares Internet-based cognitive behavioral therapy interspersed with brief personal sessions with a therapist to a waiting list control sample. Participants in the RCT were recruited for this qualitative study when ending treatment. The first and second authors of this paper took part in the planning of the RCT and conducted the interviews. The first author worked as a therapist in the RCT but did not interview her own patients. Both the first and the second author conducting the interviews were blind to the outcome of the patients at the time of the interviews.

Description of the Randomized Controlled Trial and Treatment

Patients included in the RCT after an initial assessment had a session with their therapist who introduced the self-help program. The introduction focused on giving brief information about the theoretical basis and the empirical support, as well as the content of the program and expected work load. Patients were asked to complete the five modules of MoodGYM in sequence, one per week. They were then followed up with weekly, face-to-face consultations with the therapist over a minimum of 7 weeks. A full course treatment included eight consultations. However, the treatment protocol was quite flexible and allowed for delays in the treatment, and it did not have a maximum limit for therapist sessions. In case of delays, the therapist contacted the patients to reschedule a new appointment.

The consultations each lasted approximately 15-30 minutes, similar to the time available in general practice. The guideline script comprised three compulsory subjects: (1) symptom monitoring, (2) discussion of the topic of the last module in MoodGYM, and (3) introducing the next module and discuss patient motivation. If time permitted, other issues that patients perceived as important to their depression were also discussed.

The self-help program used in the randomized controlled trial was MoodGYM. Its aim is to help patients prevent and cope with depression, based on principles of cognitive behavioral therapy [22,23]. It was developed at the Centre for Mental Health Research at Australian National University, and its effectiveness is empirically supported [24,25]. MoodGYM has five modules containing texts explaining the basic principles of CBT, a variety of self-tests and self-help exercises, and homework in which the patient is invited to analyze some personal experience in accordance with the principles of the program. Although some of the content in MoodGYM is generic CBT, there are also some specific sections devoted to parental

relationships, relationship break-ups, problem solving, and even relaxation.

In the RCT sample, 72.6% were female, and age ranged from 18-63 with a mean of 36.1. The number of treatment sessions ranged from 1-12, with a median of 8 sessions; 40.1% did not complete the treatment program. The baseline depression scores, as measured by the Beck Depression Inventory (BDI-II) [26], had a mean of 21.7. A full description of the sample can be obtained in the forthcoming paper presenting the results of the RCT (personal communication by Høifødt, Ragnhild, March 2013).

Recruitment to the Interviews and Procedure

Overall, the study was planned to provide complementary knowledge production on ICBT different from the kind of knowledge obtained in an RCT, hence the use of a qualitative approach. The recruitment of patients to the interviews was parallel to the randomized controlled trial, with patients receiving oral and written information about our qualitative study and an invitation to participate at their final consultation. Recruitment was continuous until the desired total of 14 interviews was reached. The recruitment procedure was strategic in the sense that we aimed to include men and women, younger and older, completers and noncompleters. The therapist would provide the interviewer with contact details for consenting patients. Patients could choose the location, either at their home or at the university in comfortable everyday like settings. All patients preferred to be interviewed at the university. The interviews lasted for approximately 60 minutes, were recorded using a digital voice recorder, transcribed verbatim by the second author or a clerical assistant, and then coded using NVIVO software. The initial coding procedure comprised a separate coding by the first, second, and last author of two interviews, and subsequently checked for consensus. During the entire process of analysis, the coding and the subsequent themes were discussed and reflected upon.

The researcher's interview guide consisted of open questions inviting the patients to narrate different aspects of their ICBT experience: their motivation during progression, any changes they made in their everyday life, and any changes they perceived in their condition. This prompted answers concerning, eg, sociality, temporality, and spatiality of one's lifeworld (the sum total of physical surroundings and everyday experiences that make up an individual's world). As a whole, the interview was performed as an open dialogue interview. Table 1 presents a list of questions within each of the topic areas.

Table 1. Main questions in the interview guide.

Topic	Questions
Changes in everyday life	What was your life like before you got depressed? What was it like during your depression? How is it like after treatment? Can you describe any changes you have made during this time? Can you recall situations where you have acted differently as a result of the treatment? Do you think people close to you will have noticed any difference?
Motivation during treatment	What made you start treatment with MoodGYM? How did you progress through treatment? Which were important elements in treatment? Did you experience any difficulties?
The treatment	What did you think of the treatment? Was there anything you liked or disliked in particular? How did you like working with the computer? How did you experience the relationship with the therapist? In what way were you able to influence the progression through the treatment? Did you need to make any practical adjustment in your everyday life?

The qualitative study from the onset employed a phenomenological-hermeneutical methodology, which basically means that we sought to elicit the way participants related to the treatment. The approach is phenomenological in the sense that it understands human experience as founded in a basic relatedness to the world though merely living it in a naïve way, with a natural attitude. This experience, however, may be expressed in narratives, actions, or reflections, showing the intentionality of a human being, and may be described scientifically. This approach is inspired by Lindseth and Norberg [21] who again rest their methodology on both Husserl, Heidegger, and Ricoeur [21,27]. Semistructured interviews were conducted to elicit empirical information about patient experiences in treatment, aiming to set aside “the taken for granted” attitude of their lived experience.

Patients

Fourteen patients were recruited from the RCT sample. The patients interviewed were 5 men and 9 women (64%), aged 22–61 years. The number of consultations ranged from 3–11. Three patients (21.4%) did not complete the treatment program. At pretreatment assessment, Beck depression inventory scores ranged from 10–28 (mean 18.27). At posttreatment assessment, 6 patients (42.8%) had not changed, 1 (7.1%) had improved, and 7 (50%) had recovered, based on criteria for clinically significant change [28].

Analysis

Carrying out a phenomenological analysis requires the researcher to reflect carefully upon the taken for granted statements of one’s informants and to approach these with an attitude of “bracketing”, that is, to examine and question openly what is being expressed [20]. Also according to Lindseth & Norberg [21], a phenomenological-hermeneutical analysis cycles through different levels of understanding of the text material. First, the interview is read to achieve a superficial understanding of what the text is all about. Second, the text is analyzed in terms of its meaning units (semantics) with a reflective approach. The semantic units are then condensed to form themes and subthemes that disclose meaning in everyday words rather than merely portraying concepts. Also, the themes develop from the material rather than from the interview topics. The identified themes are, after the second step, compared with the initial cursory understanding for validation. Third, the themes are reviewed and reflected on as a whole, and an overall understanding may be reached through critical reflection based on theoretical literature. This procedure was adhered to, but it was necessary to compare the themes not only with the naïve understanding but also with the original interview transcript. And in the third step, comprehensive understanding was achieved in light of theories about change processes in psychotherapy. In summary, the analysis involved a dynamic process moving between abstracted themes and the interview transcripts for validation [27]. During the analytical steps, essential meaningful experiences with the treatment were identified and one of them was what counted as “helpful” for the patients. The themes revealing what helpfulness consisted of might emerge as pragmatic issues (which they were), but what was experienced as helpful was also interpreted as

capturing exactly how the treatment became meaningful to the patients.

Results

During analysis, several meaningful themes were identified from the interviews. The overall topic chosen for this paper is how patients perceive the treatment as being helpful. Another paper focuses on motivational aspects of the treatment (personal communication by Wilhelmsen, Maja, Jan. 2013).

Overall, we defined five themes reflecting perceived helpful dimensions. Two themes were related to being in treatment in general and were nonspecific to ICBT. These concerned (1) taking action to address one’s problems and (2) the value of talking to a professional. Two themes specifically addressed the patients’ experiences with the self-help program material: (3) acquiring relevant knowledge and (4) restructuring the new knowledge. A fifth theme concerned (5) the way that patients describe actual changes in perceptions and interactions with their environment that patients relate to the treatment they have undergone.

Being in Treatment

Taking Action to Address One’s Problem

In most cases, patients had been considering the pros and cons of seeking help for a long time, even for years. One typical response to taking this difficult step was the sense of relief and satisfaction that patients felt when they took action to address their problem. Some patients had not gotten very far with the process and had not experienced any improvement, while others were highly active during treatment and making deliberate changes to their lives one step at a time. In both cases, patients talked about the mere act of doing something as important:

Working with MoodGYM, the best thing about it all was that I was doing something about it. You know, coming to these sessions every week, getting to talk, starting the next chapter. You know, the things I worked with did not suffice, but I felt good working with it. I felt sort of like I was getting out from...getting back to normal. [Male (26)]

The analysis reflects the significance of moving from a state of passivity to one of activity. Patients recounted that they had felt a need to move forward that brought them to seek help. Taking action was in itself an achievement that engendered hope and brought a positive effect. The treatment might not be all they had hoped for, but patients still valued their own effort to try and improve their own situation. This theme is related to motivation in the treatment process, but it became evident in the interviews that this affected how the patients were feeling, as a meaningful event towards recovery.

Patients highlighted the easy access to the treatment as another important facilitator for taking action. The alternative for many people with depression would be referral to specialist health services, where long wait times leave patients passive. The alternative of a private clinic leaves patients worried about their financial liabilities. Our patients had low expectations about the abilities of general practitioners to deal with

depression—GPs were assumed to have little time on their hands and little knowledge of mental illness. In general, our patients were pleased to be offered a course of qualified therapy universally available online at no cost.

The Value of Talking to a Professional

There was consensus among the interviewees that the involvement of a therapist was vital in the treatment and that talking to a professional is very important to them. The level of satisfaction with the amount of contact with the therapist varied a great deal. Some patients were quite satisfied, while others came to realize that what they really wanted was conventional face-to-face therapy, not guided ICBT. For all patients, a trusting relationship with a professional was a fundamental part of the treatment:

I thought it [the relationship with the therapist] was really good! She didn't make me feel judged in any way. She was very accommodating. Almost as if she understood what I was talking about. She sometimes was ahead of me about things I was going to say, in a way. She understood very well what it was like. [Female (22)]

Patients expressed a need to talk freely about their issues and a chance to reveal things about themselves without fear of judgment. Their need for appropriate verbal communication was significant. Communication with the therapist was described as different from everyday interaction with others, and patients thought it essential to the treatment. The chance for patients to share their experiences, innermost thoughts, and feelings was something they found important:

In the first session I got to talk about it all. So...the first session was very important...the first and the second session, that was when I got to talk about the things that troubled me. So I believe...I think it was very important to be able to do that. [Male (26)]

Furthermore, the reciprocity of the relationship was emphasized: not in the sense that the therapist and the patient are equal, rather, the role of the therapist as a professional was appreciated. The importance of a dialogue was stressed, where the patient could ask questions, discuss issues with the therapist, receive trustworthy feedback, and be supported and acknowledged:

But, I have sort of looked for...what shall I say, a professional or an adult who has in some way supported me in my thoughts about all the things I do. Those things I do because I feel guilty, have to, but in reality don't need to. I would like a verbal confirmation that I'm doing enough! [Female (56)]

These aspects of the professional relationship, that is, having someone you can trust, a chance to freely express yourself, and receiving individual affirmation, were for some just as important as working with direct symptom relief or negative thought patterns. It seems that by engaging in this dialogue, these patients were reassured by the affirmation and support they received.

Internet-Based Cognitive Behavioral Therapy

One vital aspect of MoodGYM depression therapy is that patients are actively engaged; they receive homework aimed at challenging their repetitive automatic negative thoughts and cognitive distortions. The latter two themes: (3) acquiring relevant knowledge and (4) restructuring the new knowledge, describe ways in which the content of MoodGYM was experienced by patients. The extent to which patients could relate to MoodGYM varied greatly, so these responses provided insight into relevant dimensions of interacting with the program.

Acquiring Relevant Knowledge

Patients commonly describe features of the self-help material as being experienced as particularly relevant, issues presented in the program that “have to do with me”, or in contrast, issues that were irrelevant. Typically, patients were more satisfied with the standard content of CBT (eg, presenting the principles, cognitive distortions) rather than the content aimed at specific problems (eg, relationship break-ups, parental relationships). Patients could accept some parts of the content being of little relevance to them, as long as they could find other parts that they could learn from:

The first thing I came to the session and said, “this, this is huge”. It [a presentation of the basic ABC-model of CBT] was the drawing we saw early on, with situations, cognitions and feelings. I had never seen that drawing before...and if I were to draw it, feelings would never be in such a drawing. Because I've ignored feelings...So that was, well, a first...a sort of awakening. [Female (39)]

The fact that a patient could recognize something in the program content and feel its relevance was a way to feel support and recognition in a situation that otherwise was unfamiliar or hard to accept. The process of seeing relevance in the material is intertwined with the process of learning and acquiring new self-knowledge. Conversely, patients who found no relevance in the material were unlikely to learn anything. So for some respondents, the content raised awareness, reactivated once-familiar knowledge, or provided new insights; whereas for others, the feeling was that the content was not meaningful to them:

Well, so little by little, when I could only identify with the character that was not depressed, then it like became more and more...it was almost as if I felt myself getting annoyed by those modules. And I decided that this here stuff doesn't give me anything. [Female (28)]

Not all patients felt able to relate to the principles of CBT presented in MoodGYM. This was generally due to a mismatch of the program aims and what patients perceived as their most pressing problem. All our patients showed symptoms of depression, but not all of them felt depressive thoughts and ideas to be their principal problem. Some were quite clear in their mind that ICBT would not provide the answer to their difficulties, perhaps to the extent that they were unwilling or unable to relate to the MoodGYM content.

Restructuring the New Knowledge

Patients' accounts of their experiences with the program content show how they reflected on the material, adapting or processing it to suit their own perceived needs. Patients who found parts of the self-help program relevant did not necessarily accept it all uncritically, rather, they described an active process of interpreting the material. This also meant reflecting on past events, thoughts, and feelings in light of their newfound understanding:

I felt it [working with the modules] took a long time because I was sitting reading and trying to interact...interact with what I read...It was not that I struggled with the homework or with understanding what it said, but I chose to spend time on it. [Female (26)]

It was evident that the proposed approaches and techniques were not universally suitable, and some patients went to considerable lengths to restructure the content to suit their perceived needs. A table in module 2 gives an overview of distorted cognitions (Figure 1).

There were an awful lot of categories. How could I make it useful? On a daily basis, I had to merge some into larger groups and then work out, like, "Alright, now I'm making this type of mistake". It had to be restructured a bit, because I couldn't be bothered to sit and cram all of them. And I didn't need to either. [Male (33)]

The patients talked about how the program material made them more aware of their own negative thinking and that this awareness opened up for further reflections about the validity of those thoughts, and how such thoughts are incorporated in a negative cycle that also includes feelings and actions. However, not all patients were able to make adaptations of the material to fit their own needs, despite recognizing the relevance of the material. For some, bridging the gap between theoretical concepts of negative thinking and making the ideas their own, was far from easy. Sessions with the therapist helped the bridging process by enhancing patients' understanding and ideas about program content: "Indeed the conversations helped make the content of the Internet form more elastic". [Female (39)]

Actual Changes in Perceptions and Interactions

Patients described changes in both thought and behavior as essential. Some changes were clearly related to what they learned from the self-help program, eg, testing the truth value of their negative thinking or using specific techniques from the program, whereas other changes had been discussed with the therapist in the sessions and were not directly related to the self-help program. Changes in thought patterns, where patients started to question the content and validity of their depressive ideas, were a specific result of the self-help techniques:

And also in relation to the business of structuring and categorizing thoughts and mindsets, and doing a reality check on what you're thinking, related to what has happened or what you're feeling—that has been very helpful to me. What does this really mean? Where

does this come from? This has been very useful to me. [Female (39)]

The patients were describing how they came to realize that thoughts and ideas about an event are not accurate representations of reality. Moreover, they might be biased and need careful scrutiny. Questioning the validity of thoughts and ideas made room to explore other possible interpretations, and this process helped give patients more flexibility in their thinking. It showed, for example, how a bias in their mindset might be making their depression worse. Behavioral changes could follow such cognitive changes, including how patients related to events and how they discussed things with others. A technique in module 3 distinguishes observations from interpretations (Figure 2). The following quote illustrates how an exercise from MoodGYM ("The Reporter") is incorporated by one patient:

There were some things I remember, and that was when I got that far in the program, then they talked about a reporter. I've given that much thought, and that's something I'll take with me. Actually, I am an emotional person, and sometimes I get annoyed with myself...And then there was this here reporter...her name is Vold. She [a Norwegian television reporter] reports from the Middle East and Palestine, and I think she's good. And I imagined her in front of me...she's in the midst of fighting, or they are shooting, or they are fighting all around her, or at least in the background. And she's standing there trying to report accurately on what has happened. I think it's admirable. So I've had her with me in quite a few situations. [Female (61)]

Other behavioral changes served to break a negative cycle of self-recrimination, inactivity, and withdrawal. These changes could produce positive secondary benefits for the patients, including getting closer to the people around them:

I got to know her [his partner's] kids better, for instance. That's very positive. [Interviewer: "How did that happen?"] Well, I tried to make the best of it and play with them, not going into a room and hiding away all by myself. [Male (26)]

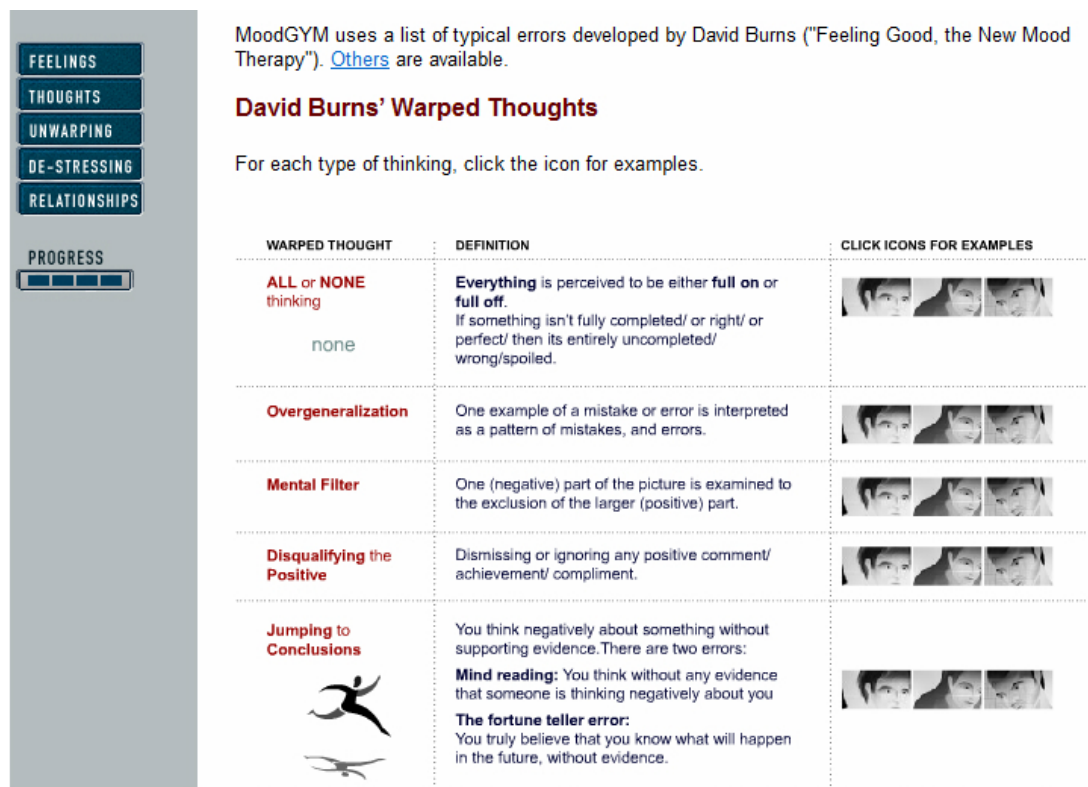
Some changes described by the patients revealed that their self-perception had tended to move in a positive direction, towards self-acceptance. This was evident from statements that revealed a more self-accepting and less self-critical attitude towards themselves:

If you have done something that you're not completely satisfied with, you should not think that you're a terrible person. That you can actually get a firm grip on it and work with it. [Female (41)]

Some of the changes patients made in their daily lives sprang out of the sessions with the therapist, but were not directly related to any of the program material. Examples might be practical solutions to problems discussed in the sessions, or specific strategies to break a negative cycle, designed specifically for the patient's own situation: "Yes, we had to find practical solutions, because it's not always that positive thinking can silence the negative." [Female (22)]

This is an illustration that patients could also experience other difficulties in their lives that needed attention. Examples are practical problems that needed to be tackled, in addition to the cognitive behavioral therapy provided in the self-help program.

Figure 1. Screenshot from module 2 in MoodGYM.



MoodGYM uses a list of typical errors developed by David Burns ("Feeling Good, the New Mood Therapy"). [Others](#) are available.

David Burns' Warped Thoughts

For each type of thinking, click the icon for examples.








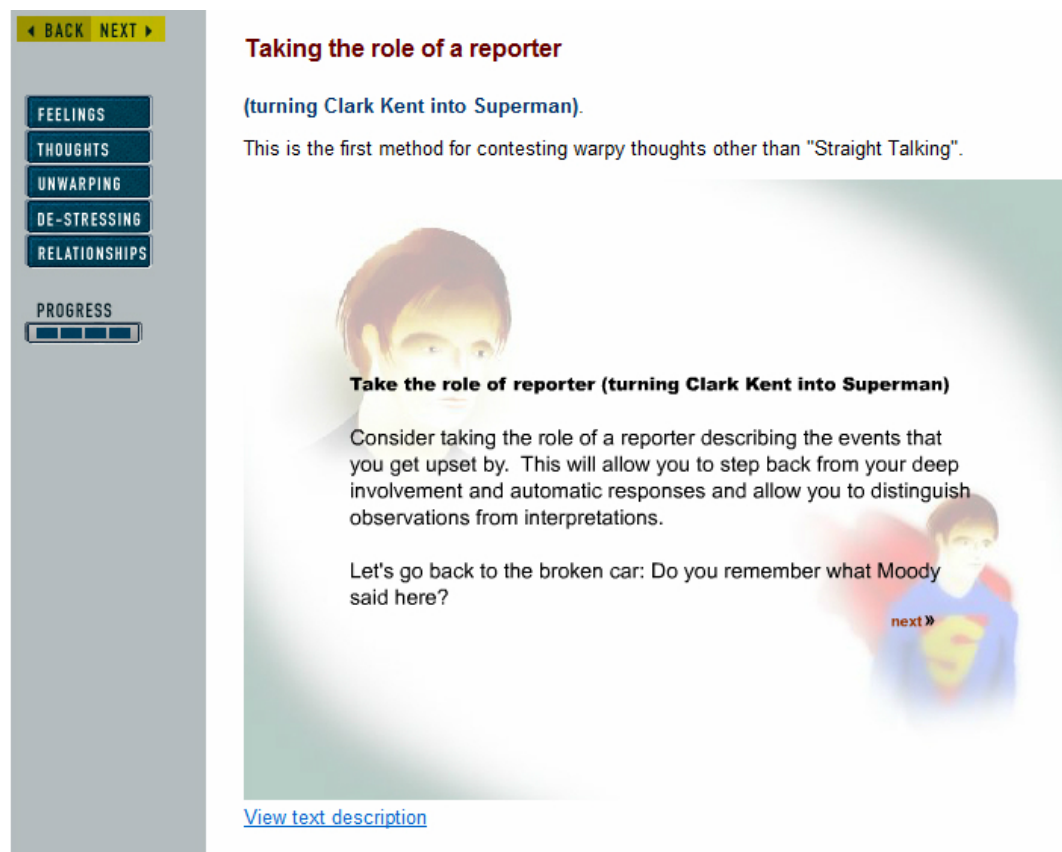
WARPED THOUGHT	DEFINITION	CLICK ICONS FOR EXAMPLES
ALL or NONE thinking none	Everything is perceived to be either full on or full off . If something isn't fully completed/ or right/ or perfect/ then its entirely uncompleted/ wrong/spoiled.	
Overgeneralization	One example of a mistake or error is interpreted as a pattern of mistakes, and errors.	
Mental Filter	One (negative) part of the picture is examined to the exclusion of the larger (positive) part.	
Disqualifying the Positive	Dismissing or ignoring any positive comment/ achievement/ compliment.	
Jumping to Conclusions  	You think negatively about something without supporting evidence. There are two errors: Mind reading: You think without any evidence that someone is thinking negatively about you The fortune teller error: You truly believe that you know what will happen in the future, without evidence.	

Figure 2. Screenshot from module 3 in MoodGYM.



← BACK NEXT →

FEELINGS
THOUGHTS
UNWARPING
DE-STRESSING
RELATIONSHIPS

PROGRESS

Taking the role of a reporter

(turning Clark Kent into Superman).

This is the first method for contesting warpy thoughts other than "Straight Talking".

Take the role of reporter (turning Clark Kent into Superman)

Consider taking the role of a reporter describing the events that you get upset by. This will allow you to step back from your deep involvement and automatic responses and allow you to distinguish observations from interpretations.

Let's go back to the broken car: Do you remember what Moody said here?

[View text description](#)

next »

Discussion

Principal Findings

The previous section has looked at what patients experienced as helpful with regards to treatment. These comprised basic dimensions as the very act of seeking help and being in a therapeutic relationship with a professional. Patients described their experiences of the specific CBT delivered by MoodGYM as a source of relevant knowledge that they adapted to their own situation and implemented in their everyday lives. Overall, the phenomenological-hermeneutical analysis gave knowledge of several experiences with ICBT concerning helpfulness but also provided a whole impression of how these experiences were related and understood. The patient identifies with and learns from MoodGYM material that stimulates them to engage in self-therapeutic activity. This process is supported through the consultations with the therapist. In addition, the relationship with the therapist had a function beyond supporting MoodGYM use, in providing an arena for sharing thoughts and feelings and receiving feedback and advice.

Regarding help-seeking, patients described that making a commitment to therapy was a great help suggesting that dimensions within the patient, not directly related to treatment content, have an influence on outcomes. Taking action to improve one's situation is a way of regaining control, referred to by previous authors as empowerment [16] or being compelled to take action [29,30]. Taking action would include making a commitment to the ICBT treatment and accepting its methodology, which again can be interpreted as an expression of the treatment relationship or the working alliance. The working alliance is shaped within the context of the therapy and represents the bond between therapist and patient, as well as a shared understanding of the work and goals of the treatment [13,31,32]. The patients were given an explanation of ICBT, and most accepted self-help therapy as a way forward with themselves being the primary agent of change. Yet there were also cases where the commitment was poor, and the patient did not accept the basic premises of ICBT or deemed the content of MoodGYM inappropriate to their needs. This resulted in poor motivation and little active engagement in the treatment. The success of the working alliance is strongly related to the success of the therapy outcome [33,34], which indicates that establishing a common understanding of the aims of the treatment is a vital key to recovery. There are findings pointing to the ability of patients to develop a meaningful relationship within a fully computerized treatment [16]. However, research findings concerning the working alliance in ICBT are inconclusive [18], and further research, particularly in relation to the type of support (face-to-face, telephone, or electronic messages) and treatment outcome is warranted.

The presence of a therapist was important in several ways. It means a trusting relationship can be developed, and patients have a channel for self-disclosure and supportive response. The need for disclosing thoughts and feelings, interaction, and feedback has been highlighted in previous studies [6,29,35], pointing to the role of the therapist in the therapeutic relationship as warm, empathic, affirming, and engaging [36]. In a previous

study, patients reported difficulties in translating computerized CBT content to their own social situation [6]. In the current study, sessions with the therapist added to the self-help program by opening up for discussion of the program content, assisting the patients in their understanding of the content. As such, the role of the therapist influenced the specific CBT dimensions. As in previous studies, some patients deemed the sessions confined and desired more conversation with a therapist to gain a deeper understanding of their problems [19,29]. Another benefit was that the therapist could step in when problems were beyond the scope of the program or tender helpful practical advice tailored to the individual. Clinicians consider a lack of these possibilities as potential drawbacks of ICBT [37]. Thus, patients and therapists value the flexibility and possibility to individually adapt the intervention in a way only human support can provide.

Acquiring new knowledge was a significant benefit from involvement in the MoodGYM program. Patients described themselves being able to relate to cases presented in the program. Patients actively sought out the parts of the program helpful to them and found ways to utilize this new knowledge. The findings show that these patients were not simply passive recipients of insights gleaned from MoodGYM, rather they were active seekers of relevant information, evaluated its validity, and adapted their new knowledge to their own personal situation. These findings are consistent with the perspective of patient involvement in psychotherapy [38] and with previous research where the self-help material stimulates new learning, making patients able to create more structure and order, and break problems down to smaller entities [6,16,19,30]. Within this perspective, the patient is an active agent, entering therapy with some ideas of what they need, selecting from therapy what they consider useful, making independent assessments of results, and integrating therapy experiences into everyday life.

Patients talked about self-therapeutic activities in terms of implementing their own adaptations of MoodGYM content and discussions with the therapist. Commonly, patients reported general insights and increased awareness of negative automatic thoughts and cognitive distortions in real life, similar to previous observations in ICBT [19] and face-to-face CBT [11]. For some patients, these insights enabled them to challenge the validity of such unruly thoughts. Changes in behavior and communication could bring secondary benefits, further strengthening the positive processes in play. Some changes were subtler, not being specifically bound to any given situation but seemed to represent a shift in self-esteem across all situations, with the patient becoming more self-accepting and less judgmental.

In light of the findings from the current qualitative study, guided ICBT with MoodGYM can be viewed as a dynamic process, in which the patient and the therapist work together within the self-help framework offered by the online treatment program.

Implications of the Study

The results of this study highlight how ICBT can be a useful treatment for depression by providing insight into everyday life experiences with this mode of treatment. This has brought forward both patient-near experiences and pragmatic solutions

to identified problems. The results specifically illustrate the dynamic interplay between the patient's lifeworld, the therapist, and the ICBT treatment program. This is important to keep in mind for the future development and implementation of guided self-help.

The collaborative nature of guided ICBT is evident from the results of this study. The patient plays an active role in the therapeutic process, and this point should be explicitly stated before and during the course of the treatment. Needless to say, individual patients have different capacities to get involved, some needing more support and encouragement than others.

The ease with which patients could adapt standard principles to suit their own situations varied a great deal. This realization should help inform continuing development of online self-help programs. In general, generic CBT modules were experienced as more relevant than modules targeting specific issues (eg, problem solving, relaxation), supporting previous findings [39]. Principles and descriptions in the program should be generic to the extent they are recognizable to the patient, yet designed to work with patients with a specific diagnosis [40], but adaptable to an everyday context. This precise diagnosis or description of the patient's problem needs to be prepared beforehand, allowing the patient to be matched to the most suitable program package [7].

Whether therapist support is provided through email, telephone calls, or face-to-face, it can be assumed to contribute to the outcome of the therapy directly and indirectly. Therapist support should offer the opportunity for a level of self-disclosure, in a nonjudgmental setting where patients can expect relevant, supportive feedback. The therapist also serves an important function to help patients understand the principles of CBT and help patients make the transformation from principle to everyday practice.

The role of the therapist was also vital when patients needed individual guidance or advice about how to deal with everyday challenges that might otherwise get in the way of treatment progress. Practical advice from a therapist, or an inspirational session with a sympathetic adult, could provide the motivation to embark on manageable, significant changes.

The findings of the present study point to the dimensions experienced as important by the patients interviewed, and the meaning of this mode of treatment in terms of moving from depression towards recovery. These findings may serve as a hint on how to continue to improve the practice of ICBT. It remains to explore what conditions in ICBT are sufficient and necessary to aid patients, and what distinguishes processes resulting in change from those ending in no change or deterioration.

Limitations

The sample of patients in this study is small, consisting of individuals who volunteered to join a randomized controlled trial and attend the interviews. Possibly they were highly motivated to participate in research, and they are in some way different from other potential candidates who chose not to be interviewed. Furthermore, recruiting noncompleters turned out to be difficult, and thus the completers are overrepresented in the sample. This is a limitation of the study. It is likely that completers and noncompleters have divergent views regarding the helpfulness of the treatment, and we recommend further research into the treatment experiences of noncompleters of ICBT. However, there were differences within the sample regarding outcome, and accordingly, the results also reflect the views of patients who did not improve.

The study examined patients' experiences with an online self-help program, incorporating occasional, brief face-to-face sessions with a therapist. By contrast, Internet-based therapy typically gives guidance through electronic messages or telephone calls. It is possible that the treatment examined in the current study has a greater resemblance to conventional psychotherapy than to Web-based self-help. Further investigation could identify other possible themes for discussion in other forms of Web-based self-help therapies.

The interviews were conducted shortly after the end of treatment, therefore it is not possible to evaluate to what extent potentially helpful elements impacted on patients' lives. It may be that patients were still "fired up" by the therapy at the time of the interview, but this does not necessarily mean they experienced any long-term improvements. Some findings suggest that patients can continue to engage in self-therapeutic CBT for as much as 3 months after ending treatment [11], but further research is needed to shed light on patient agency in CBT.

Conclusion

Elements in ICBT that are perceived as helpful represent the essence of the patients' experience with ICBT. This can be viewed from the perspective of the contextual model of psychotherapy, which highlights the dynamic and collaborative nature of Internet-based self-help. The findings of the current study pointed to MoodGYM as a source of relevant knowledge, providing a structured approach to working with depression. The role of the patient as the primary agent of change is highlighted, through his/her engagement in treatment, seeking knowledge, and employing it to the personal context. During the intervening guidance sessions, the therapist played a useful role by facilitating the understanding and explaining the relevance of the generic MoodGYM content, providing professional feedback and interpersonal support, as well as giving practical advice.

Acknowledgments

The authors gratefully acknowledge the help of the 14 participants in this study. Funding was provided by the Research Council of Norway (196423/V50). Ethical approval was received from the Regional Ethical Committee, Tromsø (2011/2163).

Conflicts of Interest

None declared.

References

1. Spek V, Cuijpers P, Nyklíček I, Smits N, Riper H, Keyzer J, et al. One-year follow-up results of a randomized controlled clinical trial on internet-based cognitive behavioural therapy for subthreshold depression in people over 50 years. *Psychol Med* 2008 May;38(5):635-639. [doi: [10.1017/S0033291707002590](https://doi.org/10.1017/S0033291707002590)] [Medline: [18205965](https://pubmed.ncbi.nlm.nih.gov/18205965/)]
2. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. *Cogn Behav Ther* 2009;38(4):196-205. [doi: [10.1080/16506070903318960](https://doi.org/10.1080/16506070903318960)] [Medline: [20183695](https://pubmed.ncbi.nlm.nih.gov/20183695/)]
3. Johansson R, Andersson G. Internet-based psychological treatments for depression. *Expert Rev Neurother* 2012 Jul;12(7):861-9; quiz 870. [doi: [10.1586/ern.12.63](https://doi.org/10.1586/ern.12.63)] [Medline: [22853793](https://pubmed.ncbi.nlm.nih.gov/22853793/)]
4. Cuijpers P, Andersson G. Special issue on computerized treatments of depression. *Cogn Behav Ther* 2009;38(4):193-195. [doi: [10.1080/16506070903302063](https://doi.org/10.1080/16506070903302063)] [Medline: [20183694](https://pubmed.ncbi.nlm.nih.gov/20183694/)]
5. Cuijpers P, Donker T, van Straten A, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. *Psychol Med* 2010 Dec;40(12):1943-1957. [doi: [10.1017/S0033291710000772](https://doi.org/10.1017/S0033291710000772)] [Medline: [20406528](https://pubmed.ncbi.nlm.nih.gov/20406528/)]
6. Gerhards SA, Abma TA, Arntz A, de Graaf LE, Evers SM, Huibers MJ, et al. Improving adherence and effectiveness of computerised cognitive behavioural therapy without support for depression: a qualitative study on patient experiences. *J Affect Disord* 2011 Mar;129(1-3):117-125. [doi: [10.1016/j.jad.2010.09.012](https://doi.org/10.1016/j.jad.2010.09.012)] [Medline: [20889214](https://pubmed.ncbi.nlm.nih.gov/20889214/)]
7. Andersson G, Carlbring P, Berger T, Almlöv J, Cuijpers P. What makes Internet therapy work? *Cogn Behav Ther* 2009;38 Suppl 1:55-60. [doi: [10.1080/16506070902916400](https://doi.org/10.1080/16506070902916400)] [Medline: [19675956](https://pubmed.ncbi.nlm.nih.gov/19675956/)]
8. Farrand P, Woodford J. Impact of support on the effectiveness of written cognitive behavioural self-help: a systematic review and meta-analysis of randomised controlled trials. *Clin Psychol Rev* 2013 Feb;33(1):182-195. [doi: [10.1016/j.cpr.2012.11.001](https://doi.org/10.1016/j.cpr.2012.11.001)] [Medline: [23238023](https://pubmed.ncbi.nlm.nih.gov/23238023/)]
9. Peck DF. The therapist-client relationship, computerized self-help and active therapy ingredients. *Clin Psychol Psychother* 2010;17(2):147-153. [doi: [10.1002/cpp.669](https://doi.org/10.1002/cpp.669)] [Medline: [19946947](https://pubmed.ncbi.nlm.nih.gov/19946947/)]
10. Clark DA, Beck AT. Scientific foundations of cognitive theory of depression. New York: John Wiley; 1999.
11. Glasman D, Finlay WM, Brock D. Becoming a self-therapist: using cognitive-behavioural therapy for recurrent depression and/or dysthymia after completing therapy. *Psychol Psychother* 2004 Sep;77(Pt 3):335-351. [doi: [10.1348/1476083041839385](https://doi.org/10.1348/1476083041839385)] [Medline: [15355585](https://pubmed.ncbi.nlm.nih.gov/15355585/)]
12. Wampold BE. The great psychotherapy debate: models, methods, and findings. Mahwah, N.J: L. Erlbaum Associates; 2001.
13. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice* 1979;16(3):252-260. [doi: [10.1037/h0085885](https://doi.org/10.1037/h0085885)]
14. Flückiger C, Del Re AC, Wampold BE, Symonds D, Horvath AO. How central is the alliance in psychotherapy? A multilevel longitudinal meta-analysis. *J Couns Psychol* 2012 Jan;59(1):10-17. [doi: [10.1037/a0025749](https://doi.org/10.1037/a0025749)] [Medline: [21988681](https://pubmed.ncbi.nlm.nih.gov/21988681/)]
15. Richards D. Self-help: Empowering service users or aiding cash strapped mental health services? *J Ment Health* 2004 Jan;13(2):117-123. [doi: [10.1080/09638230410001669246](https://doi.org/10.1080/09638230410001669246)]
16. Purves DG, Dutton J. An exploration of the therapeutic process while using computerised cognitive behaviour therapy. *Counselling and Psychotherapy Research* 2013 Jan 18:1-9. [doi: [10.1080/14733145.2012.761259](https://doi.org/10.1080/14733145.2012.761259)]
17. Warmerdam L, van Straten A, Jongsma J, Twisk J, Cuijpers P. Online cognitive behavioral therapy and problem-solving therapy for depressive symptoms: Exploring mechanisms of change. *J Behav Ther Exp Psychiatry* 2010 Mar;41(1):64-70. [doi: [10.1016/j.jbtep.2009.10.003](https://doi.org/10.1016/j.jbtep.2009.10.003)] [Medline: [19913781](https://pubmed.ncbi.nlm.nih.gov/19913781/)]
18. Andersson G, Paxling B, Wiwe M, Vernmark K, Felix CB, Lundborg L, et al. Therapeutic alliance in guided internet-delivered cognitive behavioural treatment of depression, generalized anxiety disorder and social anxiety disorder. *Behav Res Ther* 2012 Sep;50(9):544-550. [doi: [10.1016/j.brat.2012.05.003](https://doi.org/10.1016/j.brat.2012.05.003)] [Medline: [22728647](https://pubmed.ncbi.nlm.nih.gov/22728647/)]
19. Bendelin N, Hesser H, Dahl J, Carlbring P, Nelson KZ, Andersson G. Experiences of guided Internet-based cognitive-behavioural treatment for depression: a qualitative study. *BMC Psychiatry* 2011;11:107 [FREE Full text] [doi: [10.1186/1471-244X-11-107](https://doi.org/10.1186/1471-244X-11-107)] [Medline: [21718523](https://pubmed.ncbi.nlm.nih.gov/21718523/)]
20. Dahlberg K, Dahlberg H, Nyström M. Reflective Lifeworld Research. Reflective Lifeworld Research Lund: Studentlitteratur AB; 2010.
21. Lindseth A, Norberg A. A phenomenological hermeneutical method for researching lived experience. *Scand J Caring Sci* 2004 Jun;18(2):145-153. [doi: [10.1111/j.1471-6712.2004.00258.x](https://doi.org/10.1111/j.1471-6712.2004.00258.x)] [Medline: [15147477](https://pubmed.ncbi.nlm.nih.gov/15147477/)]
22. Groves C, Griffiths K, Christensen H. Working out MoodGYM: A user's guide. Canberra: Centre for Mental Health Research; 2003.
23. Christensen H, Griffiths KM, Korten A. Web-based cognitive behavior therapy: analysis of site usage and changes in depression and anxiety scores. *J Med Internet Res* 2002;4(1):e3 [FREE Full text] [doi: [10.2196/jmir.4.1.e3](https://doi.org/10.2196/jmir.4.1.e3)] [Medline: [11956035](https://pubmed.ncbi.nlm.nih.gov/11956035/)]

24. 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](#)]
25. Powell J, Hamborg T, Stallard N, Burls A, McSorley J, Bennett K, et al. Effectiveness of a web-based cognitive-behavioral tool to improve mental well-being in the general population: randomized controlled trial. *J Med Internet Res* 2013;15(1):e2 [[FREE Full text](#)] [doi: [10.2196/jmir.2240](#)] [Medline: [23302475](#)]
26. Beck AT, Steer RA, Brown G. *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation; 1996.
27. Ricoeur P. *Interpretation theory: discourse and the surplus of meaning*. Fort Worth: Texas Christian University Press; 1976.
28. Seggar LB, Lambert MJ, Hansen NB. Assessing clinical significance: Application to the beck depression inventory. *Behavior Therapy* 2002 Mar;33(2):253-269. [doi: [10.1016/S0005-7894\(02\)80028-4](#)]
29. Macdonald W, Mead N, Bower P, Richards D, Lovell K. A Qualitative Study of Patients' Perceptions of a 'Minimal' Psychological Therapy. *International Journal of Social Psychiatry* 2007 Jan 01;53(1):23-35. [doi: [10.1177/0020764006066841](#)]
30. Khan N, Bower P, Rogers A. Guided self-help in primary care mental health: meta-synthesis of qualitative studies of patient experience. *Br J Psychiatry* 2007 Sep;191:206-211 [[FREE Full text](#)] [doi: [10.1192/bjp.bp.106.032011](#)] [Medline: [17766759](#)]
31. Horvath AO, Luborsky L. The role of the therapeutic alliance in psychotherapy. *J Consult Clin Psychol* 1993 Aug;61(4):561-573. [Medline: [8370852](#)]
32. Hatcher RL, Barends A, Hansell J, Gutfreund MJ. Patients' and therapists' shared and unique views of the therapeutic alliance: an investigation using confirmatory factor analysis in a nested design. *J Consult Clin Psychol* 1995 Aug;63(4):636-643. [Medline: [7673541](#)]
33. Martin DJ, Garske JP, Davis MK. Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. *J Consult Clin Psychol* 2000 Jun;68(3):438-450. [Medline: [10883561](#)]
34. Horvath AO, Symonds BD. Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Counseling Psychology* 1991;38(2):139-149. [doi: [10.1037/0022-0167.38.2.139](#)]
35. Donkin L, Glozier N. Motivators and motivations to persist with online psychological interventions: a qualitative study of treatment completers. *J Med Internet Res* 2012;14(3):e91 [[FREE Full text](#)] [doi: [10.2196/jmir.2100](#)] [Medline: [22743581](#)]
36. Norcross JC. The therapeutic relationship. In: Duncan BL, editor. *The heartsoul of change: Delivering what works in therapy* (2nd ed.). Washington, DC: American Psychological Association; 2010:113-141.
37. MacLeod M, Martinez R, Williams C. Cognitive Behaviour Therapy self-help: who does it help and what are its drawbacks? *Behav Cogn Psychother* 2009 Jan;37(1):61-72. [doi: [10.1017/S1352465808005031](#)] [Medline: [19364408](#)]
38. Bohart AC, Tallman K. Clients: The neglected common factor in psychotherapy. In: Duncan BL, editor. *The heartsoul of change: Delivering what works in therapy* (2nd ed.). Washington, DC: American Psychological Association; 2010:83-111.
39. Christensen H, Griffiths KM, Mackinnon AJ, Brittliffe K. Online randomized controlled trial of brief and full cognitive behaviour therapy for depression. *Psychol Med* 2006 Dec;36(12):1737-1746. [doi: [10.1017/S0033291706008695](#)] [Medline: [16938144](#)]
40. Hind D, O'Cathain A, Cooper CL, Parry GD, Isaac CL, Rose A, et al. The acceptability of computerised cognitive behavioural therapy for the treatment of depression in people with chronic physical disease: a qualitative study of people with multiple sclerosis. *Psychol Health* 2010 Jul;25(6):699-712. [doi: [10.1080/08870440902842739](#)] [Medline: [20204955](#)]

Abbreviations

BDI: Beck Depression Inventory
CBT: cognitive behavioral therapy
ICBT: Internet-based cognitive behavioral therapy
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 17.01.13; peer-reviewed by D Cox, L Palmili, J Huber, P Farrand, H O'Mahen; comments to author 14.02.13; revised version received 27.03.13; accepted 08.06.13; published 20.06.13.

Please cite as:

Lillevoll KR, Wilhelmsen M, Kolstrup N, Høifødt RS, Waterloo K, Eisemann M, Risør MB
 Patients' Experiences of Helpfulness in Guided Internet-Based Treatment for Depression: Qualitative Study of Integrated Therapeutic Dimensions
J Med Internet Res 2013;15(6):e126
 URL: <http://www.jmir.org/2013/6/e126/>
 doi: [10.2196/jmir.2531](#)
 PMID: [23786763](#)

©Kjersti R Lillevoll, Maja Wilhelmsen, Nils Kolstrup, Ragnhild Sørensen Høifødt, Knut Waterloo, Martin Eisemann, Mette Bech Risør. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 20.06.2013. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Incidence of Online Health Information Search: A Useful Proxy for Public Health Risk Perception

Bo Liang¹, MS, MBA; Debra L Scammon¹, PhD

Department of Marketing, David Eccles School of Business, University of Utah, Salt Lake City, UT, United States

Corresponding Author:

Bo Liang, MS, MBA
Department of Marketing
David Eccles School of Business
University of Utah
1655 East Campus Center Drive
Salt Lake City, UT
United States
Phone: 1 480 309 3478
Fax: 1 801 581 3152
Email: bo.liang@utah.edu

Abstract

Background: Internet users use search engines to look for information online, including health information. Researchers in medical informatics have found a high correlation of the occurrence of certain search queries and the incidence of certain diseases. Consumers' search for information about diseases is related to current health status with regard to a disease and to the social environments that shape the public's attitudes and behaviors.

Objective: This study aimed to investigate the extent to which public health risk perception as demonstrated by online information searches related to a health risk can be explained by the incidence of the health risk and social components of a specific population's environment. Using an ecological perspective, we suggest that a population's general concern for a health risk is formed by the incidence of the risk and social (eg, media attention) factors related with the risk.

Methods: We constructed a dataset that included state-level data from 32 states on the incidence of the flu; a number of social factors, such as media attention to the flu; private resources, such as education and health insurance coverage; public resources, such as hospital beds and primary physicians; and utilization of these resources, including inpatient days and outpatient visits. We then explored whether online information searches about the flu (seasonal and pandemic flu) can be predicted using these variables. We used factor analysis to construct indexes for sets of social factors (private resources, public resources). We then applied panel data multiple regression analysis to exploit both time-series and cross-sectional variation in the data over a 7-year period.

Results: Overall, the results provide evidence that the main effects of independent variables—the incidence of the flu ($P<.001$); social factors, including media attention ($P<.001$); private resources, including life quality ($P<.001$) and health lifestyles ($P=.009$); and public resources, such as hospital care utilization ($P=.008$) and public health funds ($P=.02$)—have significant effects on Web searches for queries related to the flu. After controlling for the number of reported disease cases and Internet access rate by state, we estimate the contribution of social factors to the public health risk perception levels by state ($R^2=23.37\%$). The interaction effects between flu incidence and social factors for our search terms did not add to the explanatory power of our regression models ($R^2<1\%$).

Conclusions: Our study suggests a practical way to measure the public's health risk perception for certain diseases using online information search volume by state. The social environment influences public risk perception regardless of disease incidence. Thus, monitoring the social variables can be very helpful in being ready to respond to the public's behavior in dealing with public health threats.

(*J Med Internet Res* 2013;15(6):e114) doi:[10.2196/jmir.2401](https://doi.org/10.2196/jmir.2401)

KEYWORDS

health risk perception; social influence; ecological system

Introduction

The Internet has rapidly become an important source of health information: 61% of American Internet users have searched for health information online [1]. Most American Internet users primarily use search engines to look for information including health information [2-4]. Researchers in medical informatics have found a high correlation of the occurrence of certain search queries and the incidence of certain diseases, especially infectious diseases (eg, the flu), and thus have suggested the use of search query data for syndromic surveillance, or early detection of outbreaks [5-10]; this body of research has been well framed [11-16] and termed *infodemiology* by Eysenbach [6,17].

The existence of a correlation between search query volume and disease outbreaks raises a number of questions: Is the occurrence of certain search queries fully accounted for by the incidence of certain diseases? Do consumers search for online information related to a certain disease only when they have symptoms related to the disease? Are there situations in which consumers without any symptoms related to the disease search for online information related to the disease?

The perception of risk can play a role in many consumer decisions. Risk perception is the judgment that people make about the characteristics and severity of risks [18]. Over the past few decades, considerable research has been conducted on risk perception. The traditional theories of risk perception (eg, expected utility theory, prospect theory) were established by work in behavioral economics that focused on individuals' statistical or heuristic estimation of the value of alternative choices [19-22]. Understanding the risks perceived by individuals, and collectively by populations, is very helpful as the basis for designing effective strategies for communicating about risks. As a result, risk perception and risk communication have been used extensively in the field of public health [23,24].

Ecological systems theory holds that people interact with multiple social systems (eg, cultures, communities) in an environment [25,26]. Since its first introduction, ecological systems theory has been applied in various areas, such as health promotion [27]. From the ecological systems perspective, members of a specific population are influenced by the same sociocultural factors. Thus, their collective behavior is shaped by common factors.

We suggest that online information searches related to a health risk reflect the public's collective perception for the risk, which is associated not only with current health status (eg, the incidence of a disease), but also with the social environments related to the risk (eg, availability of public health resources) [28]. Our study extends previous work by exploring the association between online health information search and multiple sociocultural factors related to public health risk perception.

We selected online information searches related to the flu as the object of this study. The seasonal flu occurs on a regular basis; in the United States, an average of 5% to 20% of the population gets seasonal flu and more than 200,000 people are

hospitalized annually from seasonal flu-related complications. Thus, a significant proportion of the population has direct experience with the flu. Further, the flu can cause mild to severe illness and consumers are generally aware of the risks related to flu. In this study, we demonstrate the extent to which online information searches for the flu is explained by the incidence of the flu (including seasonal flu and pandemic flu) and sociocultural components of a population's environment. We suggest that the occurrence of online information search related to a health risk can be a practical way to assess the public's general concern for the risk, or public health risk perception.

The Ecological View of Risk Perception

From an ecological systems point of view, individuals grow and develop in different layered environmental systems, such as family, school, neighborhood, and community [25]. Because risk perception is a sociocultural construct [29,30], individuals form their perception of risks under the influence of the sociocultural components within these systems. Individuals form their risk perception through 2 types of experiences: direct (personal) and indirect (social) experience with the risk [30,31]. For a specific risk, some members of the population may directly experience the risk (eg, patients during a pandemic and victims of a natural disaster) whereas others may experience the risk only vicariously (eg, through exposure to media accounts of the event).

Both groups have social experience with the risk through sociocultural activities, such as receiving information about the risk from multiple social sources (eg, news media, personal social networks), and interpreting the information based on certain values or cultural biases [29-31]. Thus, individuals in a particular ecological system form their perceptions about a risk in response to a set of sociocultural components in the environment (eg, news coverage by mass media, demographics). Because these individuals share the same sociocultural environment, their risk perceptions have features in common. Individuals' risk perceptions are formed on the basis of a constellation of direct and indirect experience with a risk (sociocultural factors), and the interaction of the 2 types of experiences. The dynamic socioculturalization process through which individuals' risk perceptions evolve over time leads to the formation of population risk perception.

As an example, individuals who live in a specific community may form their perception about the risk of smoking influenced by their own experience with smoking and a set of shared sociocultural factors, such as the news coverage by the local and national media about the risks of smoking and the behavioral norms of other individuals in their community. As this dynamic process continues over time for individuals within a community, public risk perception for smoking will develop.

Building from this, our main proposition is that public risk perception is predicted by individuals' direct (personal) and indirect (social) experiences with a risk, and the interaction of the 2 experiences over time within an ecological system.

Previous research has identified a strong correlation between online information search and the incidence of diseases [5-10]. In the following, we establish our hypotheses related to the

social experiences that shape public risk perception. Building from previous literature on risk perception, especially health risk perception, we identify 2 major categories of social factors that are associated with public health risk perception: (1) news coverage by mass media, and (2) the availability of resources (including private and public resources).

Media Coverage

Agenda-setting theory proposes that mass media have an important influence on what issues the public considers to be important [32]. A number of studies have found powerful effects of mass media on individual risk perception [33-36]. For example, one study found that the number of news articles about the H1N1 pandemic was positively associated with individual preventive pharmaceutical intervention and engagement in information seeking [36]. It is important to note that media coverage may not always be factually correct. When the coverage of health risks by mass media is misleading (eg, exaggeration or stigmatization), the public may form misperceptions of the characteristics of the risks [37,38]. Regardless of the content of media coverage, the extent of coverage likely affects public risk perception.

We argue that when mass media pay more attention to a health risk by increasing coverage of the risk, the public will have higher awareness of the risk. Further, we suggest that when there is a high incidence of a health risk, the public will become more sensitive to the attention paid by media reporting. Their concern for the risk will be higher as the media attention increases. Thus, our first hypotheses are:

H1a: Online information search related to a health risk will be higher when mass media attention to the risk is higher;

H1b: The effect of the incidence of a health risk on online information search related to the risk will be greater when mass media attention to the risk is greater.

Availability of Resources

Research has shown that the availability of resources can reduce an individual's perceived risk [39-42]. We classify the resources related to health risk perception into 2 categories: private resources and public resources. Private resources are the resources that individuals can acquire through their own efforts, such as financial, informational, physiological, and physical resources. Studies have found that the availability of private resources is negatively associated with health risk perceptions. Three types of private resources are particularly important in the context of health risk perceptions: life quality (eg, education [43], family income [44]), health status [45,46], and health lifestyles (eg, tobacco and alcohol consumption [47]).

Public resources are the resources that the public obtains from organizations such as charities and government. Little research has been conducted on the association between public resource support and risk perception. We classify public resources into 4 groups: natural resources (eg, population density, especially risks that are related to natural disasters), financial resources (eg, funding for public health), capacity of public resources (eg, hospital beds), and utilization of public resources (eg, hospital admissions). In our study, we use capacity and utilization of

public health resources as measures of availability of health care resources. Because natural and financial resources and capacity and utilization of public resources are important resources with which the public can deal with health risks, we assume that the availability of these public resources is negatively associated with public health risk perception. Further, for those experiencing a health risk the availability of public resources may be particularly critical. Thus, we expect that their risk perception will be more likely to be influenced by the availability of resources.

We propose our second hypotheses:

H2a: Online information search for a health risk will be lower when the availability of private resources represented by life quality, health status, and health lifestyles, and public resources represented by natural and financial resources and capacity and utilization of public services is greater;

H2b: The effect of the incidence of a health risk on online information search related to the risk will be greater when the availability of private and public resources is lower.

Methods

Our study aims to explore the relationship between online information searches related to the flu and factors related to public health risk perception for the flu, including the incidence of the flu and the social factors related to the flu (news coverage and availability of resources). We used data from 2004 to 2011 from multiple published sources as detailed in the following section. The unit of analysis of our study is state population. In the following sections, we first detail the measures and data collection process for each variable necessary to test our hypotheses, and then present our analysis.

Measures

Online Information Search

Following methods used in previous studies [47-49], in this study we use Google Insights for Search (GIFS) to identify the changing patterns of Web searches used by consumers for queries related to the flu. Details of GIFS methodology are presented in [Multimedia Appendix 1](#).

Research has shown that Internet users usually include 1 or 2 terms in a search query [50]. Thus, for each of our search queries we include 1 or 2 related terms. People may have specific concerns related to prevention, diagnosis, and treatment of the flu. Thus, we preselected 96 search queries based on 3 categories: prevention (eg, flu shots, flu prevention), diagnosis (eg, flu symptoms, flu fever), and treatment of the flu (eg, Tamiflu). Our criterion for query selection was the availability of weekly search volume data for queries for 25 states in the United States (if there is not enough search volume for each query by state, GIFS shows only monthly search volume or no results). After checking the search volumes for these preselected queries, we identified 2 queries that fulfilled our criterion: flu shot(s) and flu symptom(s). The prevalence of the 2 search queries shows that the most common response by the public to the flu is to take preventive actions and to determine whether they have contracted the flu. In our study, we use the search

volumes for flu shot(s) and flu symptom(s) to represent the state population's risk perception for prevention and diagnosis of the flu, respectively.

Flu Incidence

Public health agencies in the United States often track the percentage of outpatient visits related to influenza-like illness (ILI), collected through the US Influenza Sentinel Provider Surveillance Network [51]. A high ILI percentage indicates that a large fraction of patients are experiencing flu-like symptoms. Based on previous studies of the correlation of Web search and flu surveillance [9,52] and the availability of data, we used weekly ILI outpatient visit rates to measure the weekly incidence of the flu by state (like search data, this measure is automatically normed for the state's population). We gathered these data from the official website of the Department of Health for each state. The data are not available for all states and all observed years. In all, the dataset includes the weekly ILI rate data for between 15 and 31 states over the time period of 2004 to 2011.

Mass Media Attention

Previous research on the influence of mass media on risk perception has used the number of news articles to measure mass media attention at the national level [36]. Because states vary in their population, we use the number of news articles per 1 million population to measure the relative media attention for state populations. We collected weekly data on the number of news articles by using the news search function in LexisNexis Academic [53], a comprehensive database of national and regional news media. To find news articles that focused on the topic *flu*, we set the search term in GIFS as flu and the restriction as "headline & lead." We set the time intervals as those used for search volume data from GIFS, and the sources of news as US newspapers and wires. We also set the article location (articles about a geographic location) as each state, indicating that the articles cover the population of a specific state. We collected annual state population data from the website of the [54]. The data for media attention covers all the states and all the observed years.

Private and Public Resources

For private resources, we have variables indicating health status, life quality, and health lifestyles. According to the Centers for Disease Control and Prevention (CDC), the population groups most vulnerable to the flu are young children under age 5 years, the population aged over 65 years, pregnant women, and the population with chronic diseases, such as HIV [55]. Based on the availability of data, we included 2 variables indicating age-related health status: the percentages of the population under 5 years and over 65 years, and 2 variables indicating chronic disease-related health status: the percentages of the population that have asthma and diabetes. We also included variables indicating life quality: the percentage of the population that has completed a bachelor's degree, median household income, the percentage of the population that reported good health status, and the health insurance coverage rate. For health lifestyles, we included variables indicating the percentage of the population

that used tobacco, exercised regularly, and was overweight or obese. Preventive health behavior is an important part of health lifestyles. Thus, we included a variable indicating the percentage of people over 65 years old who have had flu shots.

For public resources, we included variables indicating natural and financial resources, and capacity and utilization of public resources. Because the flu is a contagious respiratory disease, we used population density as a measure of natural resources. The flu is a health-related risk; therefore, we used public health funding as a measure of financial resources. We use the number of primary physicians per 1 million population as a measure of capacity of ambulatory care, outpatient visits per 1000 population as a measure of utilization of ambulatory care, the number of hospital beds per 1000 population as a measure of hospital care capacity, and hospital admissions, emergency room visits per 1000 population, and inpatient days as a measures of utilization of hospital care. We collected the data for private and public resources from the websites of the US Census Bureau [56], the CDC [57], Kaiser State Health Facts [58], and Trust for America's Health [59]. A majority of the data about health status is captured through state residents' self-report surveys conducted by the CDC. In all, our dataset for private and public sources includes 20 annual variables. The data for each variable are available for all states (Hawaii is an exception because the data were lacking for the years 2004 and 2005) and for at least 4 years for the observed time period. We present the details about our measures in Tables 1 and 2.

Analysis

To include as many observations and variables as possible, we use unbalanced panel data in our analyses. According to the CDC, the official annual flu season starts in October and ends in May covering 33 weeks [60]. As most of the states in our dataset have missing values for the incidence of the flu in some weeks outside of each flu season (especially in the years before the H1N1 flu pandemic occurred), we dropped the observations for these weeks. Because the Web search volume data has been normalized by the total Internet traffic from each respective state, we included control variables representing household Internet usage, the percentage of households with an Internet connection, and the percentage of households with an Internet connection through broadband for each state. These data were obtained from the website of the US Department of Commerce and were available for all of the states for 3 years: 2007, 2009, and 2010 [61].

To account for weekly variations in search volumes, we included 33 dummy variables to indicate the specific weeks in each flu season. We also observed that in the 2 flu seasons following the 2009 H1N1 flu pandemic, search volumes for flu-related queries were higher than in the flu seasons before the pandemic occurred. To account for this variation, we used a dummy variable to indicate the weeks before and after the H1N1 flu pandemic. We present the trends of the means of the search volumes for flu and the incidence of the flu for all the states across the 33 weeks in each flu season in Figure 1.

Table 1. Study variables: measures and types of data.

Variable	Measure	Type of data
Search for symptoms	Web search volume for flu symptom(s)	Nonofficial public databases
Search for shots	Web search volumes for flu shot (s)	Nonofficial public databases
Flu incidence	Influenza-like illness (ILI) outpatient visits	Official reports
Mass media attention	The number of news articles per 1 million population	Nonofficial public databases
Private resources		
Age-related health status		
Population <5 years	The percent of population <5 years	Official reports
Population >65 years	The percent of population >65 years	Official reports
Chronic disease-related health status		
Asthma	Percent of adults who have been diagnosed with asthma	Self-report surveys
Diabetes	Percent of adults who have been diagnosed with diabetes	Self-report surveys
Life quality		
Bachelor's degree	Percent of population that has completed a bachelor's degree	Official reports
Income	Median household income	Official reports
Good health	Percent of population that has fair or better health	Self-report surveys
Health insurance coverage	Percent of adults aged 18-64 who have any kind of health care coverage	Self-report surveys
Exercises	Percent of adults who indicated that they participated in physical activities during the past month	Self-report surveys
Health lifestyle		
Smoking	Percent of adults who are currently smokers	Self-report surveys
Overweight or obese	Percent of adults who are overweight or obese	Self-report surveys
Flu shots	Percent of adults ≥65 who have had a flu shot within the past year	Self-report surveys
Public resources		
Natural resources		
Population density	Population per square mile (land area)	Official reports
Financial resources		
Public funds	Total of public health funding per capita ^a	Official reports
Ambulatory care capacity		
Primary physicians	Number of primary care physicians per 1 million population ^b	Nonofficial public databases
Ambulatory care utilization		
Outpatient visits	Number of outpatient visits per 1000 population ^c	Nonofficial public databases
Hospital care capacity		
Hospital beds	Number of beds per 1000 population	Nonofficial public database
Hospital care utilization		
Hospital admissions	Number of hospital admissions per 1000 population ^c	Nonofficial public databases
Emergency department visits	Number of emergency department visits per 1000 population ^c	Nonofficial public databases
Inpatient days	Number of inpatient days per 1000 population ^c	Nonofficial public databases

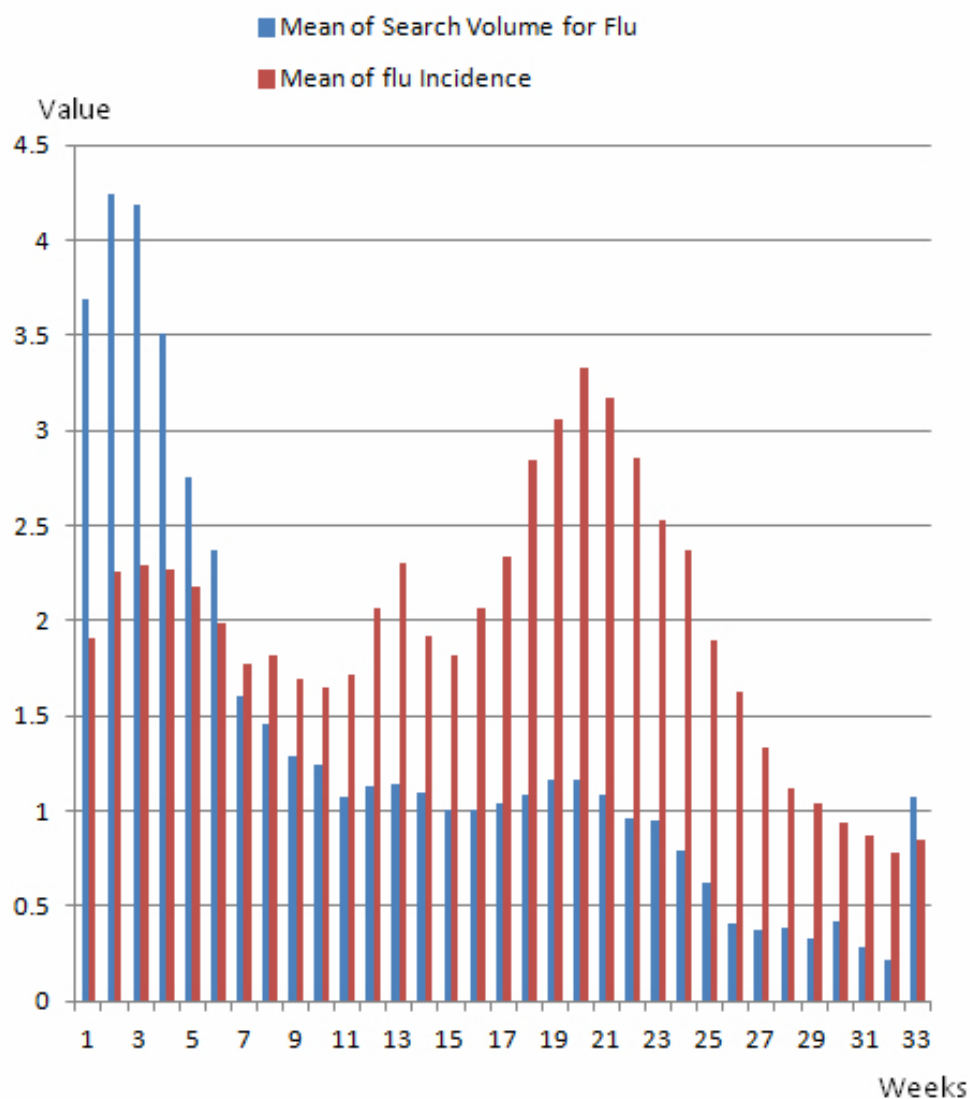
^aData includes funds from CDC and State health agencies.

^bData includes physicians for general practice, family practice, OB-GYN, pediatrics, and internal medicine.

^cData are for community hospitals, which represent 85% of all hospitals. Federal hospitals, long-term care hospitals, psychiatric hospitals, institutions for persons with mental disabilities, and hospitals for alcoholism and other chemical dependencies are not included.

Table 2. Study variables: data availability and sources.

Variable	Frequency data were collected	Number of years data are available	Data sources
Search for symptoms	Weekly	7	Google Insights for Search
Search for shots	Weekly	7	Google Insights for Search
Flu incidence	Weekly	7	Public health agencies
Media	Weekly	7	LexisNexis Academic
Private resources			
Age-related health status			
Population <5 years	Annually	7	US Census Bureau
Population >65 years	Annually	7	US Census Bureau
Chronic disease-related health status			
Asthma	Annually	7	CDC
Diabetes	Annually	7	CDC
Life quality			
Bachelor's degree	Annually	5	US Census Bureau
Income	Annually	7	US Census Bureau
Good health	Annually	7	CDC
Health insurance coverage	Annually	7	CDC
Exercises	Annually	7	CDC
Health lifestyle			
Smoking	Annually	7	CDC
Overweight or obese	Annually	7	CDC
Flu shots	Annually	7	CDC
Public resources			
Natural resources			
Population density	Annually	7	US Census Bureau
Financial resources			
Public funds	Annually	5	Trust for America's Health
Ambulatory care capacity			
Primary physicians	Annually	6	Henry J Kaiser Family Foundation
Ambulatory care utilization			
Outpatient visits	Annually	6	Henry J Kaiser Family Foundation
Hospital care capacity			
Hospital beds	Annually	6	Henry J Kaiser Family Foundation
Hospital care utilization			
Hospital admissions	Annually	6	Henry J Kaiser Family Foundation
Emergency department visits	Annually	6	Henry J Kaiser Family Foundation
Inpatient days	Annually	6	Henry J Kaiser Family Foundation

Figure 1. Trends of the means of search volumes for flu and the incidence of the flu.

Data Transformation

We normalized the data for all continuous variables included in the models by natural log transformations. To make the coefficients for interaction effects more interpretable, we centered all the continuous independent variables by subtracting the mean from each value.

Factor Analysis

Table 3 presents summary statistics for all dependent and independent variables included in the study. We used Stata (StataCorp LP, College Station, TX, USA) to perform factor analysis with varimax rotation to reduce the number of independent variables. Because factor analysis by Stata is conducted on the correlations (as opposed to the covariances), it is not a concern that the variables have different means and/or standard deviations (eg, variables are measured in different scales). Based on the components identified by factor analysis, 5 composite indexes were constructed: (1) life quality index, (2) age-related health status index, (3) chronic disease-related health status index, (4) health lifestyle index, and (5) hospital

care utilization index. We used the values of the composite indexes generated by factor analysis in our regression models. The life quality index includes positive factors indicating the percentage of the population with a bachelor's degree, good health status, health insurance, and median household income. The age-related health status index includes a positive factor indicating the percentage of the population younger than 5 years and a negative factor indicating the percentage of the population older than 65 years. The chronic disease-related health status index includes positive factors indicating the percentage of the population that has been diagnosed with asthma and diabetes. The health lifestyle index includes positive factors indicating the percentage of the population that consumes tobacco and are overweight or obese, and negative factors indicating the percentage of the population that exercises regularly and the percentage of the population older than 65 years who have had a flu shot. The hospital care utilization index includes positive factors indicating the number of hospital admissions, emergency department visits, and inpatient days. All other variables are represented by single data items. We present factor loadings and the uniqueness for each variable in Table 4.

Model Construction

We applied panel data multiple regression analysis to exploit both time-series and cross-sectional variation in the data using Stata. We built regression models to examine the effects of the incidence of flu, social factors including media attention, and private and public resources and their interaction on the Web search volumes for the 2 queries: flu symptom(s) and flu shot(s). Because we assumed that variation across states was random and uncorrelated with the independent variables, we used random state-specific effects in our models. We also used robust Huber–White standard errors to address any potential heteroscedasticity and autocorrelation in our estimation. To avoid collinearity, we examined the correlation matrix of independent variables (Table 5). We found that each pair of variables has a correlation coefficient of less than 0.8 with most less than 0.6.

To investigate the separate effects of control variables and independent variables on dependent variables, we ran 6 models for each dependent variable (each search query) sequentially as shown in Table 6.

Results

Overall, the results provide substantial evidence that the main effects of the independent variables we analyzed—the incidence of the flu, media attention, and private and public resources—have significant affects on Web search for queries related to the flu. Specifically, the results provide full support for hypotheses H1a and H1b, and partial support for hypotheses H2a and H2b.

Control Variables

The models including control variables (Models 1) were significant ($P<.001$) with coefficient of determination (R^2) value of 37.88% and 59.11% for flu symptoms and flu shots, respectively. Specifically, the dummy variables for the occurrence of H1N1 and seasonality were significant in the models for both search queries. However, with the independent variables sequentially added in Models 2 to Models 6, H1N1 occurrence showed significant effects only in the model for flu symptoms, but not flu shots. Seasonality contributed significantly to the variance of Web search for flu shots. We present the results for the control variables in Tables 7 and 8.

Flu Incidence and Media Attention

With flu incidence as an independent variable added in model 2, the r^2 values increased slightly by approximately 2% from

model 1 for flu symptoms ($P<.001$) and flu shots ($P<.001$). Further, with media attention as an independent variable added in model 3, r^2 values increased approximately 7% and 1% from model 2 for flu symptoms ($P<.001$) and ($P<.001$) flu shots, respectively. With the interaction of flu incidence and media attention added in model 4, the R^2 values showed an increase of less than 1% from model 3 for flu symptoms ($P<.001$) and flu shots ($P=.003$).

All these models were significant ($P<.001$) with flu incidence ($P<.001$), media attention ($P<.001$), and their interaction ($P<.001$ for flu symptoms; $P=.001$ for flu shots) showing positive effects on Web search volume. The changes in R^2 values from model 3 to model 4 show that media attention has a stronger positive influence on a population's search for flu symptoms than for flu shots.

Private and Public Resources

With variables indicating private and public resources added in models 5, the r^2 values showed a substantial increase from model 4 of 23.37% for flu symptoms ($P<.001$) and of 6.28% for flu shots ($P<.001$). For private resources, the life quality index ($P=.001$), health lifestyle index ($P=.009$), and chronic disease index ($P=.004$) had negative effects on search volume for flu symptoms.

For public resources, the number of outpatient visits ($P<.001$) and hospital care utilization index ($P=.008$) had positive effects, and the number of hospital beds had negative effects ($P<.001$) on search volume for flu symptoms. Public health funds ($P=.02$) had a negative effect, whereas population density ($P=.001$) and number of primary physicians ($P=.006$) had positive effects on search volume for flu shots.

With the interaction of these social factors and flu incidence added in models 6, the R^2 values increased slightly from model 5, 1.2% for flu symptoms ($P<.001$), and 0.55% for flu shots ($P=.004$). The interaction of flu incidence and number of primary physicians had a negative effect on search volume for flu symptoms. The interaction of flu incidence and life quality index had a negative effect on search volume for flu shots. We present the results for the changes in the R^2 values and the coefficient results for the independent variables in Tables 9 and 10.

Table 3. Summary statistics.

Variable	Observations (n)	Mean	SD	Min	Max
Dependent variables					
Search for symptoms	5364	2.89	6.14	0.00	70.00
Search for shots	8217	4.29	9.12	0.00	100.00
Independent variables					
Flu incidence	5226	2.04	2.10	0.00	20.43
Media	8364	0.95	1.69	0.00	31.12
Life quality					
Bachelor's degree	6006	27.98	4.35	18.70	38.20
Income	8217	49788.35	7725.76	32875.00	68059.00
Health insurance coverage	8184	85.39	4.42	71.50	95.70
Good health	8184	84.78	4.73	35.80	90.00
Age-related health status					
Population <5 years	8217	6.65	0.78	5.23	9.69
Population >65 years	8217	12.82	1.57	8.61	17.34
Chronic disease-related health status					
Asthma	8151	8.454	1.167	5.900	11.100
Diabetes	8184	3.96	0.93	0.00	4.62
Health lifestyle					
Exercise	8184	76.59	3.85	66.60	85.80
Smoking	8184	18.77	3.17	9.10	26.10
Overweight and obesity	8184	36.43	1.42	25.60	40.70
Flu shots	8151	69.99	4.29	55.60	80.00
Natural resources					
Population density	8184	224.88	293.28	5.24	1185.32
Financial resources					
Public funds	6006	56.81	31.91	18.39	197.76
Ambulatory care capacity					
Primary physicians	7128	120.69	27.07	78.50	191.30
Ambulatory care utilization					
Outpatient visits	7128	2134.60	676.25	942.00	4370.00
Hospital care capacity					
Hospital beds	7128	2.76	0.83	1.70	5.60
Hospital care utilization					
Hospital admissions	7128	114.49	17.37	81.00	154.00
Emergency department visits	7128	401.94	72.33	258.00	601.00
Inpatient days	7128	652.81	173.13	360.00	1210.00

Table 4. Composite indexes from factor analysis.

Index and variable		Loading	Uniqueness ^a
Life quality			
	Bachelor's degree	0.89	0.21
	Good health	0.54	0.71
	Income	0.91	0.17
	Health insurance coverage	0.79	0.37
Age-related health status			
	Population <5 years	−0.93	0.13
	Population >65 years	0.93	0.13
Chronic disease-related health status			
	Asthma	0.75	0.44
	Diabetes	0.75	0.44
Health lifestyle			
	Exercise	−0.81	0.34
	Smoking	0.71	0.49
	Overweight and obesity	0.74	0.45
	Flu shots	−0.52	0.73
Hospital care utilization			
	Hospital admissions	0.93	0.13
	ED visits	0.75	0.44
	Inpatient days	0.87	0.24

^aUniqueness is defined as $1 - \text{communality}$, or the portion not explained by common factor analysis.

Table 5. Correlation matrix of independent variables.

Independent variable	1	2	3	4	5	6	7	8	9	10	11	12
1 Flu incidence	1.00											
2 Media	0.16	1.00										
3 Public funds	0.03	0.11	1.00									
4 Population density	−0.02	−0.04	0.05	1.00								
5 Life quality	−0.23	0.09	0.24	0.41	1.00							
6 Age-related health status	−0.18	−0.02	0.03	0.33	0.21	1.00						
7 Chronic disease-related health status	−0.21	0.03	0.12	0.14	0.35	0.33	1.00					
8 Health lifestyle	0.21	−0.10	−0.21	0.01	−0.62	0.19	−0.28	1.00				
9 Primary physicians	−0.13	0.03	0.27	0.64	0.70	0.48	0.43	−0.30	1.00			
10 Outpatient visits	−0.24	0.07	−0.01	−0.06	0.38	0.44	0.43	−0.05	0.30	1.00		
11 Hospital beds	0.07	0.03	0.00	−0.33	−0.19	0.30	−0.26	0.53	−0.17	0.30	1.00	
12 Hospital care utilization	−0.02	−0.02	−0.09	0.17	−0.04	0.57	−0.07	0.60	0.21	0.40	0.80	1.00

Table 6. Model construction.

Model	Control variables	Flu incidence	Media	Flu incidence \times media	Other social factors	Flu incidence \times other social factors
1	Included					
2	Included	Included				
3	Included	Included	Included			
4	Included	Included	Included	Included		
5	Included	Included	Included	Included	Included	
6	Included	Included	Included	Included	Included	Included

Table 7. Coefficients for control variables for flu symptom(s).

Control variable	Flu symptom(s)					
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
H1N1 mark	1.666 ^a	1.956 ^a	1.590 ^a	1.483 ^a	1.027 ^a	0.823 ^a
Internet connection	-5.397 ^a	1.779	3.69	5.237	0.918	1.339
Broadband connection	-0.024	-5.050 ^a	-4.516 ^a	-4.748 ^a	-0.586	-0.105
Flu season week						
1	-0.980 ^a	-0.819 ^a	-0.710 ^a	-0.847 ^a	-0.536 ^a	-0.568 ^a
2	-0.843 ^a	-0.756 ^a	-0.608 ^a	-0.771 ^a	-0.396 ^a	-0.454 ^a
3	-0.790 ^a	-0.684 ^a	-0.562 ^a	-0.744 ^a	-0.439 ^a	-0.477 ^a
4	-0.877 ^a	-0.829 ^a	-0.608 ^a	-0.741 ^a	-0.422 ^a	-0.439 ^a
5	-0.875 ^a	-0.803 ^a	-0.565 ^a	-0.640 ^a	-0.351 ^a	-0.348 ^a
6	-0.845 ^a	-0.792 ^a	-0.538 ^a	-0.602 ^a	-0.343 ^a	-0.339 ^a
7	-1.036 ^a	-0.904 ^a	-0.634 ^a	-0.655 ^a	-0.445 ^a	-0.431 ^a
8	-1.010 ^a	-0.859 ^a	-0.631 ^a	-0.625 ^a	-0.405 ^a	-0.400 ^a
9	-1.057 ^a	-0.849 ^a	-0.680 ^a	-0.643 ^a	-0.523 ^a	-0.504 ^a
10	-1.123 ^a	-0.883 ^a	-0.667 ^a	-0.648 ^a	-0.559 ^a	-0.527 ^a
11	-1.099 ^a	-0.874 ^a	-0.676 ^a	-0.595 ^a	-0.630 ^a	-0.580 ^a
12	-0.992 ^a	-0.897 ^a	-0.568 ^a	-0.547 ^a	-0.371 ^a	-0.345 ^a
13	-1.031 ^a	-1.014 ^a	-0.678 ^a	-0.630 ^a	-0.515 ^a	-0.492 ^a
14	-1.203 ^a	-1.098 ^a	-0.823 ^a	-0.769 ^a	-0.646 ^a	-0.592 ^a
15	-1.187 ^a	-1.078 ^a	-0.742 ^a	-0.703 ^a	-0.580 ^a	-0.546 ^a
16	-1.061 ^a	-1.042 ^a	-0.703 ^a	-0.670 ^a	-0.558 ^a	-0.519 ^a
17	-0.798 ^a	-1.112 ^a	-0.730 ^a	-0.631 ^a	-0.505 ^a	-0.480 ^a
18	-0.606 ^a	-1.056 ^a	-0.713 ^a	-0.657 ^a	-0.496 ^a	-0.462 ^a
19	-0.660 ^a	-1.184 ^a	-0.755 ^a	-0.642 ^a	-0.421 ^a	-0.371 ^a
20	-0.662 ^a	-1.191 ^a	-0.799 ^a	-0.705 ^a	-0.461 ^a	-0.416 ^a
21	-0.835 ^a	-1.317 ^a	-0.964 ^a	-0.910 ^a	-0.598 ^a	-0.559 ^a
22	-1.071	-1.395 ^a	-0.931 ^a	-0.856 ^a	-0.578 ^a	-0.549 ^a
23	-1.334 ^a	-1.533 ^a	-1.078 ^a	-1.000 ^a	-0.837 ^a	-0.807 ^a
24	-1.454 ^a	-1.601 ^a	-1.171 ^a	-1.121 ^a	-0.902 ^a	-0.889 ^a
25	-1.618 ^a	-1.582 ^a	-1.246 ^a	-1.220 ^a	-1.196 ^a	-1.169 ^a
26	-1.852 ^a	-1.748 ^a	-1.306 ^a	-1.284 ^a	-1.274 ^a	-1.245 ^a
27	-2.063 ^a	-1.804 ^a	-1.307 ^a	-1.354 ^a	-1.302 ^a	-1.292 ^a
28	-2.199 ^a	-1.873 ^a	-1.450 ^a	-1.509 ^a	-1.506 ^a	-1.487 ^a
29	-2.372 ^a	-1.912 ^a	-1.510 ^a	-1.600 ^a	-1.579 ^a	-1.576 ^a
30	-2.499 ^a	-2.034 ^a	-1.649 ^a	-1.735 ^a	-1.815 ^a	-1.790 ^a
31	-2.531 ^a	-1.986 ^a	-1.534 ^a	-1.693 ^a	-1.729 ^a	-1.762 ^a

Control variable	Flu symptom(s)					
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
32	-2.658 ^a	-2.117 ^a	-1.694 ^a	-1.880 ^a	-2.016 ^a	-2.045 ^a
33	-2.908 ^a	-2.225 ^a	-1.761 ^a	-1.992 ^a	-2.071 ^a	-2.094 ^a

^a $P < .05$.

Table 8. Coefficients for control variables for flu shot(s).

Control variable	Flu shot(s)					
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
H1N1 mark	0.606 ^a	0.493	0.319	0.286	0.091	0.018
Internet connection	−0.887	−1.419	−0.745	−0.376	3.958	0.806
Broadband connection	0.312	0.478	0.832	0.821	−0.049	1.338
Flu season week						
1	2.459 ^a	2.254 ^a	2.292 ^a	2.237 ^a	2.297 ^a	2.231 ^a
2	2.728 ^a	2.355 ^a	2.416 ^a	2.333 ^a	2.469 ^a	2.422 ^a
3	2.612 ^a	2.212 ^a	2.265 ^a	2.163 ^a	2.359 ^a	2.308 ^a
4	2.342 ^a	1.896 ^a	2.000 ^a	1.925 ^a	2.101 ^a	2.060 ^a
5	2.244 ^a	1.887 ^a	2.008 ^a	1.966 ^a	2.222 ^a	2.193 ^a
6	1.811 ^a	1.525 ^a	1.668 ^a	1.631 ^a	2.024 ^a	2.007 ^a
7	1.102 ^a	0.931 ^a	1.066 ^a	1.056 ^a	1.457 ^a	1.437 ^a
8	1.032 ^a	0.865 ^a	1.004 ^a	1.002 ^a	1.432 ^a	1.400 ^a
9	0.654 ^a	0.587 ^a	0.687 ^a	0.694 ^a	0.935 ^a	0.906 ^a
10	0.383 ^a	0.423 ^a	0.519	0.541 ^a	0.758	0.723 ^a
11	−0.201 ^a	−0.218 ^a	−0.11	−0.075	−0.108	−0.139
12	−0.392 ^a	−0.517 ^a	−0.333	−0.33	−0.371	−0.393
13	−0.312 ^a	−0.498 ^a	−0.306	−0.29	−0.396	−0.407
14	−0.393 ^a	−0.484 ^a	−0.333	−0.307	−0.437 ^a	−0.44
15	−0.654 ^a	−0.722 ^a	−0.556 ^a	−0.530 ^a	−0.647 ^a	−0.659 ^a
16	−0.884 ^a	−1.038 ^a	−0.847 ^a	−0.829 ^a	−0.976 ^a	−1.005 ^a
17	−0.984 ^a	−1.255 ^a	−1.061 ^a	−1.011 ^a	−1.188 ^a	−1.193 ^a
18	−1.182 ^a	−1.532 ^a	−1.339 ^a	−1.296 ^a	−1.386 ^a	−1.367 ^a
19	−1.491 ^a	−1.913 ^a	−1.673 ^a	−1.613 ^a	−1.684 ^a	−1.649 ^a
20	−1.551 ^a	−2.017 ^a	−1.808 ^a	−1.766 ^a	−1.845 ^a	−1.814 ^a
21	−1.759 ^a	−2.240 ^a	−2.025 ^a	−2.000 ^a	−2.014 ^a	−1.978 ^a
22	−2.096 ^a	−2.584 ^a	−2.318 ^a	−2.284 ^a	−2.344 ^a	−2.313 ^a
23	−2.106 ^a	−2.550 ^a	−2.280 ^a	−2.250 ^a	−2.263 ^a	−2.251 ^a
24	−2.253 ^a	−2.653 ^a	−2.392 ^a	−2.377 ^a	−2.359 ^a	−2.360 ^a
25	−2.300 ^a	−2.629 ^a	−2.427 ^a	−2.409 ^a	−2.489 ^a	−2.497 ^a
26	−2.300 ^a	−2.564 ^a	−2.312 ^a	−2.302 ^a	−2.386 ^a	−2.410 ^a
27	−2.371 ^a	−2.550 ^a	−2.281 ^a	−2.292 ^a	−2.268 ^a	−2.306 ^a
28	−2.394 ^a	−2.527 ^a	−2.281 ^a	−2.313 ^a	−2.271 ^a	−2.303 ^a
29	−2.418 ^a	−2.465 ^a	−2.229 ^a	−2.270 ^a	−2.190 ^a	−2.200 ^a
30	−2.441 ^a	−2.413 ^a	−2.179 ^a	−2.233 ^a	−2.166 ^a	−2.212 ^a
31	−2.488 ^a	−2.446 ^a	−2.162 ^a	−2.256 ^a	−2.173 ^a	−2.203 ^a

Control variable	Flu shot(s)					
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
32	-2.488 ^a	-2.347 ^a	-2.080 ^a	-2.182 ^a	-2.167 ^a	-2.222 ^a
33	-2.512 ^a	-2.299 ^a	-2.013 ^a	-2.142 ^a	-2.112 ^a	-2.187 ^a

^a $P < .05$.

Table 9. Coefficient of determination (R^2) change and coefficient results for dependent variable flu symptom(s).

Independent variables	Model 2		Model 3		Model 4		Model 5		Model 6	
	Coefficient	<i>P</i>	Coefficient	<i>P</i>	Coefficient	<i>P</i>	Coefficient	<i>P</i>	Coefficient	<i>P</i>
Flu and media										
Flu incidence	0.644	<.001	0.528	<.001	0.498	<.001	0.371	<.001	0.473	<.001
Media			0.293	<.001	0.241	<.001	0.334	<.001	0.323	<.001
Flu incidence × media					0.194	<.001	0.177	<.001	0.181	<.001
Private resources										
Life quality							-0.401	.001	-0.389	.004
Age							-0.143	.06	-0.208	.05
Chronic disease							-0.279	.004	-0.309	.002
Health lifestyles							-0.287	.009	-0.205	.02
Public resources										
Public funds							-0.315	.05	-0.375	.04
Population density							0.188	.23	0.169	.14
Primary physicians							0.123	.06	0.106	.07
Outpatient visits							1.455	<.001	1.523	<.001
Hospital beds							-3.356	<.001	-3.601	<.001
Hospital care utilization							0.601	.008	0.651	.02
Interactions										
Flu × life quality									0.001	.03
Flu incidence × age									0.173	.08
Flu incidence × chronic disease									0.109	.09
Flu incidence × healthy lifestyles									0.072	.30
Flu incidence × public funds									-0.018	.88
Flu incidence × population density									-0.095	.37
Flu incidence × primary physician									0.150	.03
Flu incidence × outpatient visits									-0.666	.52
Flu incidence × hospital beds									0.708	.08
Flu incidence × hospital care utilization									-0.115	.43

Table 10. Coefficient of determination (R^2) change and coefficient results for dependent variable flu shot(s).

Independent variables	Model 2		Model 3		Model 4		Model 5		Model 6	
	Coefficient	P	Coefficient	P	Coefficient	P	Coefficient	P	Coefficient	P
Flu and media										
Flu incidence	0.477	<.001	0.414	<.001	0.383	<.001	0.251	<.001	0.173	.04
Media			0.162	<.001	0.144	<.001	0.189	<.001	0.206	<.001
Flu incidence \times media					0.089	.001	0.112	<.001	0.128	.001
Private resources										
Life quality							−0.070	.43	−0.018	.32
Age							−0.013	.27	−0.008	.29
Chronic disease							−0.083	.41	−0.036	.51
Health lifestyles							0.114	.62	0.078	.73
Public resources										
Public funds							−0.413	.02	−0.544	.02
Population density							0.191	.001	0.117	.05
Primary physicians							1.439	.006	1.561	.01
Outpatient visits							0.022	.08	−0.234	.13
Hospital beds							−0.662	.64	−0.918	.72
Hospital care utilization							−0.05	.19	0.073	.05
Interactions										
Flu \times life quality									−0.223	.03
Flu incidence \times age									0.045	.08
Flu incidence \times chronic disease									0.018	.54
Flu incidence \times healthy lifestyles									0.018	.43
Flu incidence \times public funds									−0.165	.09
Flu incidence \times population density									0.073	.06
Flu incidence \times primary physicians									0.823	.47
Flu incidence \times outpatient visits										0.0112
Flu incidence \times hospital beds									0.808	.06
Flu incidence \times hospital care utilization									−0.281	.43

Discussions

Principal Results

Research on the correlation between the incidence of certain diseases and online information searches related to those diseases has increased in recent years. However, there has been little research on the effects of social factors on online information searches related to disease. In this paper, we demonstrate the usefulness of online search for queries related to a health risk as a measure of public health risk perception. We use publicly available data to demonstrate how such data can be used to provide insights into the factors that influence the public's perception of health risks.

The results of our regression analyses provide strong support for our hypotheses: Web search volumes for flu-related queries as a measure of public health risk perception is predicted by the

incidence of the flu and social factors, including media attention, and private and public resources. In addition to the independent impact of these social variables, we anticipated that the effect of incidence of the flu on public risk perception would be heightened by factors in the social environment. However, our models that incorporated the interaction effects between flu incidence and social factors did not add much to the explanatory power of our regression models. The social environment affects public health risk perception regardless of the incidence of diseases.

We modeled information searches for both risk prevention (flu shots) and risk diagnosis (flu symptoms). In our analyses, independent variables, especially media attention and private and public resources, had significant influence on search volumes for flu symptoms; however, seasonality variables had significant influence on search volumes for flu shots. As we

anticipated, different factors appear to influence public perception of risk diagnosis and risk prevention.

Both the model for flu symptoms and that for flu shots demonstrate positive main effects for the incidence of the flu on search volumes. When a population's flu incidence is higher, the population's concerns for both prevention and diagnosis of the risk are higher.

Our data also support the expected positive effects of media attention and its interaction with the incidence of the flu on search volumes for risk prevention and risk diagnosis. Because mass media pays more attention to the risks related to a specific population, the overall population and the population with the flu both have more concern for prevention and diagnosis of the risk. Thus, our results show that the media play a significant role in setting the public agenda for health risk (agenda-setting theory [32]).

Private resources represented by life quality and public resources represented by hospital beds were negatively related to search volumes for flu symptoms. For the risk of the flu, a population with higher life quality and more access to hospital services demonstrated less searches for symptoms, whereas a population with lower life quality and less access to hospital services demonstrated more searches for symptoms. These results suggest consumers may use information from the Internet as a substitute for health care resources. Specifically, consumers vulnerable because of lower life quality and less available hospital services engage in more Internet searches, perhaps because information available on the Internet represents a relatively low-cost and easy-access source for information related to health risks.

With regard to private and public resources related specifically to health, our analyses suggest that there may be some synergistic effects of the 2 types of resources. Private resources represented by healthy lifestyles and public resources represented by outpatient visits and hospital care utilization are positively related to search volume for flu symptoms. Further, public resources represented by primary physicians and public funds were positively related to search volume for flu shots.

Based on these findings, we suggest that when a population has healthier lifestyles and more contact with health care professionals (through outpatient visits, emergency department visits, inpatient stays, availability of primary care physicians, and dedicated public health funds), it may be more conscious about current health risks. These results raise a question about the direction of the relationship between access to health care professionals and consumers' searches for health information on the Internet. It could be that access to health care professionals stimulates consumers to be more vigilant about risk protective behaviors. If this is the case, primary care physicians and public health agencies play an important role in educating the public to take protective actions.

Private resources represented by chronic diseases had a negative effect on search volume for flu symptoms. For the risk of the flu, a population with higher incidences of chronic diseases demonstrated less searches for flu symptoms. This finding may reflect an environmental constraint on Internet access rather than a lack of interest in such information. As noted in a report

from the Pew Research Center [62], adults living with chronic disease are significantly less likely than healthy adults to have access to the Internet. Individuals with chronic diseases are more likely to have regular contact with health professionals, also highlighting the important role health care providers play in patient education about health risks.

Finally, in our analyses, public resources represented by population density had a positive effect on search volume for flu shots. This finding is important in that, as a contagious illness, the search patterns for flu may also emerge for other communicable diseases.

Implications

Our study has important implications for public policy makers and health care professionals theoretically and practically. First, based on ecological systems theory, we proposed that there is a correlation between online health information search and public health risk perception. Recognition of this relationship by policy makers and health care professionals is important. In designing health risk communications strategies and policies, it is critical to take the social environments in which the public engages in online health information search into consideration.

Second, we suggest that the analysis of Internet search query data related to a particular health risk can provide a bell-weather of public health risk perception. Our analysis suggests that online health information search is a reflection of public health risk perception that can be predicted by social context variables. We demonstrate a practical way for policy makers and health professionals to monitor these contextual factors. Previous work has shown that aggregate search data reflect public concerns or interests (eg, [63]). Following this stream of research, we demonstrate that when a population is concerned about a specific health risk, they engage in online searches about that risk. This online search is predicted by contextual factors. Monitoring these contextual variables on a regular basis can assist policy makers in identifying areas and/or populations that could benefit from enhanced education. It may also help identify areas on which to focus the development/expansion of resources.

The search volume data for queries representing different stages of risk response, such as prevention and diagnosis, can inform policy makers and health professionals about the likely response of a population to current and emerging threats. Social marketing resources should be allocated based on an understanding of public risk perception of prevention and diagnosis of a health risk. For example, seasonality had more influence on search volume related to prevention than did other variables. Social marketing efforts should be timed to coincide with the seasonal variation in public risk perception for flu prevention. Those states with high levels of public perception for risk prevention can use this finding to help prepare for a new flu season by arranging for extra supplies of flu vaccine and planning effective systems for distributing the vaccine. States with low levels of public risk perception for prevention might benefit from more health education and health promotion prior to the onset of a new flu season to help increase awareness of the impending risk.

We found that media attention and private and public resources had strong effects on public risk perception for symptoms. Populations with higher risk perception for the diagnosis of the flu are likely to have higher demand for products or services related to treatment of the flu (eg, vitamin C supplements, primary care visits). Retailers in states with high levels of public risk perception for response to the flu may want to ensure that they have adequate supply of over-the-counter medications for dealing with flu symptoms. Ambulatory care clinics and primary care providers can assist a population in dealing with the flu by providing educational materials focused on identification of symptoms and by ensuring same-day access to provider visits for patients experiencing flu symptoms. To respond to public risk perception for diagnosis of the flu, social marketing efforts should use sociocultural segmentation (eg, vulnerable and health conscious consumers) to target resources most needed by each segment.

With the popularity of mobile devices (eg, smartphones, iPads), mobile searches are growing among consumers. Surveys have shown that a search engine is the most used application by 77% of smartphone users, and 90% of mobile search activities result in actions (eg, purchasing, recommending) [64]. This suggests that search data from mobile devices may reflect the public's perception of the urgency of the risks and their ability to manage the risks. We suggest that policy makers and health care professionals use mobile search data related to health risks to establish more actionable and timely strategies.

Online information searching is a bidirectional communication process, including sending search requests and receiving search results. Sending search requests reflects the public's perception of the severity and urgency of risks, whereas receiving search results reflects the public's perception of their ability to manage or respond to the risks. This study focused on public risk perception as demonstrated by the patterns of search requests. Policy makers and health professionals may further explore public risk perception by examining the patterns of responses to the returns to search requests. We suggest that the public's perception of the management of health risks may be revealed through behaviors that reflect 4 types of social relations (ie, hierarchical, egalitarian, individualist, and fatalist social relations) [29]. People with a hierarchical approach to social relations (ie, supporting patriotism, law, and order) may be more likely to click on search results links from official government websites, whereas people with an individualist way of life (ie, supporting individual efforts) may be more likely to click on search results from citizen media (eg, independent journalists). We suggest that investigating the association between the public's response to search results and social-cultural factors may be a practical way to assess the public's perception of the management of health risks. Policy makers and health care professionals may combine the patterns of search requests and

response to search results to generate a composite index for public health risk perception.

Limitations

Our study has several limitations. First, data gaps exist for the variables we used to indicate flu incidence and online information searches related to the flu. By using a different unit of analysis, additional relevant data may be available for study. For example, for each flu season from 1997 to the current year, the CDC posts the data of ILI outpatient visit rates for 9 flu surveillance regions on its official websites [65]. Regional data are available for Pacific, Mountain, West South Central, East South Central, West North Central, East North Central, New England, mid-Atlantic, and South Atlantic. Using these data regional variations regarding public health risk perception could be explored. Similarly, with data that are available at the state level, state data could be combined, facilitating regional analysis. Such regional analyses may be particularly relevant to public health risks that occur most commonly in particular geographic areas, such as those related to hurricanes.

Second, we base our findings on aggregate data. One limitation of aggregate data is that they represent the characteristics of a group as a whole but do not allow for analysis of individual variation. We cannot establish how individuals perceive their social environments related to health risks. Future research is needed to investigate individual responses to social factors related to health risks by collecting data from self-report surveys.

Next, our study has only shown the usefulness of sets of variables for the prediction of public risk perception related to the flu. Different types of health risks vary in their characteristics such as immediacy, frequency, and severity. These factors may lead to variations not only in the effects of disease incidence, but also the relationship of sociocultural factors to public risk perception as demonstrated by online information search. More research is needed to identify common and unique variables for the measurement of public risk perception related to different types of health risks. For example, food-borne illnesses and the flu are both common health risks. Vaccination is available for the flu but not for food-borne illnesses. Future research should consider the availability of preventive and treatment options for different health risks as they may affect public perception for the health risks.

Finally, our study has shown the strong effects of traditional mainstream mass media (ie, newspaper and news wires) on public risk perception. Research is needed to investigate the influence of multiple forms of mass media, especially social media (eg, blogs, online social networks), on public risk perception. May's [66] report about information channels and networks during Hurricane Katrina has identified the prominence of digital communication for risk management.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Google Insights for Search (GIFS) Methodology.

[[PDF File \(Adobe PDF File\), 21KB - jmir_v15i6e114_app1.pdf](#)]

References

1. Fox S, Jones S. The social life of health information. Washington, DC: Pew Internet & American Life Project; 2009 Jun. URL: http://pewinternet.org/~media/Files/Reports/2009/PIP_Health_2009.pdf [accessed 2012-10-11] [[WebCite Cache ID 6BLDY5Hs1](#)]
2. 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](#)]
3. Schwartz KL, Roe T, Northrup J, Meza J, Seifeldin R, Neale AV. Family medicine patients' use of the Internet for health information: a MetroNet study. *J Am Board Fam Med* 2006 Feb;19(1):39-45 [[FREE Full text](#)] [Medline: [16492004](#)]
4. Fallows D. Search engine use. Washington, DC: Pew Internet & American Life Project; 2008. URL: http://www.pewinternet.org/~media/Files/Reports/2008/PIP_Search_Aug08.pdf [accessed 2012-10-12] [[WebCite Cache ID 6BMXHFTIc](#)]
5. Cooper CP, Mallon KP, Leadbetter S, Pollack LA, Peipins LA. Cancer Internet search activity on a major search engine, United States 2001-2003. *J Med Internet Res* 2005 Jul 1;7(3):e36 [[FREE Full text](#)] [doi: [10.2196/jmir.7.3.e36](#)] [Medline: [15998627](#)]
6. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. *AMIA Annu Symp Proc* 2006:244-248 [[FREE Full text](#)] [Medline: [17238340](#)]
7. Brownstein JS, Freifeld CC, Madoff LC. Influenza A (H1N1) virus, 2009--online monitoring. *N Engl J Med* 2009 May 21;360(21):2156 [[FREE Full text](#)] [doi: [10.1056/NEJMp0904012](#)] [Medline: [19423868](#)]
8. Pelat C, Turbelin C, Bar-Hen A, Flahault A, Valleron A. More diseases tracked by using Google Trends. *Emerg Infect Dis* 2009 Aug;15(8):1327-1328 [[FREE Full text](#)] [doi: [10.3201/eid1508.090299](#)] [Medline: [19751610](#)]
9. 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](#)]
10. Hulth A, Rydevik G, Linde A. Web queries as a source for syndromic surveillance. *PLoS One* 2009;4(2):e4378 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0004378](#)] [Medline: [19197389](#)]
11. Wong PW, Fu KW, Yau RS, Ma HH, Law YW, Chang SS, et al. Accessing suicide-related information on the internet: a retrospective observational study of search behavior. *J Med Internet Res* 2013;15(1):e3 [[FREE Full text](#)] [doi: [10.2196/jmir.2181](#)] [Medline: [23305632](#)]
12. Zheluk A, Gillespie JA, Quinn C. Searching for truth: internet search patterns as a method of investigating online responses to a Russian illicit drug policy debate. *J Med Internet Res* 2012;14(6):e165 [[FREE Full text](#)] [doi: [10.2196/jmir.2270](#)] [Medline: [23238600](#)]
13. Burton SH, Tanner KW, Giraud-Carrier CG, West JH, Barnes MD. "Right time, right place" health communication on Twitter: value and accuracy of location information. *J Med Internet Res* 2012;14(6):e156 [[FREE Full text](#)] [doi: [10.2196/jmir.2121](#)] [Medline: [23154246](#)]
14. 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](#)]
15. 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](#)]
16. Hill S, Mao J, Ungar L, Hennessy S, Leonard CE, Holmes J. Natural supplements for H1N1 influenza: retrospective observational infodemiology study of information and search activity on the Internet. *J Med Internet Res* 2011;13(2):e36 [[FREE Full text](#)] [doi: [10.2196/jmir.1722](#)] [Medline: [21558062](#)]
17. 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](#)]
18. Slovic P. Perception of risk. *Science* 1987 Apr 17;236(4799):280-285. [Medline: [3563507](#)]
19. Arrow KJ. Aspects of the Theory of Risk Bearing. Chicago, IL: Markham Publication; 1965.
20. Duxbury D, Summers B. Financial risk perception: are individuals variance averse or loss averse? *Economics Letters* 2004;84(1):21-28.
21. Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. *Econometrica* 1979;47(2):263-292.
22. Viscusi WK. The risks and rewards of criminal activity: a comprehensive test of criminal deterrence. *Journal of Labor Economics* 1986;4(3):317-340.
23. Brewer NT, Chapman GB, Gibbons FX, Gerrard M, McCaul KD, Weinstein ND. Meta-analysis of the relationship between risk perception and health behavior: the example of vaccination. *Health Psychol* 2007 Mar;26(2):136-145. [doi: [10.1037/0278-6133.26.2.136](#)] [Medline: [17385964](#)]

24. Finucane M, Slovic P, Mertz CK, Flynn J, Satterfield T. Gender, race, and perceived risk: the "white male" effect. *Health Risk Society* 2000;2(2):159-172.
25. Bronfenbrenner U. Toward an experimental ecology of human development. *American Psychologist* 1977;32:513.
26. Klein KJ, Tosi H, Cannella AA. Multilevel theory building: benefits, barriers, and new developments. *Academy of Management Review* 1999;24:243.
27. McLeroy KR, Bibeau D, Steckler A, Glanz K. An ecological perspective on health promotion programs. *Health Educ Q* 1988;15(4):351-377. [Medline: [3068205](#)]
28. Edberg M. *Essential Readings in Health Behavior: Theory and Practice*. Burlington, MA: Jones & Bartlett Publishers; 2010.
29. Douglas M, Wildavsky A. *Risk and Culture: An Essay on the Selection of Technical and Environmental Dangers*. Berkeley, CA: University of California Press; 1982:1982.
30. Short JF. The social fabric at risk: Toward the social transformation of risk analysis. *American Sociological Review* 1984;49(6):711-725.
31. Kasperson RE, Renn O, Slovic P, Brown HS, Emel J, Goble R, et al. The social amplification of risk: a conceptual framework. *Risk Analysis* 1988;8(2):177-187 [FREE Full text]
32. McCombs ME, Shaw DL. The agenda-setting function of mass media. *Public Opinion Quarterly* 1972;36:176-187. [doi: [10.1086/267990](#)]
33. Coleman C. The Influence of Mass Media and Interpersonal Communication on Societal and Personal Risk Judgments. *Communication Research* 1993 Aug 1993;20(4):611-628. [doi: [10.1177/009365093020004006](#)]
34. Sjöberg L. Factors in risk perception. *Risk Anal* 2000 Feb;20(1):1-11. [Medline: [10795334](#)]
35. Wahlberg AAF, Sjöberg L. Risk perception and the media. *Journal of Risk Research* 2000;3(1):31-50.
36. 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](#)]
37. 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 2011;33(2):241-259 [FREE Full text] [doi: [10.1093/aepp/ppr008](#)]
38. Holland K, Blood RW, Pirkis J, Dare A. Postpsychiatry in the Australian media: the "vulnerable" talk back. *Asia Pacific Media Educator* 2009;19:142-157 [FREE Full text]
39. Pancioli AM, Broderick J, Kothari R, Brott T, Tuchfarber A, Miller R, et al. Public perception of stroke warning signs and knowledge of potential risk factors. *JAMA* 1998;279(16):1288-1292. [Medline: [9565010](#)]
40. Maswanya ES, Moji K, Horiguchi I, Nagata K, Aoyagi K, Honda S, et al. Knowledge, risk perception of AIDS and reported sexual behaviour among students in secondary schools and colleges in Tanzania. *Health Educ Res* 1999 Apr;14(2):185-196 [FREE Full text] [Medline: [10387499](#)]
41. Sjöberg L. Political decisions and public risk perception. *Reliability Engineering & System Safety* 2001 May 2001;72(2):115-123. [doi: [10.1016/S0951-8320\(01\)00012-6](#)]
42. Wildavsky A, Dake K. Theories of risk perception: Who fears what and why? *Daedalus* 1990;119(4):41-60.
43. Grasmück D, Scholz RW. Risk perception of heavy metal soil contamination by high-exposed and low-exposed inhabitants: the role of knowledge and emotional concerns. *Risk Anal* 2005 Jun;25(3):611-622. [doi: [10.1111/j.1539-6924.2005.00628.x](#)] [Medline: [16022694](#)]
44. Rountree PW, Land KC. Perceived risk versus fear of crime: Empirical evidence of conceptually distinct reactions in survey data. *Social Forces* 1996 Jun;74(4):1353-1376.
45. Redelmeier DA, Rozin P, Kahneman D. Understanding patients' decisions. Cognitive and emotional perspectives. *JAMA* 1993 Jul 7;270(1):72-76. [Medline: [8510300](#)]
46. Haomiao J, Santana A, Lubetkin EI. Measuring risk perception among low-income minority primary care patients. *J Ambul Care Manage* 2004 Dec;27(4):314-327. [Medline: [15495744](#)]
47. Schmidt T, Vosen S. Social Science Research Network. 2009. Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1514369 [accessed 2012-10-12] [WebCite Cache ID 6BMX39E0J]
48. Breyer BN, Eisenberg ML. Use of Google in study of noninfectious medical conditions. *Epidemiology* 2010 Jul;21(4):584-585 [FREE Full text] [doi: [10.1097/EDE.0b013e3181e080cd](#)] [Medline: [20539114](#)]
49. Baram-Tsabari A, Segev E. Exploring new web-based tools to identify public interest in science. *Public Understanding of Science* 2009 Oct 2009;20(1):130-143. [doi: [10.1177/0963662509346496](#)]
50. Jansen BJ, Spink A, Bateman J, Saracevic T. Real life information retrieval: a study of user queries on the Web. *SIGIR Forum* 1998 Apr 1998;32(1):5-17. [doi: [10.1145/281250.281253](#)]
51. US Outpatient Influenza-like Illness Surveillance Network (ILINet). URL: <http://www2a.cdc.gov/ilinet/> [accessed 2013-04-10] [WebCite Cache ID 6FmC8uJc]
52. Doornik JA. Improving the timeliness of data on influenza-like illnesses using Google search data. 2009 Oct 8. URL: [http://www.doornik.com/flu/Doornik\(2009\)_Flu.pdf](http://www.doornik.com/flu/Doornik(2009)_Flu.pdf) [accessed 2012-10-12] [WebCite Cache ID 6BMZIDysv]

53. LexisNexis Academic. URL: <http://www.lexisnexis.com/hottopics/lnacademic/> [accessed 2013-04-10] [WebCite Cache ID 6FmCRXDd4]
54. US Census Bureau. Data access tools URL: <http://www.census.gov/main/www/access.html> [accessed 2013-04-10] [WebCite Cache ID 6FmCnRPW6]
55. Centers for Disease Control and Prevention. Seasonal influenza (flu): information for specific groups URL: <http://www.cdc.gov/flu/groups.htm> [accessed 2013-04-10] [WebCite Cache ID 6FmD95B7a]
56. US Census Bureau. Fast facts for Congress URL: <http://www.census.gov/fastfacts/> [accessed 2013-04-10] [WebCite Cache ID 6FmDqxldH]
57. Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System URL: <http://apps.nccd.cdc.gov/gisbrfss/default.aspx> [accessed 2013-04-10] [WebCite Cache ID 6FmE2Ovuj]
58. The Henry J. Kaiser Foundation. State health facts URL: <http://www.statehealthfacts.org/> [accessed 2013-04-10] [WebCite Cache ID 6FmEBAaUy]
59. Trust for America's Health. State data URL: <http://healthyamericans.org/states/> [accessed 2013-04-10] [WebCite Cache ID 6FmMJprOw]
60. Centers for Disease Control and Prevention. Seasonal influenza (flu): the flu season URL: <http://www.cdc.gov/flu/about/season/flu-season.htm> [accessed 2013-04-10] [WebCite Cache ID 6FmNJNjiP]
61. US Census Bureau. Computer and Internet use URL: <http://www.census.gov/hhes/computer/publications/2009.html> [accessed 2013-04-10] [WebCite Cache ID 6FmNiWx2h]
62. Fox S, Purcell K. Chronic disease and the Internet. Washington, DC: Pew Internet & American Life Project; 2010 Mar 24. URL: http://pewinternet.org/~media/Files/Reports/2010/PIP_Chronic_Disease_with_topline.pdf [accessed 2012-10-12] [WebCite Cache ID 6BMZsb79V]
63. Askatas N, Zimmermann KF. Google econometrics and unemployment forecasting. *Applied Economics Quarterly* 2009;55:107 [FREE Full text]
64. Google Mobile Ads Blog. 2011 Apr 26. Smartphone user study shows mobile movement under way URL: <http://googlemobileads.blogspot.com/2011/04/smartphone-user-study-shows-mobile.html> [accessed 2013-02-18] [WebCite Cache ID 6EXAhORQj]
65. Centers for Disease Control and Prevention. Seasonal influenza (flu): past weekly surveillance reports URL: <http://www.cdc.gov/flu/weekly/pastreports.htm> [accessed 2013-02-27] [WebCite Cache ID 6EkdaCR1t]
66. May AL. First informers in the disaster zone: the lessons of Katrina. Washington, DC: The Aspen Institute; 2007. URL: <http://www.policyarchive.org/handle/10207/bitstreams/4525.pdf> [accessed 2013-02-19] [WebCite Cache ID 6EXhcjXHe]

Abbreviations

GIFS: Google Insights for Search

ILI: influenza-like illness

Edited by G Eysenbach; submitted 16.10.12; peer-reviewed by N Bragazzi, E Augustson; comments to author 09.02.13; accepted 25.02.13; published 17.06.13.

Please cite as:

Liang B, Scammon DL

Incidence of Online Health Information Search: A Useful Proxy for Public Health Risk Perception

J Med Internet Res 2013;15(6):e114

URL: <http://www.jmir.org/2013/6/e114/>

doi: [10.2196/jmir.2401](https://doi.org/10.2196/jmir.2401)

PMID: [23773974](https://pubmed.ncbi.nlm.nih.gov/23773974/)

©Bo Liang, Debra L Scammon. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 17.06.2013. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Postmarket Drug Surveillance Without Trial Costs: Discovery of Adverse Drug Reactions Through Large-Scale Analysis of Web Search Queries

Elad Yom-Tov¹, PhD; Evgeniy Gabrilovich¹, PhD

Yahoo Research, New York, NY, United States

Corresponding Author:

Elad Yom-Tov, PhD

Yahoo Research

111 W 40th st.

New York, NY, 10018

United States

Phone: 1 6462136000

Fax: 1 6462136000

Email: eladyt@yahoo.com

Abstract

Background: Postmarket drug safety surveillance largely depends on spontaneous reports by patients and health care providers; hence, less common adverse drug reactions—especially those caused by long-term exposure, multidrug treatments, or those specific to special populations—often elude discovery.

Objective: Here we propose a low cost, fully automated method for continuous monitoring of adverse drug reactions in single drugs and in combinations thereof, and demonstrate the discovery of heretofore-unknown ones.

Methods: We used aggregated search data of large populations of Internet users to extract information related to drugs and adverse reactions to them, and correlated these data over time. We further extended our method to identify adverse reactions to combinations of drugs.

Results: We validated our method by showing high correlations of our findings with known adverse drug reactions (ADRs). However, although acute early-onset drug reactions are more likely to be reported to regulatory agencies, we show that less acute later-onset ones are better captured in Web search queries.

Conclusions: Our method is advantageous in identifying previously unknown adverse drug reactions. These ADRs should be considered as candidates for further scrutiny by medical regulatory authorities, for example, through phase 4 trials.

(*J Med Internet Res* 2013;15(6):e124) doi:[10.2196/jmir.2614](https://doi.org/10.2196/jmir.2614)

KEYWORDS

machine learning; information retrieval; side effects; infoveillance; infodemiology

Introduction

Existing mechanisms for postmarket drug surveillance work well in many cases, but failures resulting in harm to patients and even fatalities are widely documented [1], including the withdrawal of thalidomide in the 1960s [2], and more recently of cerivastatin [3], troglitazone [4], and rofecoxib [5]. Two main kinds of postmarket drug surveillance mechanisms exist today. One kind is run by regulatory agencies, such as MedWatch by the US Food and Drug Administration (FDA) and the Vaccine Adverse Event Reporting System (VAERS) by the FDA and the Centers for Disease Control and Prevention (CDC) in the United States, the Yellow Card Scheme by the Medicines and

Healthcare products Regulatory Agency (MHRA) in the United Kingdom, and the International Drug Monitoring Programme by the World Health Organization (WHO). These are supplemented by public (or public-private cooperation) initiatives, such as Research on Adverse Drug Events and Reports (RADAR), and Web sites, such as eHealthMe.com, which collect patient-reported information on drug outcomes. The most serious limitation of these data collection initiatives is that they rely on the patients and their health care providers to make the association between the adverse drug reaction (ADR) and the drug. This can be especially difficult when the adverse reaction appears only after the drug is taken for a lengthy period of time, or when the patient takes several

medications concurrently. To alleviate this problem, projects such as the FDA's Sentinel Initiative [6], the EU-ADR initiative [7], and the Observational Medical Outcomes Partnership (OMOP) [8-11] are beginning to use observational data, including administrative claims and electronic health records, to identify adverse drug reactions.

Our proposed approach uses a novel kind of observational data, namely, Web search query logs. Search queries contain a cornucopia of world knowledge, and prior studies have used query logs to track tropical storms [12], certain life events [13], and the spread of disease [14]. As such, this approach is an example of infodemiology [15], and is enabled by the fact that as many as 80% of US Internet users seek health information online [16]. Consequently, our methodology allows analyzing the data from literally hundreds of millions of people, and in some cases, a significant percentage of the patients using a given drug. Performing such analysis continually allows for long-term monitoring, whereas grouping search requests by geographical location facilitates demographic segmentation of the population [17].

Existing drug surveillance mechanisms often depend on the need for medical providers or patients to realize the connection between the treatment and its side effects (adverse or otherwise). This inherent limitation poses a challenge to testing new methods for ADR discovery because existing data are not comprehensive enough to be considered a gold standard, considering that patients and medical providers might not realize the connection between treatments and some ADRs. Therefore, we adopted a 2-pronged approach for validating our method. First, we showed that it can reliably identify currently known ADRs. Although the findings of our method are positively correlated with existing data, this correlation is not perfect as we discover new, previously unknown ADRs. Second, we characterized the differences between the known and the newly discovered ADRs, and identified the most discordant ADRs (MDADRs) between these 2 sources. We show that the ADRs found by our method are usually less acute reactions (ie, not requiring immediate medical attention) with much later onset, which is exactly why they elude detection by conventional mechanisms. For all the drugs examined, we found that the ADRs apnea and cramps are consistently overlooked in the FDA data, as reported in the Adverse Event Reporting System (AERS), described subsequently, whereas tiredness and weight loss are frequent ADRs of vaccines that are overlooked in VAERS reports. We propose that the ADRs newly discovered by our method be further investigated in carefully designed clinical trials, which should be lengthy enough to allow detection of late-onset reactions.

Methods

Our method, called the query log reaction score (QLRS), quantifies the prevalence of ADRs for a given drug, as explained subsequently. We used QLRS to identify ADRs of top-selling drugs and vaccines on the basis of queries submitted to the Yahoo US Web search engine during 6 months in 2010. A total

of 176 million unique users, as identified by a unique signature of the users' browser, were included in this study. The search logs were anonymized according to the Yahoo privacy policy by scrambling actual user identifiers. This was achieved by using a 1-way cryptographic hash function, which makes it impossible to map the resultant hash values back to the original user identifiers, while keeping the probability of collisions very low. As explained subsequently, only the search counts were considered, which were aggregated across users. Furthermore, the research described herein was carried out according to the Yahoo guidelines on human subject research.

We investigated 20 drugs (additional results for the top 100 drugs are provided in [Multimedia Appendix 1](#)), which are the top-selling drugs in the United States by revenue [18]. We analyzed these drugs for 2 reasons. First, these findings would likely affect the largest number of people. Second, data are more abundant for these drugs; thus, results are likely to be more significant for these drugs. We note that all these drugs are usually taken for long periods of time; however, we have also demonstrated the applicability of our method to vaccines, as detailed in [Multimedia Appendix 1](#), which are usually administered a limited number of times to each patient. We limited our work to nongeneric versions of these drugs to reduce the chance of additional confounding influences, and because brand names are mentioned 88% more often than generic names in the query log (not statistically significant). However, we also discuss the differences in ADRs of similar drugs in the Results section.

A total of 195 symptoms from *The International Statistical Classification of Diseases and Related Health Problems* 10th Revision (ICD-10) [19] were studied as manifestations of possible ADRs. We filtered the symptoms according to Wikipedia's List of Medical Symptoms [20] to facilitate replicability of our method in other languages. This list of symptoms was further expanded with synonyms (described subsequently), because patients frequently use nonmedical terminology to describe their symptoms. Basing our work on terms from Wikipedia (a popular information source) and identifying synonyms using behavioral data makes our approach suitable for identifying ADRs as described by nonprofessionals.

We limited the symptoms under consideration for each drug to the 50 most frequently queried symptoms for that drug. We identified possible ways nonprofessionals described their health symptoms by using 2 query expansion methods. First, we selected the most frequent search terms that led users to click on the Wikipedia page that described each symptom [21-23]. Second, we extracted frequently occurring lexical affinities [24], namely, word pairs appearing in close proximity in the 50 highest ranked Web search results returned when the symptom name was used as a query. The 2 top terms from each of the 2 methods were used as alternative names for each symptom. For each symptom discussed in the paper, the various possible search terms expressing it have been mapped to the same medical term. For example, the ADR *diplopia* could have been searched for by using the colloquial term *double vision*.

Table 1. The chi-square contingency table for a given drug and adverse drug reactions (ADRs) used for computing the query log reaction score (QLRS).

When user queried for the ADR	User queried for the drug?	
	No	Yes
Before Day 0	N11	N12
After Day 0	N21	N22

Three medical professionals (2 medical doctors, 1 nurse practitioner) independently labeled the expansion terms with respect to their relevance as an expansion term to each specific medical term. The interannotator agreement estimated using the Fleiss' kappa statistic [25] was 0.44 ($P < .001$). This is a medium-level agreement. However, for 88% of the terms, most annotators (ie, ≥ 2) agreed that the term was an appropriate expansion of the medical term. Thus, our expansion method constructs a high-precision dictionary of terms. To maintain the automated nature of our method, the results reported here are based on all the expansion terms, not just those marked as relevant by the annotators.

For each drug, we first identified all the users who had searched for the drug name. For those, we define Day Zero for each user as the day when that user first searched for the drug. Day Zero for all other users (who did not search for the drug) was defined as the midpoint of their observed query history. We then counted the number of times each symptom was queried before and after Day Zero by each user. The purpose of using the data from people who did not search for the drug was to normalize against environmental effects, eg, seasonal allergies. This is in contrast with most prior infoveillance research [15], which is concerned with whole-population prevalence rather than the comparison of specific subpopulations, eg, people using or not using the drug.

For each drug-symptom pair, we constructed a 2-way contingency table counting the number of times a symptom was searched for before and after Day Zero, for users who did and did not search for the drug (see Table 1). For each symptom, we scored its prevalence as a reaction to the drug using the Pearson's goodness of fit test, the chi-square test statistic [26]. We refer to this score as the query log reaction score (QLRS). Additional results for the top 100 drugs are provided in Multimedia Appendix 1.

We used 2 reference datasets to assess the validity of our findings. Adverse Event Reporting System (AERS, currently known as the FDA Adverse Event Reporting System) is the database of the FDA's postmarket safety surveillance program for approved pharmaceutical drugs. The Side Effect Resource (SIDER) lists known ADRs for marketed drugs, extracted from public documents and package inserts [27].

The AERS data were downloaded from the FDA AERS website [28], and included reports submitted between January 2004 and June 2010. Reports were mapped to the same list of symptoms as QLRS using the same synonym list. In total, 47% of the cases in AERS were matched to at least 1 of the 195 symptoms or their synonyms, indicating good coverage by the symptoms list used in our study. Similar analysis was performed for SIDER. To assess the overall quality of ADR discovery by our method, we computed the Spearman rank correlation (ρ) between the 2

lists of ADRs for each drug, 1 ordered by QLRS and 1 by the number of AERS reports.

The AERS data are complicated by the fact that multiple reports can be submitted to the FDA for the same case, and that reports can pertain to side effects of the drug, the underlying disease, or another drug taken concurrently [29]. Therefore, ADR prevalence according to AERS should be considered a noisy reference. We employed several approaches to computing the correlation. First, we used the raw report counts in AERS, and denote the correlation between these and QLRS by ρ_1 . We also used the AERS data to compute 2 regularized measures of disproportionality that are commonly employed for analyzing adverse side effect reports. Specifically, we used the empirical Bayes geometric mean (EBGM) [30,31], and denote the correlation between EBGM computed for AERS and the QLRS by ρ_2 . Finally, the correlation between the information component (IC) [32] computed for AERS and QLRS is denoted by ρ_3 .

We hypothesized that some ADRs are more likely to be reported to the FDA, whereas others tend to be self-addressed by patients through online research. Consequently, if our method was to discover previously unknown ADRs, the correlation can never be perfect. Therefore, we first analyzed the commonalities among the ADRs we discovered and those already known. Then, we analyzed the properties of the newly discovered ADRs.

To focus on the ADRs identified by both our method and the AERS data, we removed the 5 symptoms that most reduced the value of 1 of the metrics, ρ_1 , using a greedy selection process. We call the removed symptoms the most discordant ADRs (MDADRs). Specifically, we iteratively identified and removed the ADR that most reduced the Spearman rank correlation between the AERS counts and the QLRS ranking of ADRs. An alternative method of removing discordant ADRs would focus on reaching statistically significant values of ρ_1 . However, we chose to use a fixed number to facilitate the analysis of MDADRs, as performed subsequently.

Identification of Adverse Drug Reactions for Multiple Drugs

Some individuals are prescribed multiple drugs to be taken simultaneously. The interaction between these drugs may give rise to specific ADRs that are not present (or are present at different severity) if each drug is taken individually. Thus, in the following we show how our method can be used to identify ADRs that are associated with taking pairs of drugs. Our method attempts to remove the ADRs attributed to individual drugs so as to identify those ADRs that arise from the combination thereof.

For each pair of drugs that were analyzed, we identified their characteristic ADRs caused by the interaction by discounting the probability of the ADRs arising from each of the individual drugs. This was done by subtracting the contribution of ADRs of the individual drugs as predicted by a linear regression model. We hypothesized that the ADRs observed in patients who only take 1 of the drugs, will appear at a similar ratio for the patients who take both drugs. However, new ADRs that are caused by the interaction of the 2 drugs will not be reliably predicted by modeling each drug separately and will, therefore, appear at a substantially different ratio than the prediction.

For each pair of drugs, we identified 3 disjoint groups of users: the first 2 groups are those who searched for only 1 of the 2 drugs, and the third group searched for both. For the first 2 groups, we counted the number of times each ADR was searched for before and after Day Zero. For the third group (ie, users who searched for both drugs), we defined Day Zero as $\max(\text{date-first-search}(\text{drug}_i), \text{date-first-search}(\text{drug}_j))$, where $\text{date-first-search}()$ is the earliest date on which the user searched for a given drug. We denote these numbers (before/after Day Zero) for the i -th ADR in population p by $n_{i,b}^p$ ($n_{i,a}^p$). Next, we defined the ratio of change in the ADR prevalence (after the commencement of treatment with the second drug) as $n_{i,a}^p / (n_{i,b}^p + n_{i,a}^p)$. Finally, we built a regression model to predict the probability of change in the third population (patients taking both drugs) given the corresponding values in the first 2 populations. This regression model effectively discounts the effect of the ADRs caused by each drug separately. We also identified MDADRs for pairs of drugs in a similar way as for individual drugs.

Results

We counted the number of times each drug appeared in AERS, and found it to be highly correlated with the number of online searches for that drug. For the drugs listed in Table 2, $\rho = 0.66$ ($P=.002$; $n=20$). The correlation becomes even more pronounced for pairs of drugs, $\rho = 0.73$ ($P<.001$; $n=380$). The correlation between the sales figures (as represented by the number of prescriptions sold) and the number of Web searches is $R^2=0.26$ ($P=.005$). A linear model that uses both the number of AERS reports and the sales figures to predict Web search volume yields a R^2 2.4% greater than the one using only AERS reports. We believe these findings mean that the search volume is more indicative of the prevalence of ADRs rather than actual sales. Thus, the popularity of a drug in Web queries is highly representative of its appearance in AERS, suggesting that Web queries are strongly reflective of real-world phenomena.

As noted in the Methods section, we assessed the overall quality of ADR discovery by our method by computing the Spearman rank correlation coefficient between 2 lists of ADRs for each drug, 1 ordered by QLRs and 1 by the number of AERS reports.

Table 2 reports the values of ρ_1 after removing 5 MDADRs for each drug. QLRs predictions are relatively highly correlated with the AERS counts, and the correlation is statistically significant ($P<.05$) for 12 of the 20 drugs using the Olkin-Pratt (DSL) fixed-effect meta-analytical approach [33] ($P<.001$; $n=20$). Positive correlation was not found in only 1 of the drugs (Singulair). Interestingly, removing 15 MDADRs for this drug (instead of 5) resulted in a statistically significant correlation of $\rho_1 = 0.48$ ($P=.02$), suggesting a particularly high discrepancy between the prevalence of ADRs as predicted by QLRs and as registered in AERS for this drug.

We also note that although most of the observed correlation values are significant, they are far from indicating perfect correlation. This is to be expected because the correlation would only have been perfect if our method were exactly rediscovering the known ADRs. However, as we discovered previously unknown ADRs, we obviously achieved an imperfect match to the list of known ones in AERS. In the following section, we analyze the differences between the known ADRs and those identified by our method. Additionally, there is a small negative correlation ($\rho=-0.22$, $P=.02$) between the number of users who queried for a drug and ρ_1 . This demonstrates that higher correlations are obtained when more data are available, and is an additional cause for the imperfect correlations.

Statistically significant correlations with EBGM and IC were also found (see Multimedia Appendix 1), and the meta-analysis is statistically significant ($P<.001$; $n=100$). However, EBGM and IC are measures designed to enhance the detection of ADRs that are especially prevalent in a given drug under study compared with all other drugs. At the same time, raw AERS counts (used for the computation of ρ_1) are more likely to be associated with the appearance of an ADR regardless of any other drug. This explains the higher correlation we observed of QLRs with the raw AERS counts (ρ_1) than with EBGM and IC (ρ_2 and ρ_3 , respectively).

SIDER [27] contains information on ADRs extracted from public documents and package inserts. Because of regulatory and legal requirements, it is overly inclusive in its listings, which makes it a noisy reference as well. The SIDER data are essentially binary, without relative frequency or absolute counts, which makes the previous correlation analysis inapplicable. We used SIDER to assess the accuracy of QLRs by computing the area under the curve (AUC) [34] of a receiver operating characteristic (ROC) curve and the F score [35], taking as positive examples all the ADRs listed in SIDER for the drug. AUC measures the method's ability to correctly identify known ADRs, whereas the F score simultaneously considers precision and recall. Only 8 of the 20 drugs we analyzed appeared in SIDER, and the corresponding accuracies are reported in Table 3 (after removing MDADRs). The results suggest that our method is able to reconstruct known ADRs, as measured with AUC and the F score.

Table 2. Spearman rank correlation (ρ) between query log reaction score (QLRS) and the number of adverse drug reaction (ADR) reports in the Adverse Event Reporting System (AERS), with the most discordant ADRs (MDADRs) removed.

Drug	ρ	<i>P</i> value ^a	MDADRs ^b
Advair	0.28		Anxiety, apnea, ^c chest pain, cough, weight gain ^c
Aranesp	0.30		Asthenia, back ache, back pain, edemac
Diovan	0.34		Chest pain, cramp, ^c sleepy, ^c wound
Effexor	0.54	<.001	Nausea, phobia, ^c sleepy, ^c weight gain ^c
Enbrel	0.39	.02	Back pain, cough, diarrhea, fever, weight gain ^c
Lipitor	0.54	<.001	Asthenia, constipation, diarrhea, dizziness, nausea
Mabthera	0.38	.01	Chest pain, fever, headache, malaise, wound ^c
Nexium	0.45	.008	Abdominal pain, tired, ^c weak, ^c weight gain ^c
Norvasc	0.34		Apnea, ^c constipation, cramp, ^c tired, ^c weight loss ^c
Pantoloc	0.49	.001	Chest pain, fever, headache, malaise, nausea
Pantozol	0.51	.006	Chest pain, fever, headache, malaise, nausea
Plavix	0.25		Back pain, chest pain, cough, paresthesia
Protonix	0.25		Abdominal pain, diarrhea, nausea, vomit ^c
Remicade	0.37	.04	Chest pain, fever, infertility, ^c paresthesia, rash
Risperdal	0.40	.02	Diarrhea, headache, insomnia, weight gain ^c
Rituxan	0.23		Abdominal pain, diarrhea, paresthesia, weak ^c
Seretide	0.41	0.004	Chest pain, dyspnea, ^c headache, malaise, nausea
Seroquel	0.48	0.004	Apnea, ^c dizziness, headache, weight gain ^c
Singular	-0.06		Apnea, ^c dizziness, insomnia, tired ^c
Zyprexa	0.61	0.002	Constipation, diarrhea, nausea, paresthesia, somnolence

^a*P* values are provided for statistically significant correlations ($n=45$).^bUnless otherwise indicated, MDADRs are those prominent in AERS.^cMDADRs emphasized in QLRS.**Table 3.** Accuracy of adverse drug reaction (ADR) identification by using QLRS, tested against the SIDER dataset with most discordant ADRs (MDADRs) removed.

Drug	<i>F</i> score (df)	AUC
Advair	0.77	0.67
Diovan	0.43	0.71
Effexor	0.94	0.67
Lipitor	0.76	0.7
Pantoloc	0.44	0.57
Pantozol	0.44	0.64
Plavix	0.55	0.59
Singular	0.52	0.64

Most Discordant Adverse Drug Reactions

Analyzing the MDADRs revealed characteristic differences among the known ADRs (registered in AERS and SIDER) and those identified by our method. The ADRs identified as most discordant are not random; instead, they belong to 1 of the

following 2 classes, as shown in Table 4. The first class includes ADRs that are readily recognized by patients and medical professionals because of their acuteness and fast onset. The other class includes later onset, less acute ADRs, which are more difficult to identify using self-reporting methods.

As noted previously, upon removing as few as 5 MDADRs, the correlation between QLRs and AERS counts (ρ_1) frequently becomes statistically significant. Conversely, removing a random subset of 5 symptoms only results in a negligible, statistically insignificant change in the correlation.

Although the MDADRs were identified separately for each drug, they were highly consistent across drugs. Of the 32 MDADRs we identified overall, 22 were chosen for more than 1 drug (mean 3.1, SD 2.3). Significantly, these ADRs were always overemphasized either in the query log or in AERS, but never in both (for different drugs). The likelihood of such behavior at random is smaller than 1:105. A typical example is the ADR nausea, which appeared at a far higher rank (ie, more prevalent) in the AERS dataset than in the QLRs ranking for 7 out of the 20 drugs, and was never found at a rank below that of QLRs for the other drugs.

Most importantly, MDADRs that are prominent in queries and in AERS have notable differences in their temporal behavior. As an illustrative example, we used the query log to compute the cumulative density functions (CDFs) over time for 2 MDADRs for the drug Effexor, 1 overemphasized in AERS (nausea) and 1 overemphasized by QLRs (sleepiness). Figure 1 shows the difference between the CDFs of the 2 MDADRs, starting from the time the drug is first searched for (day 0). As Figure 1 demonstrates, each of these 2 MDADRs is more likely

to occur in a different time range. Observe that the symptom prominent in AERS (nausea) is usually searched for shortly after the first query about the drug (ie, several days after Day Zero), when it is much more likely than the other symptom. In contrast, the symptom ranked highly by QLRs (sleepiness) appears much more prominently 45 to 75 days after the commencement of treatment, when the likelihood of nausea drops significantly.

We measured the difference in the time of onset (defined as the number of days between the first search for the drug and the first search for the ADR in the query log) for the MDADRs that were overemphasized by QLRs and in the AERS data. Averaged over all the drugs, the difference was 7.3 days (2-sided Wilcoxon signed rank test [36], $P=.01$; $n=15$). Based on these findings, we conclude that ADRs are more likely to be reported to the regulatory authorities if they appear shortly after commencing the treatment (as it might be easier for patients and caregivers to link the ADRs to the treatment), and that are serious enough to warrant reporting. Conversely, ADRs identified by our method usually appear much later after the beginning of treatment; hence, their possible association to the drug is often overlooked.

Thus, the MDADRs overemphasized by QLRs represent an interesting class of reactions that are harder to discover using traditional methods.

Figure 1. Temporal behavior of adverse drug reactions (ADRs). The difference between the cumulative probabilities of the ADRs “nausea” and “sleepiness” for the drug Effexor. The ADR highly ranked by QLRs (sleepiness) has a much later onset (45-75 days).

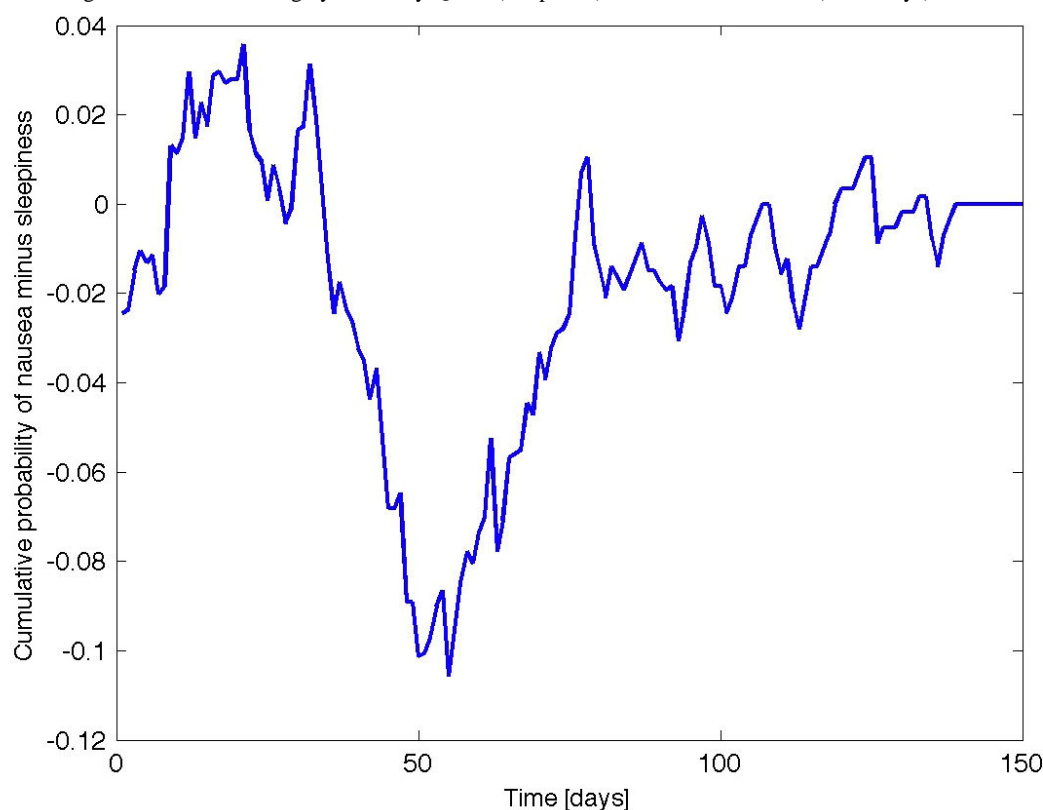


Table 4. Most discordant adverse drug reactions (MDADRs) identified (out of the 20 drugs analyzed).

MDADRs	n ^a
Overrepresented in AERS	
Abdominal pain	3
Asthenia	2
Back pain	3
Chest pain	8
Constipation	3
Cough	3
Diarrhea	8
Dizziness	3
Fever	5
Headache	6
Insomnia	2
Malaise	4
Nausea	7
Paresthesia	4
Overrepresented in the query logs	
Apnea	4
Cramps	2
Sleepy	2
Tired	3
Weak	2
Weight gain	6
Wound	2

^a Number of drugs in which MDADR appeared.

Correlation Between Query Log Reaction Scores of Similar Drugs

Several of the drugs we investigated are different brands of essentially the same drug. All other things being equal, we expected that patients taking 2 different brand versions of the same drug would experience similar ADRs. To evaluate this conjecture, we conducted 2 evaluations. First, we measured the Spearman correlation between the QLRs of ADRs for the multiple brand versions of the same drug. Second, we evaluated the correlation between the QLRs and AERS counts after aggregating the chi-square contingency tables for drugs that have the same generic names.

The intradrag correlation (different brands of the same drug) was, on average, 0.42 compared to 0.23 for all the other pairs of drugs ($P=.03$, 1-sided rank sum test). Thus, although the ADRs are somewhat different among the near-identical drugs, the correlation is statistically significantly higher than that observed for random pairings of drugs. The imperfect correlation can be explained by several factors. First, different manufacturers may produce drugs with slight variations in inactive ingredients, coloring agents, and fillers. The change in fillers in the drug Eltroxin in Israel has been associated with a

large number of patients experiencing major side effects, including changes in heart rate, dizziness, and difficulties in breathing [37]. Furthermore, there could be demographic differences between the populations taking those drugs. For example, different brand versions may be prescribed in different markets or different geographical regions. Finally, although many health care providers report that they do not employ special monitoring after switching from brand names to generic drugs, some have encountered specific ADRs caused by switching [38]. Thus, there are known differences in ADRs caused by different versions of similar drugs.

Nonetheless, the relatively high correlation between the ADRs of similar drugs provides additional supporting evidence that the ADRs discovered by QLRs are a genuine reflection of actual patient experiences.

By using RxNorm [39] we identified 30 drugs that are different brand versions of 14 generic drugs, out of the 100 drugs that we analyzed (see [Multimedia Appendix 1](#)). For example, Procrit and Eprex are 2 brand versions of the generic drug erythropoietin. This allowed us to focus the analysis on the generic component of the drug rather than the specific brand names, by computing an aggregated contingency table for all

brand name versions of the same generic drug before computing the QLRs (the AERS counts were aggregated similarly).

The resulting correlations are reported in Table 5. The values of ρ_1 aggregated over all the brand names of the same generic drug are significantly higher than those for individual brand

names (on average, aggregated $\rho_1 = 0.62$ compared to 0.35 for individual brand names). In all 14 cases, statistically significant correlations between the QLRs and AERS counts were found. MDADRs found in the aggregated data were the same as those identified in the brand name versions of the drug in 79% of the cases.

Table 5. Spearman rank correlation (ρ) between the adverse drug reports (ADRs) of generic drugs as identified by query log reaction score (QLRS) and by the number of reports in the Adverse Event Reporting System (AERS), with most discordant ADRs (MDADRs) removed according to the raw report counts in AERS. The QLRs and AERS counts for the generic drugs were computed by aggregating over multiple brand names of the same generic drug. For statistically significant correlations, P values are provided ($n=45$).

Brand names	Generic name	Averaged ρ_1 of individual brand names	ρ_1 aggregated over all the brand names of the same generic drug	P value
Procrit, Eprex	Erythropoietin	0.35	0.72	<.001
Neulasta, Neupogen	Filgrastim	0.43	0.54	.003
Lantus, Humalog	Insulin analog	0.43	0.60	<.001
Avonex, Rebif	Interferon beta-1a	0.60	0.83	<.001
AcipHex, Pariet	Rabeprazole	0.21	0.53	.002
Protonix, Pantozol, Pantoloc	Pantoprazole	0.42	0.67	<.001
TriCor, Lipanthyl	Fenofibrate	0.40	0.70	<.001
Rituxan, MabThera	Rituximab	0.31	0.52	.003
Advair, Flovent	Fluticasone	0.31	0.64	<.001
Cozaar, Hyzaar	Losartan	0.42	0.56	<.001
Losec, Prilosec	Omeprazole	0.27	0.65	<.001
Paxil, Seroxat	Paroxetine	0.13	0.44	.002
Avandamet, Avandaryl, Avandia	Rosiglitazone	0.38	0.67	<.001
Imigran, Imitrex	Sumatriptan	0.26	0.60	<.001

We attribute these findings to several reasons. First, by aggregating different brand name versions of a drug we focus on the active ingredient of the drug, which is more likely to be reported in AERS by medical personnel. Second, averaging these additional observations over multiple drugs reduces the amount of noise in the data, and thus increases the correlation with AERS. Finally, analyzing the active ingredient reduces the effect of individual manufacturing procedures and components, and thus focuses the analysis on a simpler set of chemical components, which may have a smaller set of ADRs.

Adverse Reactions to Multiple Drugs

Some ADRs occurred only when 2 drugs were taken concurrently or in close temporal proximity. These ADRs can

be especially difficult to detect because they occur infrequently, and only in a population that takes both drugs. To this end, we extended our method to identify ADRs of pairs of drugs.

The correlation between QLRs rankings and AERS (raw counts, EBGM, and IC), for the 10 most common pairs of drugs, is shown in Table 6. These correlations are lower than those for individual drugs, but are still statistically significant using the Olkin-Pratt (DSL) fixed-effect meta-analytical approach [33] ($P<.001$; $n=10$). We believe this result is noteworthy because it may be more difficult to include in a clinical trial those patients who take multiple drugs concurrently.

Table 6. Spearman rank correlation (ρ) between the adverse drug reactions (ADRs) of pairs of drugs identified by query log reaction score (QLRS) and by the number of reports in the Adverse Event Reporting System (AERS), with most discordant ADRs (MDADRs) removed according to the raw report counts in AERS.

Drug 1	Drug 2	ρ_1	ρ_2	ρ_3
Risperdal	Seroquel	0.27	-0.08	-0.16
Effexor	Advair	0.28	-0.11	0.04
Zyprexa	Seroquel	0.1	0.15	-0.05
Advair	Lipitor	-0.21	0.05	0.02
Plavix	Lipitor	0.3	0.29	0.24
Lipitor	Effexor	0.14	-0.18	-0.02
Advair	Plavix	0.14	0.15	0.05
Nexium	Plavix	0.19	0.11	0.33
Seroquel	Effexor	0.32	0.15	0.11
Lipitor	Nexium	0.12	0.07	0.14

Based on these findings, we believe our method can also be applicable to combination products (ie, drugs that contain 2 or more active substances), if each of the active substances is marketed also as a separate drug in similar doses. We plan to extend our method to combination products in our future work, and intend to investigate whether the correlation can be increased, for example, by using nonlinear correlation measures.

In their recent work, White et al [40] used search logs to study the side effects of 1 specific drug pair, paroxetine and pravastatin, whose interaction was reported to cause hyperglycemia. Their finding confirms the utility of search logs in identifying drug interactions, which were later validated by the FDA. However, there are several key differences between their study and ours. First, the method proposed by White et al performs a direct count of symptoms, thus not taking into account seasonal and other effects handled by the QLRS method. Second, to identify queries that are indicative of hyperglycemia, White et al constructed a list of hyperglycemia-related terms manually by reviewing the relevant medical literature. In contrast, we map user queries to medical terminology in an automated way, building on query expansion methods developed in the field of information retrieval. Finally, whereas White et al only analyzed 1 particular condition (hyperglycemia) and 1 specific pair of drugs (paroxetine and pravastatin), our study was conducted on a substantially larger scale. Specifically, we automatically mined side effects of 100 top-selling drugs and their combinations, and side effects of vaccines.

Discussion

Principal Findings

Clinical trials of pharmaceutical drugs are limited in their extent owing to their prohibitively high cost and insufficient diversity among participants. On the other hand, voluntary reporting of ADRs by patients and health care professionals is limited because of the extra effort required, and because of the difficulty of linking the ADRs to the drug that caused them (especially when these ADRs have a late onset or are caused by multidrug treatments). We propose a novel, low-cost method for discovering adverse drug reactions from aggregated Web search

data of large populations of Internet users. We demonstrated that our method allows analyzing the ADRs of drugs and vaccines in dramatically larger populations than typical clinical trials, and can assist in identifying ADRs that have so far eluded discovery by the existing mechanisms.

We believe our method constitutes a new, complementary approach to pharmacovigilance, because of its computational efficiency and access to vastly larger and more diverse populations. There are multiple avenues for future work. The effectiveness of our method can be validated by analyzing medical records (eg, OMOP), or by assessing its ability to predict changes in safety labels by regulatory authorities. It would also be interesting to compare the QLRS to those derived from the analysis of social media. Finally, a validation of MDADRs through clinical trials would be of significant value to validate our method. Specifically, we propose to test the MDADRs found by our method, which are underemphasized in current ADR databases (eg, AERS), in a clinical setting or through phase 4 trials. Such trials should be prioritized by the severity, volume of searches, and uniqueness of the ADRs discovered. Once verified, these MDADRs will become an important addition to the list of known ADRs of which patients are informed. Finally, quantifying the strength of the protopathic bias (if any) in our data would serve to strengthen the validity of ADRs discovered by our method.

Our work falls within the domain of infodemiology; that is, the study of Internet media to inform public health and policy [15]. Much previous work in this area has centered on detection and characterization of transient events (ie, disease outbreaks [15] and special events [40]) and the analysis of the kinds of information available to users [41,42]. Our paper is novel in that it makes use of search engine queries to identify transient events at the individual level and, more importantly, to discover associations between events [15] that eluded detection by the patients themselves or their health practitioners.

Limitations

The main drawback of relying on Web search data is that it is inherently noisy. It is often impossible to ascertain whether a

person searching for drugs and ADRs is doing so out of curiosity, or conducting research for himself, a relative, or even for a patient. Admittedly, Internet users comprise a biased sample of the population, and so the ADRs discovered may not be fully representative of the entire population. Nonetheless, our results suggest that the sheer size of the data alleviates these concerns, and the proposed method is able to identify adverse effects of drugs that are not captured by existing surveillance mechanisms.

Another limitation of this study is using a restricted set of symptoms expanded through the use of synonyms. Although a larger dictionary would have allowed identification of additional (and possibly rarer) ADRs, our focus on more common symptoms is likely to lead to better identification of the more common concerns to patients. Future work will focus on professionally used term dictionaries to focus on more knowledgeable patients and health providers. Another way to strengthen our results is the use of non-English search data, which will increase the volume of data (and the size of the observed population); thus, enabling the analysis of less frequent drugs and ADRs.

Finally, although this work is based on data from a large Internet search engine, it does not cover the entire population. However,

privacy concerns preclude conducting our analysis across search engines, as the latter never share information about their users. Nevertheless, given the sheer number of users whose data was analyzed in the study (176 million, which is especially notable compared to most other pharmacovigilance studies), we believe our findings are still significant. It should also be emphasized that QLRs discovers ADRs via aggregating queries across multiple users and query sessions. Consequently, the output of our method comes in the form of a list of newly discovered ADRs for each drug, and does not include any private, personal, or user-specific data whatsoever.

Conclusions

Our approach is mostly language-independent except for the initial list of symptoms [20], and obtaining the latter from non-English versions of Wikipedia will allow one to apply the method to additional languages, markets, and populations. Extending the coverage is particularly important for studying rare events, such as ADRs in patients who take many prescription drugs. Computational pharmacovigilance, which uses observational data such as Web search query logs, is complementary to the existing data collection mechanisms, and the ADRs it identifies should be considered as candidates for further investigation.

Acknowledgments

We thank Eugene Agichtein, Elmer Bernstam, David Carmel, Tania Fuchs, Aleksandr Kagan, Bo Pang, and Michael Schwarz for comments.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Identification of adverse drug reactions in vaccines.

[[PDF File \(Adobe PDF File\), 40KB - jmir_v15i6e124_app1.pdf](#)]

References

1. Ray WA, Stein CM. Reform of drug regulation--beyond an independent drug-safety board. *N Engl J Med* 2006 Jan 12;354(2):194-201. [doi: [10.1056/NEJMs053432](#)] [Medline: [16407517](#)]
2. Lary JM, Daniel KL, Erickson JD, Roberts HE, Moore CA. The return of thalidomide: can birth defects be prevented? *Drug Saf* 1999 Sep;21(3):161-169. [Medline: [10487395](#)]
3. Furberg CD, Pitt B. Withdrawal of cerivastatin from the world market. *Curr Control Trials Cardiovasc Med* 2001;2(5):205-207 [[FREE Full text](#)] [Medline: [11806796](#)]
4. Gale E. Troglitazone: the lesson that nobody learned? *Diabetologia* 2006 Jan;49(1):1-6. [doi: [10.1007/s00125-005-0074-6](#)] [Medline: [16362281](#)]
5. Topol E. Failing the public health--rofecoxib, Merck, and the FDA. *N Engl J Med* 2004 Oct 21;351(17):1707-1709. [doi: [10.1056/NEJMp048286](#)] [Medline: [15470193](#)]
6. Platt R, Wilson M, Chan KA, Benner JS, Marchibroda J, McClellan M. The new Sentinel Network--improving the evidence of medical-product safety. *N Engl J Med* 2009 Aug 13;361(7):645-647. [doi: [10.1056/NEJMp0905338](#)] [Medline: [19635947](#)]
7. Trifiro G, Fourrier-Reglat A, Sturkenboom MC, Díaz Acedo C, Van Der Lei J, EU-ADR Group. The EU-ADR project: preliminary results and perspective. *Stud Health Technol Inform* 2009;148:43-49. [Medline: [19745234](#)]
8. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc* 2012;19(1):54-60 [[FREE Full text](#)] [doi: [10.1136/amiajnl-2011-000376](#)] [Medline: [22037893](#)]

9. Madigan D, Ryan P. What can we really learn from observational studies?: the need for empirical assessment of methodology for active drug safety surveillance and comparative effectiveness research. *Epidemiology* 2011 Sep;22(5):629-631. [doi: [10.1097/EDE.0b013e318228ca1d](https://doi.org/10.1097/EDE.0b013e318228ca1d)] [Medline: [21811110](#)]
10. Stang PE, Ryan PB, Racoosin JA, Overhage JM, Hartzema AG, Reich C, et al. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. *Ann Intern Med* 2010 Nov 2;153(9):600-606. [doi: [10.1059/0003-4819-153-9-201011020-00010](https://doi.org/10.1059/0003-4819-153-9-201011020-00010)] [Medline: [21041580](#)]
11. Ryan PB, Welebob E, Hartzema AG, Stang PE, Overhage JM. Surveying US observational data sources and characteristics for drug safety needs. *Pharm Med* 2012 Aug 23;24(4):231-238. [doi: [10.1007/BF03256821](https://doi.org/10.1007/BF03256821)]
12. Backstrom L, Kleinberg J, Kumar R, Novak J. Spatial variation in search engine queries. 2008 Presented at: WWW2008; April 21-25, 2008; Beijing, China p. 366.
13. Richardson M. Learning about the world through long-term query logs. *ACM Trans Web* 2008 Oct 01;2(4):1-27 [FREE Full text] [doi: [10.1145/1409220.1409224](https://doi.org/10.1145/1409220.1409224)]
14. Ginsberg J, Mohebbi M, Patel R, Brammer L, Smolinski M, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2008;457:1012-1014. [doi: [10.1038/nature07634](https://doi.org/10.1038/nature07634)]
15. 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](https://doi.org/10.2196/jmir.1157)] [Medline: [19329408](#)]
16. Fox S. Health topics 2011. Washington, DC: Pew Internet & American Life Project; 2011 Feb 01. URL: <http://www.pewinternet.org/Reports/2011/HealthTopics.aspx> [accessed 2013-05-29] [WebCite Cache ID 6Gz3XPrf2]
17. Weber I, Jaimes A. Who uses web search for what? And how? In: WSDM11: Fourth ACM international conference on Web search and data mining. 2011 Presented at: WSDM; February 9-12, 2011; Hong Kong, China p. 15-24.
18. Humphreys A. MedAdNews. 2007. MedAdNews 200 World's Best Selling Medicines URL: http://en.wikipedia.org/wiki/List_of_bestselling_drugs [accessed 2013-05-29] [WebCite Cache ID 6Gz29e52G]
19. World Health Organization. 1994. ICD10 Symptom Checklist for Mental Disorders URL: http://www.who.int/substance_abuse/research_tools/icd10symptomchecklist/en/ [accessed 2013-06-07] [WebCite Cache ID 6HCWRINT6]
20. Wikipedia. List of medical symptoms URL: http://en.wikipedia.org/wiki/List_of_medical_symptoms [accessed 2013-05-29] [WebCite Cache ID 6Gz3PUHmO]
21. Hang Cui, Ji-Rong Wen, Jian-Yun Nie, Wei-Ying Ma. Query expansion by mining user logs. *IEEE Trans Knowl Data Eng* 2003 Jul;15(4):829-839 [FREE Full text] [doi: [10.1109/TKDE.2003.1209002](https://doi.org/10.1109/TKDE.2003.1209002)]
22. Baeza-Yates R, Tiberi A. Extracting semantic relations from query logs. In: KDD '07 Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. 2007 Presented at: KDD '07; 2007; Warsaw, Poland p. 76-85. [doi: [10.1145/1281192.1281204](https://doi.org/10.1145/1281192.1281204)]
23. Van Der Heijden M, Hinne M, Kraaij W, Verberne S, van der Weide T. Using query logs and click data to create improved document descriptions. In: WSCD '09 Proceedings of the 2009 workshop on Web Search Click Data. 2009 Presented at: WSCD '09; February 9-12, 2009; Barcelona, Spain p. 64-67. [doi: [10.1145/1507509.1507519](https://doi.org/10.1145/1507509.1507519)]
24. Carmel D, Farchi E, Petruschka Y, Soffer A. Automatic query refinement using lexical affinities with maximal information gain. In: SIGIR '02 Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval. 2002 Presented at: SIGIR; August 11-15, 2002; Tampere, Finland p. 290.
25. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol bull* 1971;76(5):378 [FREE Full text]
26. Upton G, Cook I. *Oxford Dictionary of Statistics*. Oxford, UK: Oxford University Press; 2002.
27. Kuhn M, Campillos M, Letunic I, Jensen L, Bork P. A side effect resource to capture phenotypic effects of drugs. *Mol Syst Biol* 2010;6:343 [FREE Full text] [doi: [10.1038/msb.2009.98](https://doi.org/10.1038/msb.2009.98)] [Medline: [20087340](#)]
28. US Food and Drug Administration. FDA Adverse Event Reporting System (FAERS): Latest Quarterly Data Files URL: <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm082193.htm> [accessed 2013-06-07] [WebCite Cache ID 6HCVNNWHO]
29. DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *American statistician* 1999;53:177-189.
30. Zorych I, Madigan D, Ryan P, Bate A. Disproportionality methods for pharmacovigilance in longitudinal observational databases. *Stat Methods Med Res* 2013 Feb;22(1):39-56. [doi: [10.1177/0962280211403602](https://doi.org/10.1177/0962280211403602)] [Medline: [21878461](#)]
31. Szarfman A, Machado SG, O'Neill RT. Use of screening algorithms and computer systems to efficiently signal higher than expected combinations of drugs and events in the US FDA's spontaneous reports database. *Drug Safety* 2004;25:381-392 [FREE Full text]
32. Wilson AM, Thabane L, Holbrook A. Application of data mining techniques in pharmacovigilance. *Brit J Clin Pharmacology* 2003;57:127-134. [doi: [10.1046/j.1365-2125.2003.01968.x](https://doi.org/10.1046/j.1365-2125.2003.01968.x)]
33. Schulze R. *Meta-analysis: A Comparison of Approaches*. Cambridge, MA: Hogrefe & Huber; 2004.
34. Duda R, Hart P, Strok DG. *Pattern Classification*. New York: John Wiley & Sons; 2001.
35. van Rijsbergen CJ. *Information Retrieval* (2nd ed). London: Butterworths; 1979.
36. Demsar J. Statistical comparisons of classifiers over multiple data sets. *JMLR* 2006;7:1-30 [FREE Full text]

37. HAARETZ. Israel drug company summoned to hearing over thyroid drug Eltroxin URL: <http://www.haaretz.com/news/national/israel-drug-company-summoned-to-hearing-over-thyroid-drug-eltroxin-1.409067> [accessed 2013-05-29] [WebCite Cache ID 6Gz3Q3n5f]
38. Rouleau B, Lavoie L, Leblanc J, Moretti S, Collin C. Reporting of adverse drug reactions by community pharmacists: a qualitative study in Quebec. *Drug Inf J* 2011;45(5):627-639. [doi: [10.1177/009286151104500613](https://doi.org/10.1177/009286151104500613)]
39. US National Library of Medicine, National Institutes of Health. RxNorm overview URL: <http://www.nlm.nih.gov/research/umls/rxnorm/overview.html> [accessed 2013-05-29] [WebCite Cache ID 6Gz3SbuCk]
40. White RW, Tatonetti NP, Shah NH, Altman RB, Horvitz E. Web-scale pharmacovigilance: listening to signals from the crowd. *J Am Med Inform Assoc* 2013;1-5.
41. 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](https://doi.org/10.2196/jmir.2148)] [Medline: [22634568](https://pubmed.ncbi.nlm.nih.gov/22634568/)]
42. Hill S, Mao J, Ungar L, Hennessy S, Leonard CE, Holmes J. Natural supplements for H1N1 influenza: retrospective observational infodemiology study of information and search activity on the Internet. *J Med Internet Res* 2011;13(2):e36 [FREE Full text] [doi: [10.2196/jmir.1722](https://doi.org/10.2196/jmir.1722)] [Medline: [21558062](https://pubmed.ncbi.nlm.nih.gov/21558062/)]

Abbreviations

ADR: adverse drug reaction
AERS: Adverse Event Reporting System
AUC: area under the curve
CDF: cumulative density functions
EBGM: empirical Bayes geometric mean
FDA: Food and Drug Administration
IC: information component
MDADR: most discordant adverse drug reactions
MHRA: Medicines and Healthcare products Regulatory Agency
QLRS: query log reaction score
RADAR: Research on Adverse Drug Events and Reports
ROC: receiver operating curve
SIDER: Side Effect Resource
VAERS: Vaccine Adverse Event Reporting System
WHO: World Health Organization

Edited by G Eysenbach; submitted 14.03.13; peer-reviewed by CH Li, S Sahay; comments to author 07.04.13; revised version received 26.04.13; accepted 23.05.13; published 18.06.13.

Please cite as:

Yom-Tov E, Gabrilovich E

Postmarket Drug Surveillance Without Trial Costs: Discovery of Adverse Drug Reactions Through Large-Scale Analysis of Web Search Queries

J Med Internet Res 2013;15(6):e124

URL: <http://www.jmir.org/2013/6/e124/>

doi: [10.2196/jmir.2614](https://doi.org/10.2196/jmir.2614)

PMID: [23778053](https://pubmed.ncbi.nlm.nih.gov/23778053/)

©Elad Yom-Tov, Evgeniy Gabrilovich. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 18.06.2013. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Collection and Visualization of Dietary Behavior and Reasons for Eating Using Twitter

Melanie Hingle^{1*}, PhD, MPH, RD; Donella Yoon^{1*}, BSc; Joseph Fowler², PhD; Stephen Kobourov², PhD; Michael Lee Schneider², BSc; Daniel Falk³, MA; Randy Burd¹, PhD

¹University of Arizona, Department of Nutritional Science, University of Arizona, Tucson, AZ, United States

²University of Arizona, Department of Computer Science, University of Arizona, Tucson, AZ, United States

³Niagara Computer Services, Inc., Jenkintown, PA, United States

*these authors contributed equally

Corresponding Author:

Randy Burd, PhD

University of Arizona

Department of Nutritional Science

University of Arizona

1177 E. 4th Street

Tucson, AZ, 85721

United States

Phone: 1 520 626 1863

Fax: 1 520 621 9446

Email: rburd@email.arizona.edu

Abstract

Background: Increasing an individual's awareness and understanding of their dietary habits and reasons for eating may help facilitate positive dietary changes. Mobile technologies allow individuals to record diet-related behavior in real time from any location; however, the most popular software applications lack empirical evidence supporting their efficacy as health promotion tools.

Objective: The purpose of this study was to test the feasibility and acceptability of a popular social media software application (Twitter) to capture young adults' dietary behavior and reasons for eating. A secondary aim was to visualize data from Twitter using a novel analytic tool designed to help identify relationships among dietary behaviors, reasons for eating, and contextual factors.

Methods: Participants were trained to record all food and beverages consumed over 3 consecutive days (2 weekdays and 1 weekend day) using their mobile device's native Twitter application. A list of 24 hashtags (#) representing food groups and reasons for eating were provided to participants to guide reporting (eg, #protein, #mood). Participants were encouraged to annotate hashtags with contextual information using photos, text, and links. User experience was assessed through a combination of email reports of technical challenges and a 9-item exit survey. Participant data were captured from the public Twitter stream, and frequency of hashtag occurrence and co-occurrence were determined. Contextual data were further parsed and qualitatively analyzed. A frequency matrix was constructed to identify food and behavior hashtags that co-occurred. These relationships were visualized using GMap algorithmic mapping software.

Results: A total of 50 adults completed the study. In all, 773 tweets including 2862 hashtags (1756 foods and 1106 reasons for eating) were reported. Frequently reported food groups were #grains (n=365 tweets), #dairy (n=221), and #protein (n=307). The most frequently cited reasons for eating were #social (activity) (n=122), #taste (n=146), and #convenience (n=173). Participants used a combination of study-provided hash tags and their own hash tags to describe behavior. Most rated Twitter as easy to use for the purpose of reporting diet-related behavior. "Maps" of hash tag occurrences and co-occurrences were developed that suggested time-varying diet and behavior patterns.

Conclusions: Twitter combined with an analytical software tool provides a method for capturing real-time food consumption and diet-related behavior. Data visualization may provide a method to identify relationships between dietary and behavioral factors. These findings will inform the design of a study exploring the use of social media and data visualization to identify

relationships between food consumption, reasons for engaging in specific food-related behaviors, relevant contextual factors, and weight and health statuses in diverse populations.

(*J Med Internet Res* 2013;15(6):e125) doi:[10.2196/jmir.2613](https://doi.org/10.2196/jmir.2613)

KEYWORDS

dietary behavior; data visualization; social media; mobile health; mHealth

Introduction

The high prevalence of obesity continues to be a major public health challenge in the United States [1], contributing to increased incidence of weight-related comorbidities (eg cancer, type 2 diabetes, cardiovascular disease) and early mortality [2]. The societal burden of obesity is substantial, including loss of productivity, increased health care costs, and decreased quality of life [3]. Collectively, these economic and social costs significantly threaten the stability of our society.

The dramatic increase in obesity over the past 3 decades is attributed to a mismatch between humans' evolved physiology and the contemporary environment, characterized by an abundance of high-calorie foods with few opportunities for equivalent energy expenditure [4]. Within this obesogenic environment, individuals must exert a significant and sustained level of self-regulation to avoid weight gain.

It has been suggested that an overall increased availability of calories (and subsequent increased energy intake) can explain a significant proportion of the increase in obesity prevalence at the population level [5]. Data from nationally representative diet and health behavior surveys (Nationwide Food Consumption Survey, 1977-78; the Continuing Survey of Food Intakes of Individuals, 1989-91; and the National Health and Nutrition Examination Surveys, 1994-98 and 2003-06) suggest that increases in energy intake over the past 30 years may be attributed to an increased frequency of eating and drinking occasions (in particular, snacking occasions) and larger portion sizes [5]. These 2 dietary behaviors represent specific and modifiable intervention targets, whereby excess energy intake may be reduced and excess weight gain prevented if strategic reductions are achieved.

Despite extensive evidence-based dietary guidance directed at consumers (eg, Healthy People 2010 and the Dietary Guidelines [6,7]), implementation of healthy lifestyle habits by Americans remains lacking [8]. Nutrition education research has established that knowledge about healthy eating is necessary but not sufficient to change behavior [9]. Qualitative and quantitative studies of eating behavior have confirmed that decisions around eating behavior are often unconscious and based on established habits [10]. As a result, dietary choices are often reactively determined by where, when, and with whom an individual eats rather than a product of conscious planning. Given the disproportionate availability of unhealthy compared to healthy foods, there is a high probability that the most readily available foods are not low calorie or nutrient-dense choices.

In the midst of this Information Age [11], the rapid advent of wireless technologies has already begun to revolutionize personal health. The number of smartphone users is increasing

and wireless devices are rapidly expected to become the number one way by which consumers will access information and resources on the World Wide Web [11]. This rise of mobile health devices and software applications is projected to continue well into the next 5 years, wherein over 80% of wireless device purchases in 2016 will be for systems to monitor and improve health [12]. The ubiquity of mobile device ownership [13] has placed the power to influence health directly into the hands of consumers, including the ability to record and access personalized health information continuously, in real time, and from any location. At the same time, ready access to the large amounts of data collected through mobile devices and applications remains one of the greatest challenges facing the utility of mobile health (mHealth). The use of mobile devices and the resultant big data to influence health behavior change can only be realized if the data are translated into easily understood and actionable steps in real life. Mobile interfaces which provide users with access to their own data in ways that will help them to make optimal choices for their age, life stage, location, culture, beliefs, and resources is a critical—and often overlooked—aspect of existing mHealth applications.

Despite a large number of software applications available in mobile devices (iOS, Android, Microsoft, and Blackberry) to help individuals track eating behavior and make healthy food choices, most lack empirical data supporting their utility as health promotion tools. In addition, commercial applications often use proprietary methods of calculating energy intake/expenditure or restrict the type and amount of data that are collected and provided to the individual user which may obscure critical information from the user and perpetuate misinformation about dietary behaviors, nutrition, and health.

Changing a pattern of established behavior—particularly diet—is a complex process, and barriers to engaging in and maintaining dietary behavior change are well documented [14]. Extensive research suggests that eliminating an undesirable behavior does not involve unlearning, but rather *new* learning, and the physical environment and social context in which learning takes place (eg, emotions, time) will influence the learning process [15]. Thus, strategies designed to help individuals engage in and maintain changes in dietary behavior should acknowledge and address this complexity, providing critical information and feedback when most needed, including when, where, and with whom food choices and eating behaviors are occurring. A first step toward understanding the complexity of eating behavior is developing and testing a method to capture these critical data.

Therefore, the purpose of this study was to test the feasibility and acceptability (including usability) of a popular and free social media application (Twitter) to capture food and beverage consumption behavior and reasons for eating over 3 days in a

group of young adults. A secondary aim was to capture and analyze user data from Twitter using a novel analytic tool, and visualize these data using behavioral maps to identify patterns in intake and behavior, and potential relationships between foods and contextual factors.

We posited that by tracking typical dietary choices and reasons for making food-related decisions simultaneously and in real time, relationships between these factors would emerge, and the addition of contextual factors, such as when, where, and with whom a person eats, would enrich these data providing additional information related to diet quality and potential mediators of dietary behavior.

Methods

Design, Participants, Recruitment, and Setting

Fifty adults were recruited from the University of Arizona between February and April 2012. Recruitment activities consisted of email advertisements sent to undergraduate listservs and announcements in undergraduate and graduate student classes. Inclusion criteria for the study included ownership of a Web-enabled mobile phone or smartphone, willingness to use a study-designated Twitter account for all study activities, and age older than 18 years at time of enrollment.

Respondents were invited to attend information sessions at the University of Arizona's Department of Nutritional Sciences and informed consent was obtained from those who were interested and eligible. Participants were provided instructions on how to access Twitter and activate a study Twitter account using a nonidentifiable study-assigned Twitter handle, and assigned a study orientation date.

Procedures

During the 1-hour study orientation session, participants were trained to use Twitter to record all food and beverages consumed over 2 weekdays and 1 weekend day, and their reasons for choosing these foods and beverages. Participants were asked to report these data using hashtags (#), Twitter's electronic shorthand that is used to describe people, places, and things in a consistent manner. A list of 24 hashtags plus examples were provided to guide reporting and categorization of foods and reasons for eating. The list was informed by nomenclature used to describe food groups in nutrition education materials developed by the US Department of Agriculture (ie, the ChooseMyPlate) and reasons for eating commonly reported in the behavioral literature [16]. Additionally, participants were instructed to annotate hashtags with descriptive information (eg, short phrases, photos, or links) to provide additional information about where, when, why, and with whom specific diet behaviors occurred.

Following the orientation session, participants recorded their diet and reasons for eating for 3 consecutive days using their study-provided username and password and their mobile phone's native Twitter application. All subsequent Twitter messages (tweets) were visible on the public Twitter stream, but authors were not identifiable.

The usability and acceptability of this approach was evaluated using a comprehensive evaluation approach that involved 2 complementary methods [17], real-time feedback from participants who were encouraged to email the study coordinator with technical problems or questions during the 3 days, combined with a 9-item exit survey that evaluated acceptability (satisfaction) of the experience after completion of data collection (see [Multimedia Appendix 1](#)). Survey responses were collected using SurveyMonkey [7], a commercially available Web-based service. Questions included participant preferences regarding Twitter access (ie, mobile platform vs desktop interface) and an overall rating of the application's usability (eg, how easy/difficult it was to record their food intake using the application). Participants were also asked whether they felt adequately trained to use the application within the context of the study, and how—if at all—they would like to see the results of their tracking displayed. Participants were provided US \$20 cash incentive upon completion of the exit survey. Permission to conduct the study was obtained from the Institutional Review Board at the University of Arizona.

Data Analysis

Participant data were downloaded from the public Twitter stream using a novel Web-based data capture application (ViBE), developed specifically to identify and retrieve data from Twitter. For each participant, ViBE automatically calculates the frequency of hashtag occurrences in the Twitter stream during the designated study period, as well as the frequency of co-occurrences with other hashtags. Occurrence and co-occurrence data were merged and compiled as a normalized matrix. The matrix was exported from ViBE and visually processed using the algorithmic framework of GMap 1988 version 2.29.0 (AT&T Research) as a complete edge-weighted graph in which the Twitter hashtags and their matrix values correspond to the vertices and edge weights of the graph, respectively (see [Multimedia Appendix 2](#)). Co-occurrences are displayed on the resulting graph or map as adjoining "countries."

The canonical map was produced by GMap [17,18], included with the Graphviz [19,20] graph visualization software (AT&T Research), which is an intuitive visualization of the relationships between patterns of foods and beverages consumed, and diet-related behaviors. GMap maps have been used for a variety of behavioral applications, including understanding purchasing of Amazon.com consumers [17]. Using these maps, users are able to see behavioral patterns and infer relationships. GMap has been further augmented by AT&T Research to handle streaming feeds from Twitter [21] and addresses the challenge of real-time data collection and visualization [22].

Results

Feasibility

All participants were able to use Twitter and the study-provided list of hashtags to report their food and beverage intake and related dietary behaviors for 3 consecutive days. A total of 773 tweets containing 2862 hashtags were reported (1756 food-related hashtags and 1106 reasons for eating hashtags). The 3 most frequently reported food-related hashtags were #grains (21%), #protein (17%), and #dairy (13%), whereas the

most frequently reported reasons for eating were #convenience (16%), #taste (13%), and #social (11%) (Figures 1 and 2). A total of 164 errors in hashtag spelling were observed (141 food-related hashtags and 23 behavioral hashtags were misspelled). In addition, participants also created 30 of their own hashtags (which were reported 40 times during the course of the study; see Table 1).

Data Visualization

The ViBE software was used to retrieve data from Twitter and determine the frequency of occurrences and co-occurrences. This matrix was imported into the GMap software platform, and several canonical maps (Figures 3 and 5) were produced. The most frequently co-occurring hash tags were clustered into “countries” designated by a unique color. The data were separated into 4-hour time increments (6 am–10 am, 10 am–2 pm, 2 pm–6 pm, 6 pm–10 pm, 10 pm–2 am, and 2 am–6 am), and frequency of hashtag occurrence by time of day was explored over 24 hours. Frequency of hash tag reporting during 4-hour time blocks are represented by the intensity of color, in which darker colors represent more frequent hashtag reporting (Figures 4 and 6). The most active reporting periods were observed between 2 pm and 6 am (92% of all tweets).

Acceptability

During the 3-day data collection period, 3 of 50 participants contacted the study coordinator via email and Twitter to receive

help on classifying specific foods (2 questions), and constructing a hashtag that involved spaces (1 question). Additional survey data was collected from participants upon completion that suggested the Twitter platform was well received by participants overall. There was ambiguity regarding the use of several hash tags (eg, #appearance, which participants interpreted as either a person’s appearance or the food’s appearance). Participants also suggested that the ability to add their own hashtags should be an option to more precisely describe food choices and reasons for eating.

Of the 50 participants, 56% used only their mobile phones to access Twitter, 8% used their desktop or laptop computer, and 36% used a combination of mobile phone and desktop or laptop computer. Most participants who used both methods preferred the mobile phone to the computer. Although 73% (38/50) rated Twitter as very easy to use, 10% (5/50) reported that the 140-character limit imposed by the application made it a challenge to accurately report foods and reasons for eating. Many participants 36% (8/50) described the use of Twitter to record diet and behavior as a positive experience. Six participants (12%) recommended increasing the study duration to obtain a more comprehensive summary of their food intake and related dietary behavior, and 18 participants (36%) also commented about providing feedback to users about diet quality and suggestions on how to improve diet quality.

Figure 1. Frequency and timing of food hashtags over a 24-hour time period.

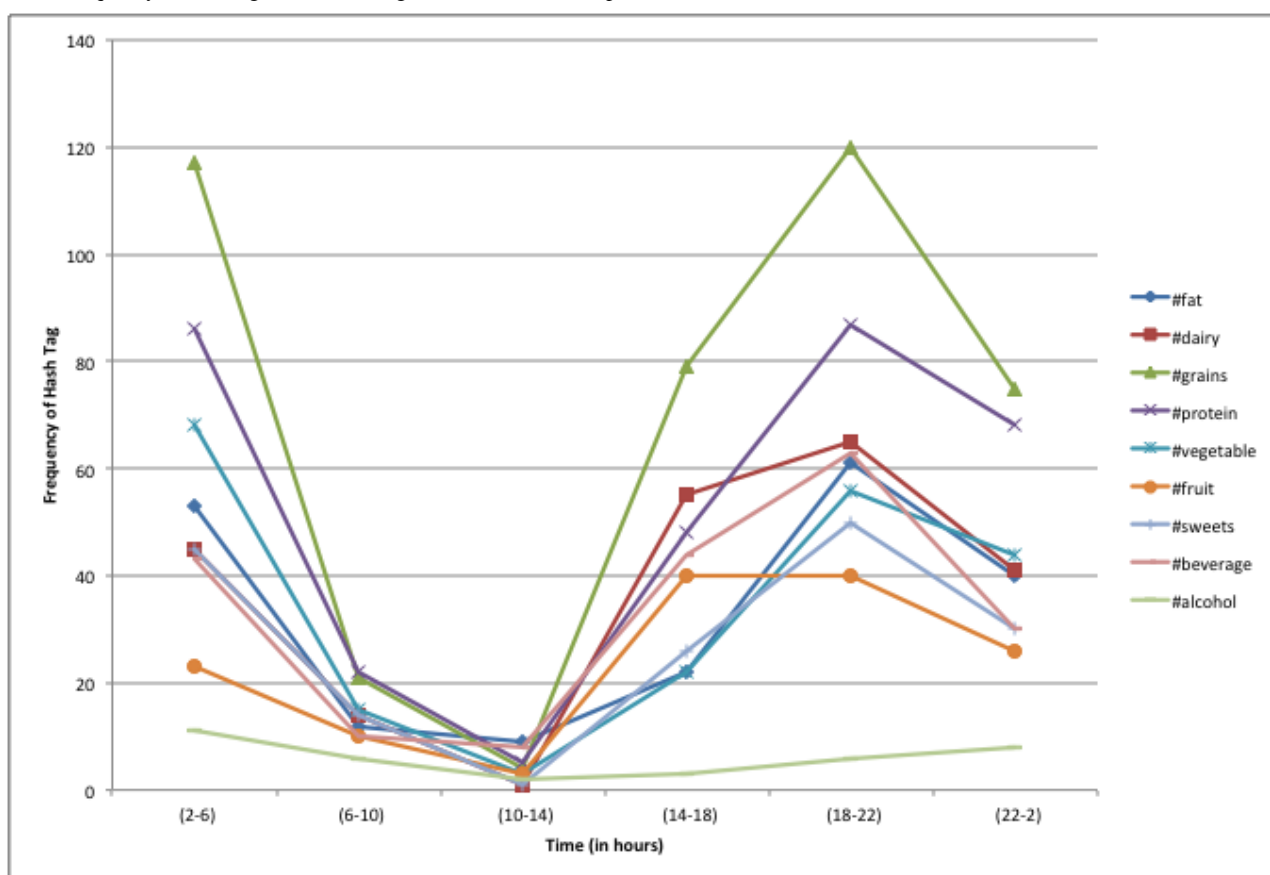


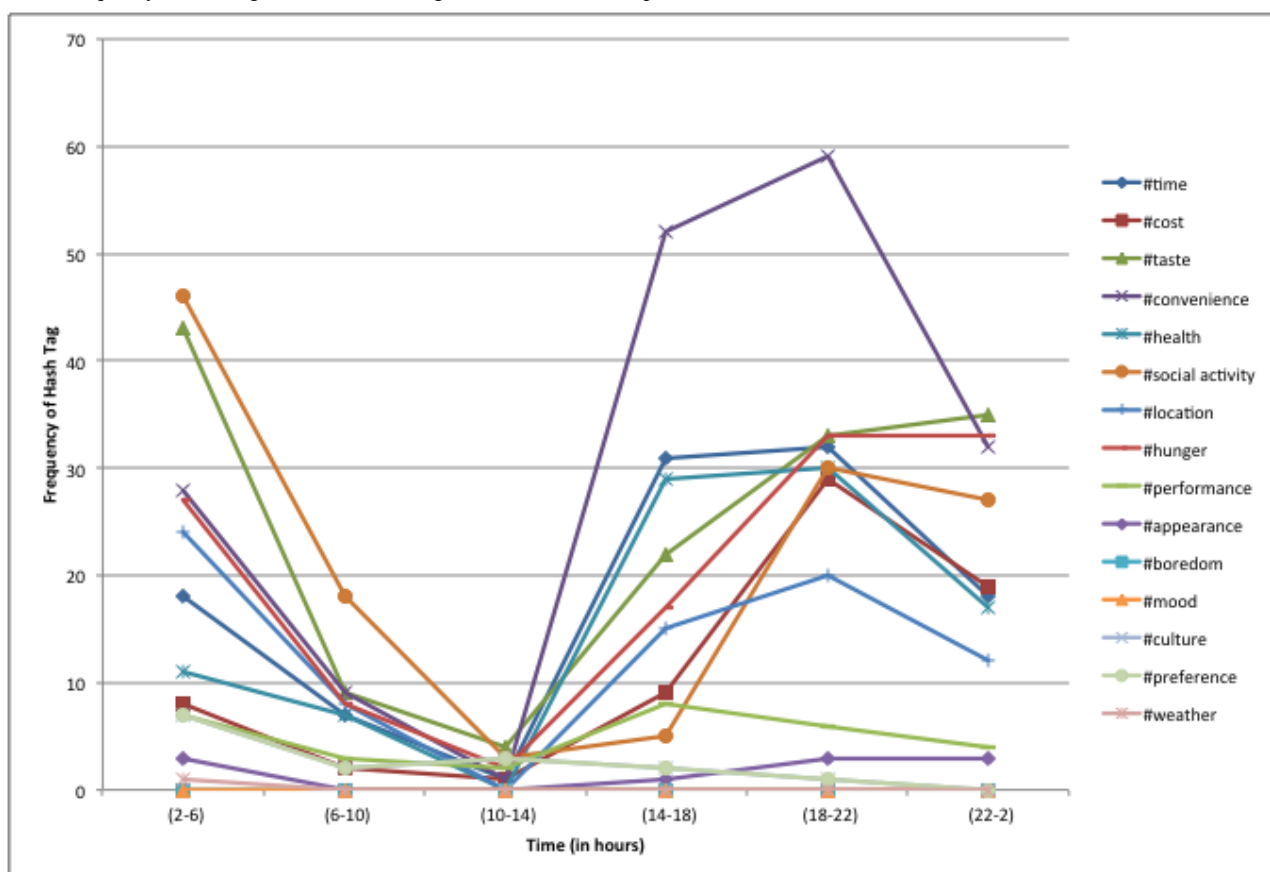
Figure 2. Frequency and timing of behavior hashtags over a 24-hour time period.

Table 1. Frequency of hashtags reported by participants (50 participants over 3 consecutive days).

Hashtags	Frequency
Study-provided	
Food-related	
#alcohol	41
#beverage	198
#dairy	221
#fat	174
#fruit	131
#grains	365
#protein	307
#sweets	150
#vegetable	169
Behavior-related	
#appearance	5
#boredom	0
#convenience	173
#cost	68
#culture	14
#health	94
#hunger	120
#location	79
#mood	103
#performance	30
#preference	41
#social	122
#taste	146
#time	109
#weather	2
Participant-generated	
Food-related	
#bread	1
#butter	1
#carbohydrates	2
#carbs	4
#coffee	1
#lotsoffat	1
#meat	4
#meats	1
Behavior-related	

Hashtags	Frequency
#2	1
#3-	1
#aapl	1
#breakfast	1
#comfort	1
#craving	2
#easy	1
#feed	1
#friend	1
#goingtofeelitinthemorning	1
#habit	2
#happy	1
#healthy	2
#home	1
#leftovers	1
#lunch	1
#price	1
#random	1
#still	1
#test	1
#timeegg	1
#vibestudy	1

Figure 3. Study-generated data collected were transformed into a co-occurrence matrix and then applied to a visually representative map. Common co-occurrence hashtags placed in a country denoted by different colors on the map. Frequency of hashtags is also shown with centrally located countries contributing to higher frequencies and peripheral countries contributing to lower frequencies of hashtags.

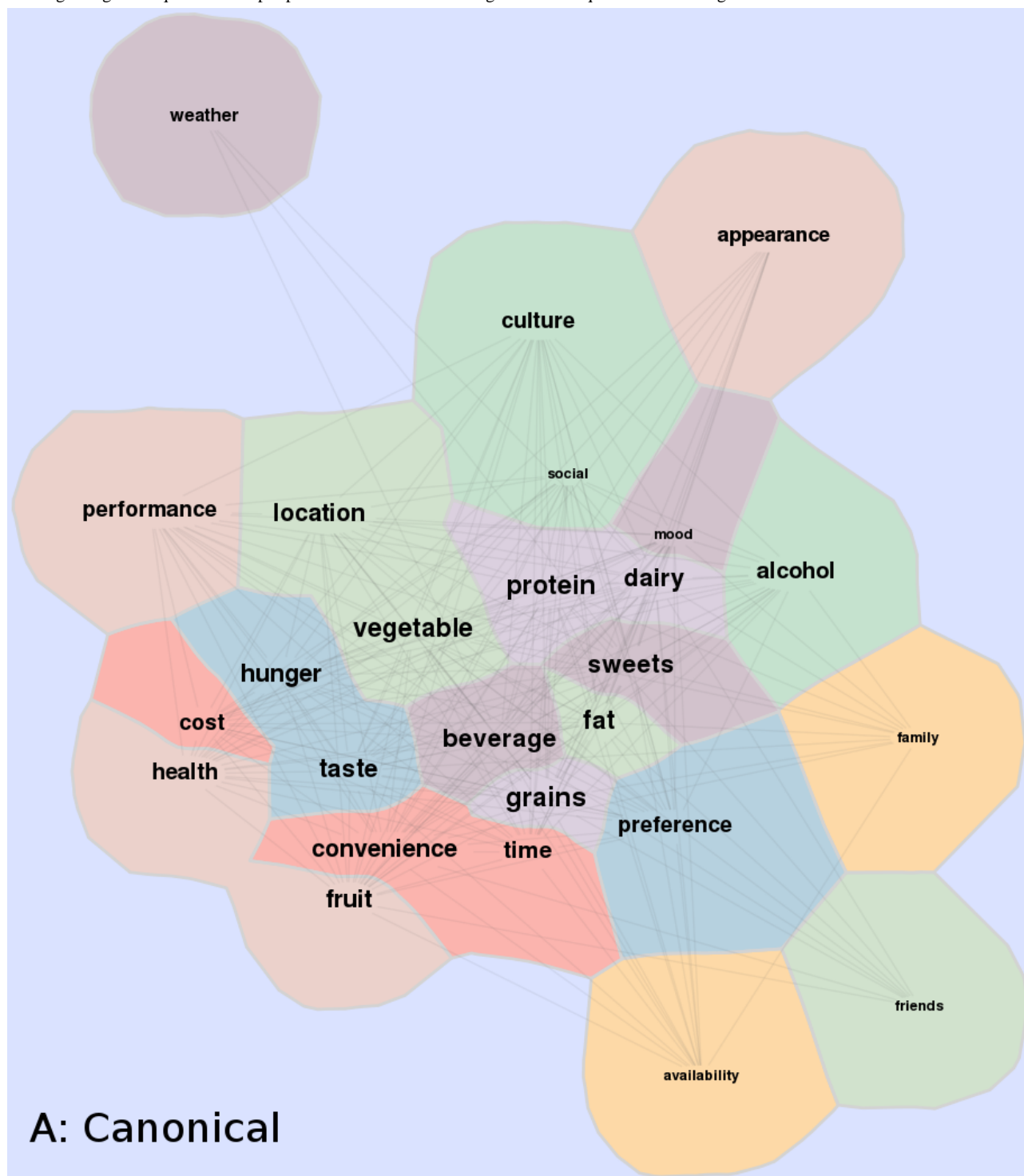


Figure 4. Heat maps showing frequencies throughout a 24-hour time period for study-generated data are shown. Higher frequencies are displayed with a darker blue hue, whereas progressively lower frequencies are lighter in blue. White coloration refers to little to no frequency activity.

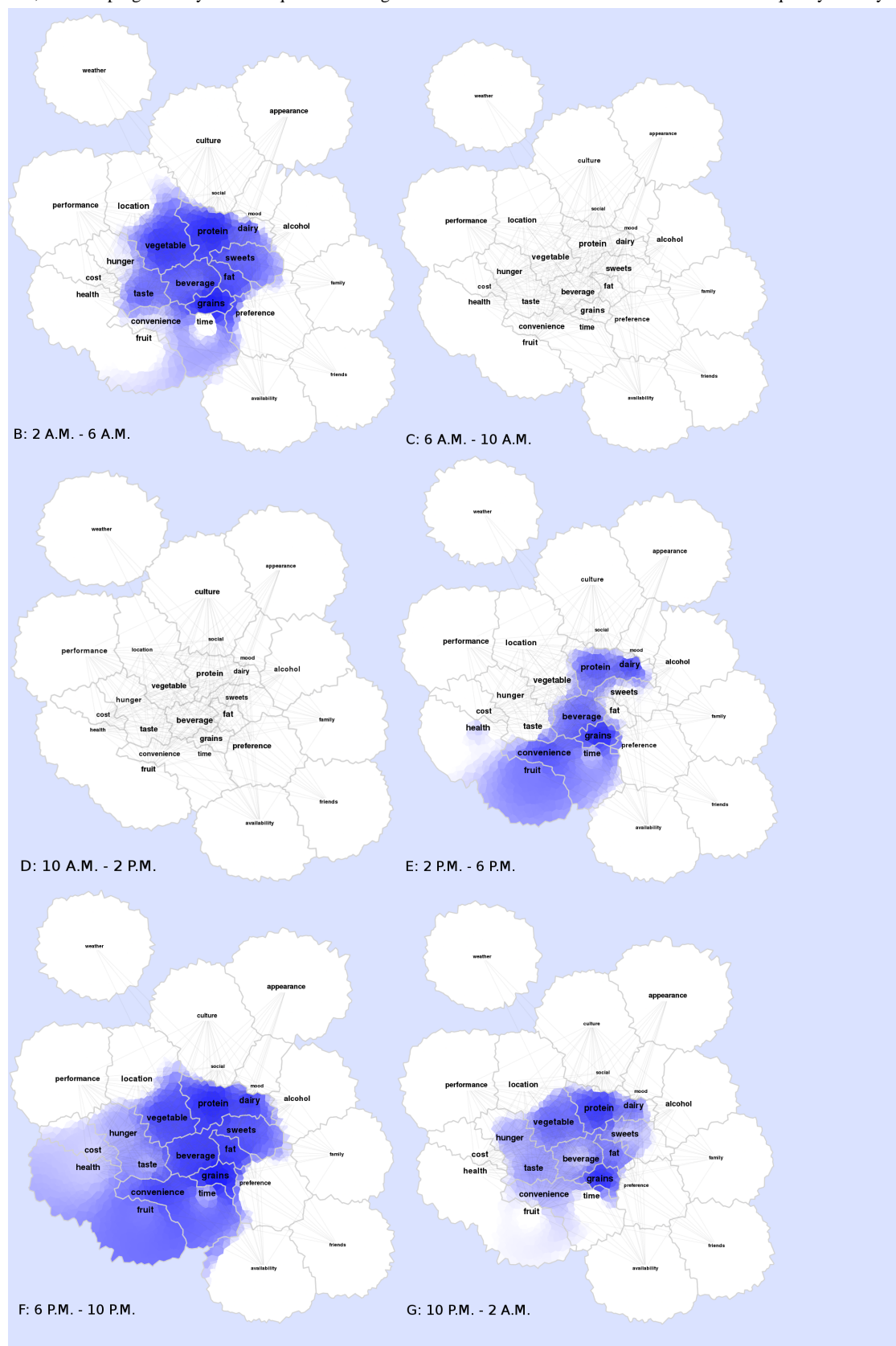


Figure 5. Study-generated hash-tags over a 24-hour time period. Study-generated data collected were transformed into a co-occurrence matrix and then applied to a visually representative map. Common co-occurrence hashtags placed in a country denoted by different colors on the map. Frequency of hashtags is also shown with centrally located countries contributing to higher frequencies and peripheral countries contributing to lower frequencies of hashtags.

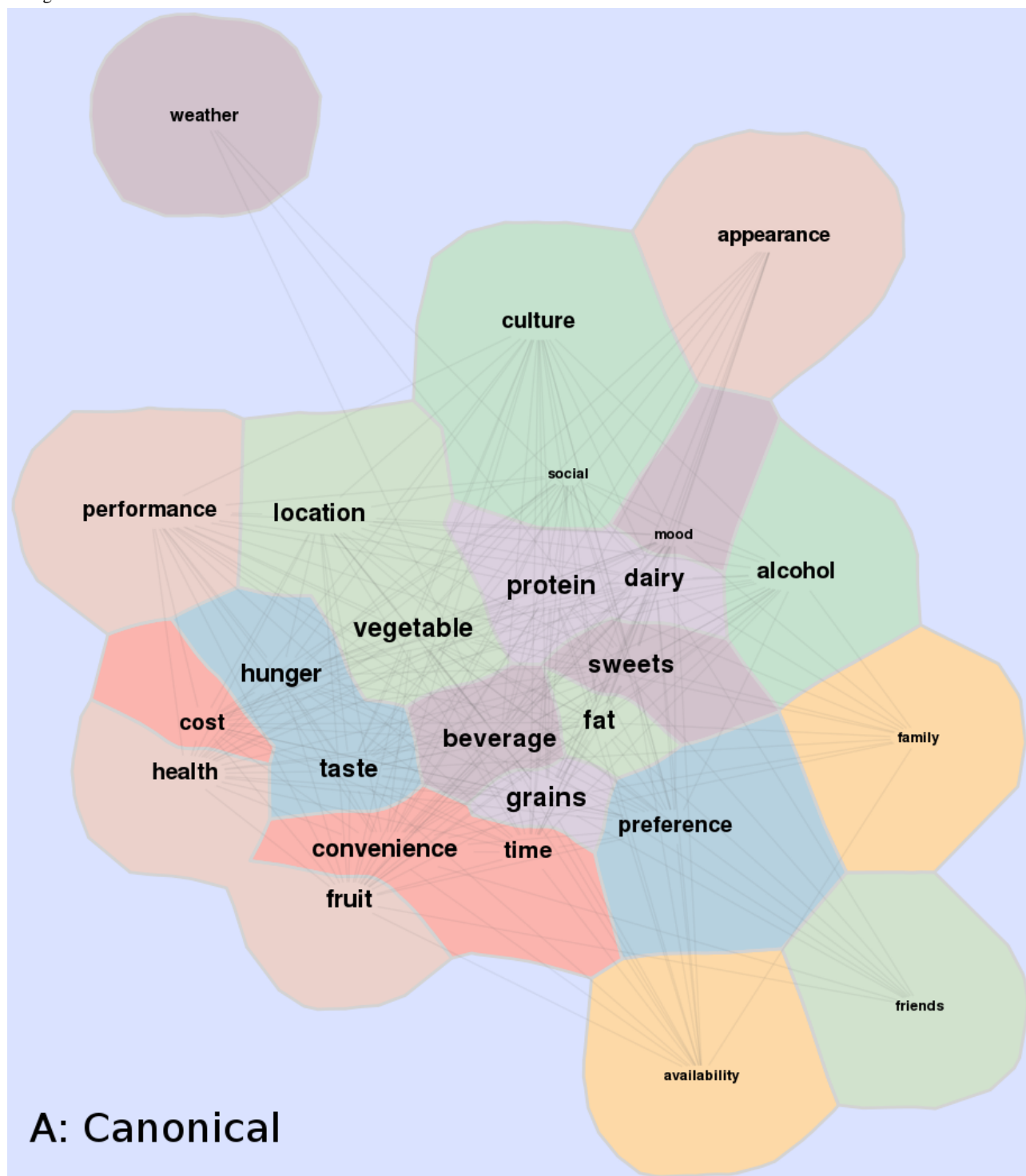


Figure 6. Study-generated hashtags over a 24-hour time period. Heat maps showing frequencies throughout a 24-hour time period for study-generated data are shown. Higher frequencies are displayed with a darker blue hue, whereas progressively lower frequencies are lighter in blue. White coloration refers to little to no frequency activity.

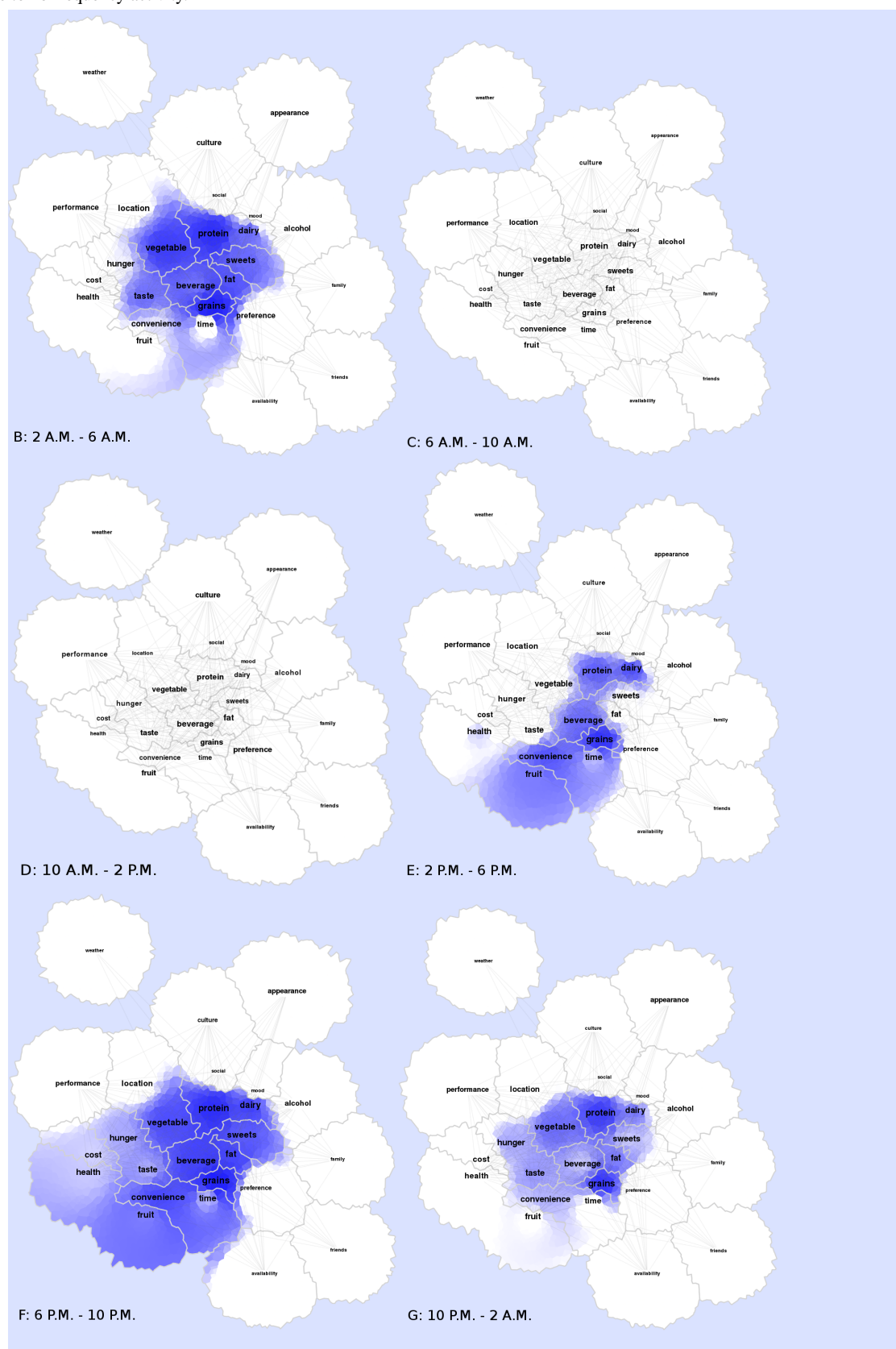


Figure 7. Study-generated and participant-provided hash tags over a 24-hour time period. Study-generated and participant-provided data collected were transformed into a co-occurrence matrix and then applied to a visually representative map. Common co-occurrence hashtags placed in a country denoted by different colors on the map. Frequency of hashtags is also shown with centrally located countries contributing to higher frequencies and peripheral countries contributing to lower frequencies of hashtags.

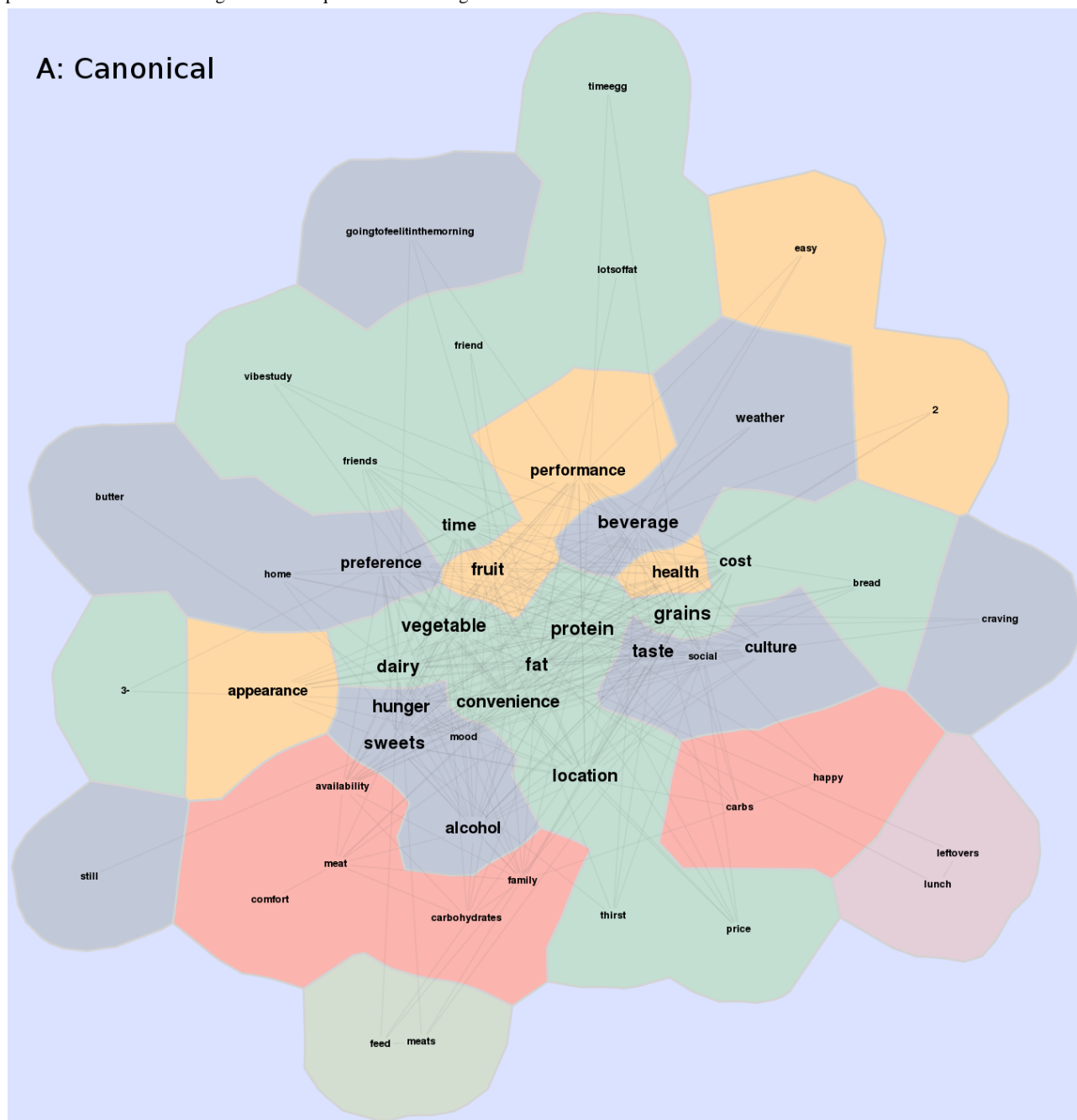
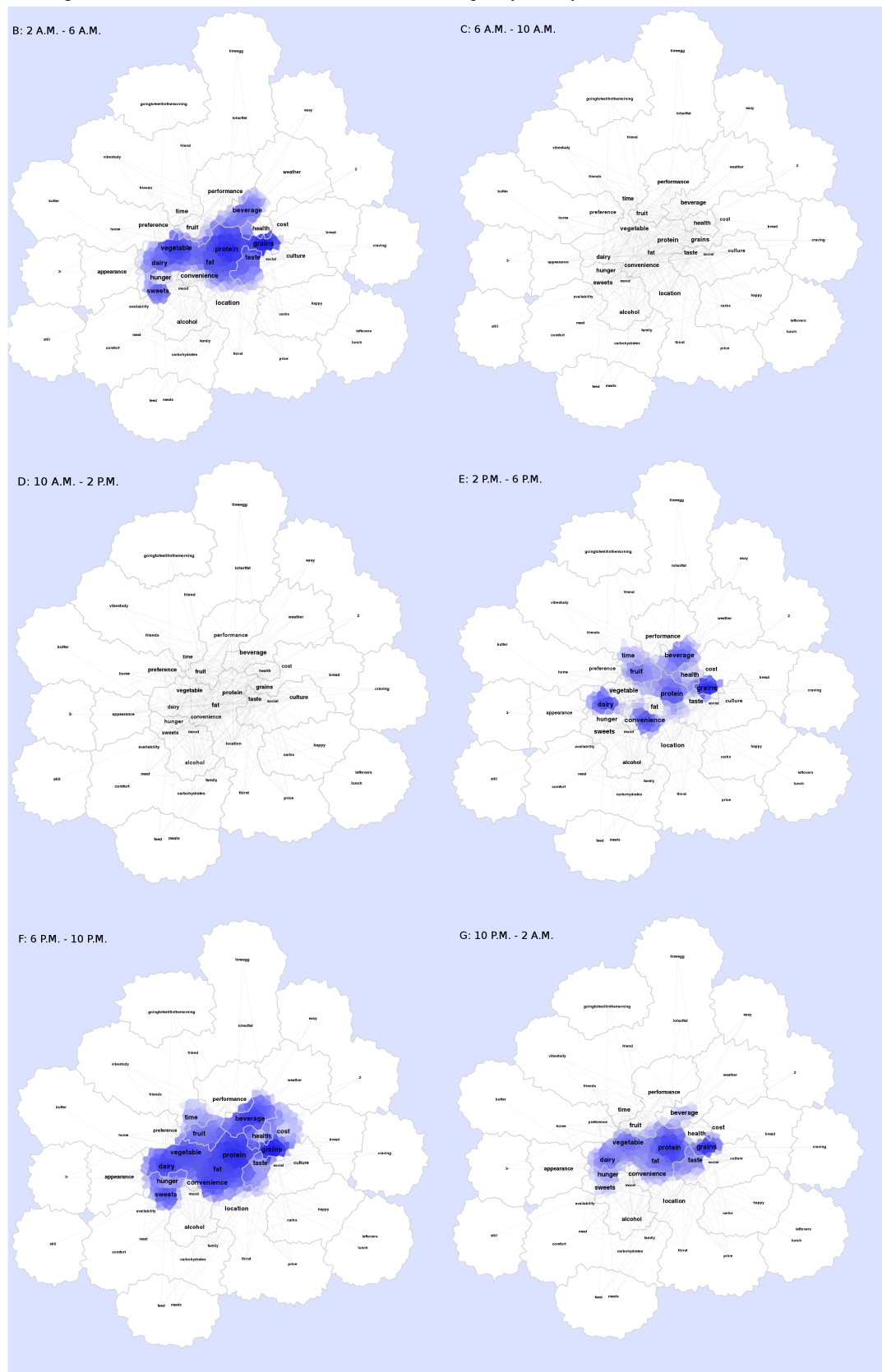


Figure 8. Study-generated and participant-provided hash tags over a 24-hour time period. Heat maps showing frequencies throughout a 24-hour time period are shown for study-generated and participant-provided data. Higher frequencies are displayed with a darker blue hue, whereas progressively lower frequencies are lighter in blue. White coloration refers to little to no frequency activity.



Discussion

Principal Findings

This study demonstrates that Twitter is a feasible and acceptable method for capturing real-time dietary intake and behavior, and that the resulting data can be used to visualize relationships between food consumption and reasons for engaging in these behaviors. This approach represents a novel method for collecting and characterizing dietary and behavioral data. In addition, this study demonstrates the capacity to visualize dietary behaviors using maps as well as adding valuable contextual dimensions, such as time of day and location. Identification of patterns of behavior and relationships between behaviors and time are important factors in understanding and creating strategies to influence established patterns of behavior, such as diet.

In the future, additional parameters could be integrated into heat maps to further enhance the visualizations, eg, phenotype, genotype, or sociodemographic data. Further illustrating the defining characteristics of a population will increase understanding of relationships between nutrition and health and disease states.

Our approach has utility for health interventions. Using these data collection and visualization techniques, data from individuals could be co-located on a map with others, leading to identification of common characteristics. As an example, tweets from a specified body mass index (BMI), such as those with a BMI greater than 25 could be color-coded and then compared to those with a BMI less than 25. An individual could see where their tweets appear on the map, the foods and behaviors associated with this location, and how they compare to others with similar or different characteristics. This could be a powerful tool for self-realization, identifying like-minded individuals to provide social or group support, or finding dissimilar groups who represent a goal likeness and exhibit specific behaviors around which goal setting could occur. Inclusion of additional parameters could also inform the design of tailored messages sent to the participant as part of specific interventions targeting at-risk individuals or groups [23,24].

The novel use of Twitter and subsequent data visualization can provide individuals and groups with what consumers of health information have come to expect from mobile health software and health information delivery—immediate, personalized, and participatory information. Twitter was chosen, in part, because it allows tweets from third-party applications (eg, from a range of mobile phone platforms) to be accepted and displayed. The personalized nature of our platform has the capacity to deliver specific and tailored interventions to individuals based on a number of complex and interrelated characteristics. The Twitter format allows for easy 2-way communication, facilitating informal interactions between individuals (eg, clinician-patient or participant-participant). The significant amounts of real-time data collected through this system could also be used to design disease prevention interventions based on predictive algorithms. The existing social media infrastructure provided by Twitter could also be capitalized upon if and when more social and group interactions are incorporated into the process. The use of

online social forums for disease prevention and health promotion activities has demonstrated utility in supporting health behavior change [25]. However, creating an appropriate environment for health behavior change through Twitter must be considered. Research suggests that as a delivery platform, the success of Twitter as a health behavior tool may be enhanced with the inclusion of tailored messaging that directly targets the individual [26,27]. We are currently incorporating this facet into our iPhone application.

A secondary objective of this study was to evaluate the acceptability of this method to collect dietary data. Participants indicated they would like to see their own data presented in a graph or similar format. Participants also wanted the ability to track their intake and reasons for eating over a longer duration to discern trends. Many users indicated that the most problematic part of the overall experience was remembering to use the study-provided list of hashtags. In this study, misspellings accounted for a small but important percentage of hashtags. This challenge has already been resolved through a feature on our iPhone application that provides participants with a list to choose from, thus eliminating need for manual entry of data.

Limitations

A technical limitation of our study was fragmentation in canonical maps (wherein countries of a similar color were scattered across the map instead of grouped together as a solitary country). This was a software limitation caused by our limited sample size (several thousand rather than several million observations, which the GMap software is capable of), and can be resolved with a larger number of observations.

Another potential limitation was the possibility that our participants had previous experience (and comfort) using Twitter. College-age individuals use Twitter more frequently than the general public (25% vs 8% in the US population). [13]. However, 25% is still relatively low; therefore, training was provided to all participants prior to data collection to insure everyone was equally capable and comfortable collecting and posting information using Twitter.

To increase usability, we have built our own custom iOS application to simplify the reporting process. Using our application, an individual has only to click once to input the appropriate hashtag. We plan to test the application to establish the evidence necessary to optimize design and demonstrate generalizability. As smartphone ownership and social media use is projected to continue to increase, we do not anticipate that participants outside college age will have major challenges using our application or approach. However, additional studies are needed to verify this.

Conclusions

Our findings suggest that the use of Twitter combined with a novel method of data visualization can provide a tool for identifying patterns in and relationships between food and reasons for eating, as well as additional parameters of interest (eg, temporal trends). Future research will establish the efficacy of this method in diverse populations, building the evidence base for mobile-inspired approaches to diet- and health-related

behavior change for preventing obesity and optimizing population health.

Acknowledgments

This work was supported in part by grant USDA#: 2010-38411-21348.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey.

[[PDF File \(Adobe PDF File\), 106KB - jmir_v15i6e125_app1.pdf](#)]

Multimedia Appendix 2

Details of data visualization.

[[PDF File \(Adobe PDF File\), 53KB - jmir_v15i6e125_app2.pdf](#)]

References

1. Flegal KM, Carroll MD, Kit BK, Ogden CL. Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999-2010. *JAMA* 2012 Feb 1;307(5):491-497. [doi: [10.1001/jama.2012.39](#)] [Medline: [22253363](#)]
2. Engbers R, de Caluwé LI, Stuyt PM, Fluit CR, Bolhuis S. Towards organizational development for sustainable high-quality medical teaching. *Perspect Med Educ* 2013 Feb;2(1):28-40. [doi: [10.1007/s40037-013-0043-6](#)] [Medline: [23670655](#)]
3. Newman ME. Modularity and community structure in networks. *Proc Natl Acad Sci USA* 2006 Jun 6;103(23):8577-8582. [[FREE Full text](#)] [doi: [10.1073/pnas.0601602103](#)]
4. Speakman JR. Obesity: the integrated roles of environment and genetics. *J Nutr* 2004 Aug;134(8 Suppl):2090S-2105S. [[FREE Full text](#)] [Medline: [15284410](#)]
5. Duffey KJ, Popkin BM. Energy density, portion size, and eating occasions: contributions to increased energy intake in the United States, 1977-2006. *PLoS Med* 2011 Jun;8(6):1-8. [doi: [10.1371/journal.pmed.1001050](#)]
6. United States Department of Agriculture. ChooseMyPlate.gov URL: <http://www.choosemyplate.gov/> [accessed 2013-06-04] [[WebCite Cache ID 6H7wvK4ty](#)]
7. SurveyMonkey. URL: <http://www.surveymonkey.com/> [accessed 2013-06-03] [[WebCite Cache ID 6H7xKLTiK](#)]
8. Glickman D, Parker L, Sim L, Cook H, Miller E. Accelerating Progress in Obesity Prevention: Solving the Weight of the Nation. Washington, DC: The National Academies Press; 2012.
9. Contento I, Balch G, Bronner Y. 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.
10. 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](#)]
11. Luxton D, McCann A, Bush N, Mishkind MC, Reger GM. mHealth for mental health: integrating smartphone technology in behavioral healthcare. *Professional Psychology: Research and Practice* 2011 Dec;42(6):505-512. [doi: [10.1037/a0024485](#)]
12. GSMA, PwC. Report: Touching Lives through Mobile Health: Assessment of the Global Market Opportunity.: PricewaterhouseCoopers Private Limited; 2012. URL: <http://www.gsma.com/connectedliving/wp-content/uploads/2012/03/gsmawpwtouchinglivessthroughmobilehealthreport.pdf> [accessed 2013-06-10] [[WebCite Cache ID 6HHOBo6v5](#)]
13. Device ownership. Washington, DC: Pew Internet & American Life Project; 2013. URL: [http://www.pewinternet.org/Trend-Data-\(Adults\)/Device-Ownership.aspx](http://www.pewinternet.org/Trend-Data-(Adults)/Device-Ownership.aspx) [accessed 2013-05-08] [[WebCite Cache ID 6GTTm6Gdj](#)]
14. Institute of Medicine. Health and Behavior: The Interplay of Biological, Behavioral, and Societal Influences. Washington, DC: National Academy Press; 2001.
15. Bouton M. A learning theory perspective on lapse, relapse, and the maintenance of behavior change. *Health Psychol* 2000 Jan;19(1 Suppl):57-63. [Medline: [10709948](#)]
16. Godin G, Kok G. The theory of planned behavior: a review of its applications to health-related behaviors. *Am J Health Promot* 1996;11(2):87-98. [Medline: [10163601](#)]
17. Horsky J, McColgan K, Pang JE, Melnikas AJ, Linder JA, Schnipper JL, et al. Complementary methods of system usability evaluation: surveys and observations during software design and development cycles. *J Biomed Inform* 2010 Oct;43(5):782-790. [doi: [10.1016/j.jbi.2010.05.010](#)] [Medline: [20546936](#)]
18. Hu Y, Gansner ER, Kobourov S. Visualizing graphs and clusters as maps. *IEEE Comput Grap Appl* 2010 Nov;30(6):54-66. [doi: [10.1109/MCG.2010.101](#)]

19. Gansner E, Hu Y, Kobourov S. Putting recommendations on the map: visualizing clusters and relations. In: ACM Conf on Recommender Systems. 2009 Presented at: ACM Conf on Recommender Systems; October 22-25, 2009; New York City p. 345-378.
20. Graphviz - Graph Visualization Software. URL: <http://www.graphviz.org/> [accessed 2013-06-09] [WebCite Cache ID 6HFWdpwCE]
21. Gansner E, North S. An open graph visualization system and its applications to software engineering. Softw: Pract Exper 2000;30(11):1203-1233. [doi: [10.1002/1097-024X\(200009\)30:11<1203::AID-SPE338>3.3.CO;2-E](https://doi.org/10.1002/1097-024X(200009)30:11<1203::AID-SPE338>3.3.CO;2-E)]
22. Gansner E, Hu Y, North S. Visualizing Streaming Text Data with Dynamic Maps. 2012 Presented at: 20th International Symposium, GD 2012; Sept 19-21, 2012; Redmond, WA p. 1-10. [doi: [10.1007/978-3-642-36763-2_39](https://doi.org/10.1007/978-3-642-36763-2_39)]
23. Mashima D, Kobourov S, Hu Y. Visualizing dynamic data with maps. 2011 Presented at: Pacific Visualization Symposium (PacificVis), 2011 IEEE; March 1-4, 2011; Hong Kong. [doi: [10.1109/PACIFICVIS.2011.5742385](https://doi.org/10.1109/PACIFICVIS.2011.5742385)]
24. 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](https://doi.org/10.1016/j.jneb.2012.05.001)] [Medline: [23103255](https://pubmed.ncbi.nlm.nih.gov/23103255/)]
25. Woolford SJ, Clark SJ, Strecher VJ, Resnicow K. Tailored mobile phone text messages as an adjunct to obesity treatment for adolescents. J Telemed Telecare 2010;16(8):458-461 [FREE Full text] [doi: [10.1258/jtt.2010.100207](https://doi.org/10.1258/jtt.2010.100207)] [Medline: [20959393](https://pubmed.ncbi.nlm.nih.gov/20959393/)]
26. Drakulich A. Knowledge is power: global coordination tactics that incorporate online technologies and social media are reshaping disease prevention and response. BioPharm International 2012 Nov 01;25(11):6 [FREE Full text]
27. Mackert M, Kim E, Guadagno M, Donovan-Kicken E. Using Twitter for prenatal health promotion: encouraging a multivitamin habit among college-aged females. Stud Health Technol Inform 2012;182:93-103. [Medline: [23138084](https://pubmed.ncbi.nlm.nih.gov/23138084/)]

Abbreviations

BMI: body mass index

mHealth: mobile health

Edited by G Eysenbach; submitted 18.03.13; peer-reviewed by J Lemacks, J Gao; comments to author 12.04.13; revised version received 01.05.13; accepted 08.05.13; published 24.06.13.

Please cite as:

Hingle M, Yoon D, Fowler J, Kobourov S, Schneider ML, Falk D, Burd R

Collection and Visualization of Dietary Behavior and Reasons for Eating Using Twitter

J Med Internet Res 2013;15(6):e125

URL: <http://www.jmir.org/2013/6/e125/>

doi: [10.2196/jmir.2613](https://doi.org/10.2196/jmir.2613)

PMID: [23796439](https://pubmed.ncbi.nlm.nih.gov/23796439/)

©Melanie Hingle, Donella Yoon, Joseph Fowler, Stephen Kobourov, Michael Lee Schneider, Daniel Falk, Randy Burd. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 24.06.2013. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Using Online Health Communities to Deliver Patient-Centered Care to People With Chronic Conditions

Martijn van der Eijk^{1*}, MSc; Marjan J Faber^{2*}, PhD; Johanna WM Aarts^{3*}, MD, PhD; Jan AM Kremer^{2,3*}, MD, PhD; Marten Munneke^{1*}, PhD; Bastiaan R Bloem^{4*}, MD, PhD

¹Radboud University Nijmegen Medical Centre, Nijmegen Centre for Evidence Based Practice, Department of Neurology (935), Nijmegen, Netherlands

²Radboud University Nijmegen Medical Centre, Scientific Institute for Quality of Healthcare (IQ healthcare), Nijmegen, Netherlands

³Radboud University Nijmegen Medical Centre, Department of Obstetrics and Gynaecology, Nijmegen, Netherlands

⁴Radboud University Nijmegen Medical Centre, Donders Center for Brain, Cognition and Behaviour, Department of Neurology (935), Nijmegen, Netherlands

* all authors contributed equally

Corresponding Author:

Bastiaan R Bloem, MD, PhD

Radboud University Nijmegen Medical Centre

Donders Center for Brain, Cognition and Behaviour

Department of Neurology (935)

PO Box 9101

Nijmegen, 6500 HB

Netherlands

Phone: 31 243615202

Fax: 31 243541122

Email: b.bloem@neuro.umcn.nl

Abstract

Background: Our health care system faces major threats as the number of people with multiple chronic conditions rises dramatically.

Objective: To study the use of Online Health Communities (OHCs) as a tool to facilitate high-quality and affordable health care for future generations.

Methods: OHCs are Internet-based platforms that unite either a group of patients, a group of professionals, or a mixture of both. Members interact using modern communication technologies such as blogs, chats, forums, and wikis. We illustrate the use of OHCs for ParkinsonNet, a professional network for Parkinson disease whose participants—both patients and professionals—use various types of OHCs to deliver patient-centered care.

Results: We discuss several potential applications in clinical practice. First, due to rapid advances in medical knowledge, many health professionals lack sufficient expertise to address the complex health care needs of chronic patients. OHCs can be used to share experiences, exchange knowledge, and increase disease-specific expertise. Second, current health care delivery is fragmented, as many patients acquire relationships with multiple professionals and institutions. OHCs can bridge geographical distances and enable interdisciplinary collaboration across institutions and traditional echelons. Third, chronic patients lack adequate tools to self-manage their disease. OHCs can be used to actively engage and empower patients in their health care process and to tailor care to their individual needs. Personal health communities of individual patients offer unique opportunities to store all medical information in one central place, while allowing transparent communication across all members of each patient's health care team.

Conclusions: OHCs are a powerful tool to address some of the challenges chronic care faces today. OHCs help to facilitate communication among professionals and patients and support coordination of care across traditional echelons, which does not happen spontaneously in busy practice.

(*J Med Internet Res* 2013;15(6):e115) doi:[10.2196/jmir.2476](https://doi.org/10.2196/jmir.2476)

KEYWORDS

community networks; Internet; patient-centered care; long-term care; chronic disease; Parkinson disease

Introduction

Background

Our health care system faces major threats. Western societies age rapidly, and as a consequence, the prevalence of people with multiple chronic conditions rises dramatically [1]. Moreover, the number of patients with complex health care needs outpaces the number of professionals with sufficient knowledge and skills to adequately care for these people [2]. Finally, health care threatens to become unaffordable due to overtreatment and costly medical advancements [3,4]. To guarantee quality and affordable health care for future generations, innovations are needed [5]. In this paper, we discuss the use of Online Health Communities (OHCs) as a tool to address some of the above challenges. We illustrate the use of OHCs for ParkinsonNet, a professional network for Parkinson disease (PD), whose participants—both patients and professionals—use various types of OHCs to deliver patient-centered care [6,7].

ParkinsonNet consists of regional allied health networks for PD in the catchment areas of Dutch hospitals. Within each network, a selected number of expert therapists are trained according to evidence-based guidelines. Neurologists are stimulated to refer PD patients to these skilled professionals. Additionally, the concept has nationwide coverage in the Netherlands with 66 regional networks and 2400 physicians, nurses, physical therapists, occupational therapists, speech-language pathologists, and dieticians involved. ParkinsonNet was developed to improve the quality of PD care delivered by allied health professionals. The implementation of ParkinsonNet has shown a profound reduction in health care utilization and costs [6]. Participants increased their PD-specific knowledge, improved the adherence to guideline recommendations, and treated a higher volume of patients per year [7].

Online Health Communities

Platforms using social media technologies, such as Wikipedia, Facebook, LinkedIn, YouTube, and Twitter, have become extremely popular among millions of people worldwide. These tools have brought new possibilities for co-creation and communication between individuals with minimal time and cost restrictions [8]. It seems logical to apply elements of this

revolution to health care. As such, social network technologies provide an Internet-based platform for communication about health and disease, for sharing care experiences, and to increase medical knowledge [9,10]. By echoing Web 2.0 principles into health care, we could help patients become active participants in their own care and more engaged partners for health professionals [11]. Moreover, Internet-based contacts are a way to expand the possibilities for communication outside the few scheduled face-to-face hospital consultations [12].

One specific example are OHCs, which consist of an Internet-based platform that unites groups of individuals with a shared goal or similar interest regardless of their whereabouts [13]. Such a group could include patients with a particular condition (eg, patients with diabetes mellitus type II), a group of professionals with a shared medical interest (eg, diabetologists), or a mixture of both patients and professionals. Members might know each other from the “real” physical world, but the strength of OHCs is their potential to connect members who would otherwise never have met because of geographical distances. Within OHCs, members interact easily using modern communication technologies such as blogs, chats, forums, and wikis (Textbox 1, an illustrative example is PatientsLikeMe, an online platform for patients with life-changing conditions who share their experiences and medical data with other patients matched for clinical conditions and geographical characteristics. This platform provides generic solutions to acquire medical information and peer support for different patient groups [14]. PatientsLikeMe is currently being used by PD patients who quantify and self-report their disease symptoms on a regular basis. These data provide health professionals with new insight into variations in symptom severity and understanding about the disease progression in PD [15].

Platforms using OHC principles are utilized by patients from various ages. Moreover, the Health and Welfare Information Portal (ZWIP) combines an electronic health record with a communication tool aiming to improve care for frail, older people. ZWIP potentially enhances patient involvement, coordination of care and collaboration among professionals [16]. Furthermore, OHCs are utilized in Dutch fertility care. Young couples gain access to their medical records containing general and personal information and communication tools with peers and their local health care team [17].

Textbox 1. Social media applications within an online health community.

Blog: a blog is a series of messages published in reverse chronological order written by one of the community members (eg, scientific developments or personal care experiences).

Chat: a chat is a real-time conversation with other community members.

Forum: a forum is used for asynchronous communication with other community members (eg, patients can put questions to professionals or peers).

Library: where documents are shared with all community members (eg, information pamphlets, newsletters, scientific articles, and guidelines).

Wiki: an application within a community where all members are allowed to adapt a certain document (eg, an address list or information pamphlet).

Open and Closed Communities

OHCs can be classified into open and closed communities based on the accessibility of the community content. The content of open OHCs can be accessed by anyone, all members are allowed to contribute to its content, and all information that is generated

is openly accessible to anyone. Within a closed OHC, the content is visible to community members only. Members are allowed to make an active contribution after a community manager, that is, an individual who leads the community, has granted them access. The platform described in this paper is

utilized by several patient groups in the field of PD, dermatology, stroke, MS, rheumatoid arthritis, fertility and cancer care [18].

Open Parkinson Communities

The Parkinson community is an open community for all people interested in PD. Members are PD patients, caregivers, and health professionals. Table 1 shows the different social media applications used within this community. Patients use the community forum for online peer support and discussions with health professionals. Often, fellow patients provide useful answers, which may alleviate the pressure on health professionals. In an open community for breast cancer patients, incorrect answers were rapidly corrected by other members [19]. A striking feature is the wiki, which, with the help of several community members, is developing into a national encyclopedia for PD.

Closed Parkinson Communities

The ParkinsonNet community is used to facilitate communication and collaboration between health professionals involved in the treatment of PD patients and is accessible to ParkinsonNet professionals only (Figure 1). After verifying the ParkinsonNet membership, the community manager grants access. The community forum has been divided into separate discussions for physical, speech, and occupational therapists, and for interdisciplinary consultation. Other applications include the community blog, where members are informed about ongoing education and guidelines; the wiki, containing an up-to-date address list of all ParkinsonNet professionals; and the community library, used for sharing presentations of multidisciplinary team meetings.

Another example is the closed community of an outpatient Parkinson clinic, which is accessible only to patients visiting the clinic and to health professionals who work there (Table 1). A distinctive feature of this community is the combination of online patient-provider and peer-to-peer communication integrated into one and the same community with both patients and health professionals from the same clinic involved. In our Parkinson center, we run such an OHC as a service to both our patients and the members of our multidisciplinary health care team. Access is restricted and controlled—to become a member, patients must first send a formal membership request. After verifying the patient identification number, the community manager grants access.

The community blog contains information about the treatment facilities that are available at our center. Within the community forum, patients are provided with facilities for communication with fellow patients and the health care team. Future patients benefit from previous discussions, which remain visible unless patients wish this to be removed. This OHC does not offer

individually tailored information because the exchange of information is not private and can be seen by all members. This community type has proved to fill the gap between patients' needs and the support our clinic can offer [18].

Some of our patients are both members of the closed outpatient Parkinson clinic and the open Parkinson community. These patients have the opportunity to ask questions on both forums. Items in the outpatient clinic are more likely to involve questions to our medical team and facts, for example, treatment options in our clinic, whereas the open national forum is more likely to contain peer contact, care experiences, and opinions.

Personal Health Communities for Parkinson Disease

A Personal Health Community (PHC) is a private community governed by individual PD patients. Apart from the patient, participants include one or more (ideally all) health professionals involved in the care process, and the immediate caregiver. The patient is the owner of the community and decides who is granted access to the community. The immediate caregiver can act as community manager if the patient is unwilling or unable to do so. Once gathered, the patient and the health care team exchange information and communicate about individual health problems. Like an electronic patient record, PHCs offer the opportunity to store all medical information in one central place, while allowing transparent communication across members of the health care team. Hereby, the patient is in the lead as an active and equal partner who contributes to his own health.

PHCs differ from other OHCs in two ways. First, PHC functionalities are customized and used in a different way. The blogging feature is used as a diary to inform other members about, eg, side effects of anti-Parkinson medication, the forum for online consultation of health professionals, the library to store medical information, and the wiki as a specific medical document, like a medication scheme or treatment overview. A second difference is “two-way authentication”, which adds an extra layer of security to the PHC. Patients have to enter their username, password, and a security code sent to their mobile phone.

Active Users

The following definition of active user is applied on our platform: “The total number of users who performed at least one activity for a given day. Activities include: blog posts, blog comments, forum posts, forum replies, library uploads, library downloads, new wiki pages, wiki revisions, wiki comments, joining a group, subscribing to content or rating a post” [20]. Table 1 shows that over a 12-month period, 54% of the Parkinson and ParkinsonNet community members generated new content or posted a comment. Other participants may have visited the community, albeit without an active contribution.

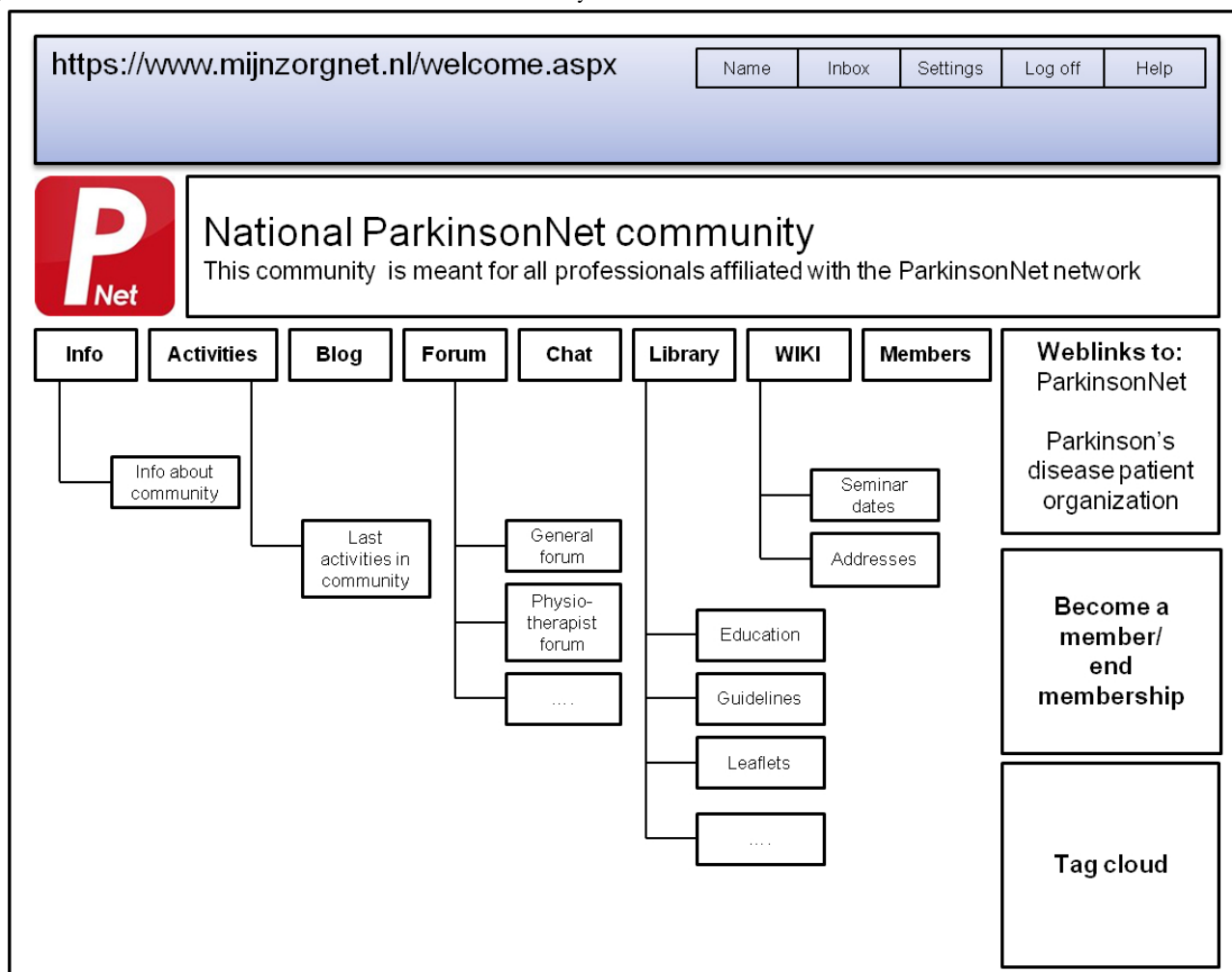
Figure 1. Schematic overview of the closed ParkinsonNet community.

Table 1. Social media applications and members of the Parkinson communities.

	Parkinson community	ParkinsonNet community	Outpatient Parkinson clinic	Personal health community
Community type	Open	Closed	Closed	Closed
Info	A national community for people interested in PD for peer-to-peer contact and patient-professional interaction	Transmural community for ParkinsonNet expert therapists for online professional interaction	Intramural community of an outpatient Parkinson clinic. Supplementary service to the existing face-to-face care relationship	A private electronic patient record of individual patients to allow personalized medicine
Members	PD patients, caregivers and health professionals 1190 (Dec. 31, 2012) 1224 (Jan. 31, 2013)	ParkinsonNet health professionals	The multidisciplinary health care team and PD patients visiting the Parkinson clinic	One PD patient, his caregiver and health professionals involved (eg physical therapist, neurologists, PCP) 147 patients 78 professionals (Jan. 31, 2013)
Active users	2012: 646 (54%) Recent month ^a : 83 (7%)	2012: 737 (54%) Recent month: 182 (13%)	2012: 56 (88%)	Data not available due to technical and privacy issues
Page views	2012: 241,093 Recent month: 24,671	2012: 76,452 Recent month: 7568	2012: 8376 Recent month: 1098	Data not available due to technical and privacy issues
Blog	Information from physicians and therapists on new trials, etiology, diagnosis and PD treatment options	Information from ParkinsonNet professionals about conferences, team meetings, new PD guidelines	Information on medication, non-motor symptoms, research in our clinic, announcement of maternity leave local PD nurse specialist	PD patients' diary about eg on/off fluctuations, wearing off of, side effects and daily experiences
Forum	Discussions between community members about medication, symptoms and peer support	Discussions about allied health therapy, food and diet and medication	Peer-to-peer contact between PD patients in a familiar setting and questions to the local health care team	Consultation between a patient and his professional care team about, eg, side effects
Library	Documents on allied health disciplines, driving abilities, medication and side effects, PD guidelines and scientific papers	Centralized up-to-date information: newsletters, presentations, scientific papers, PD guidelines, clinimetrics, courses	Information on treatment facilities provided by our clinic and regional peer contact	Individually tailored information, eg, physical exercises, medication schemes or a diary
Chat	No chat available	Real time conversation at assigned times between ParkinsonNet professionals	No chat available	No chat available
Wiki	A national encyclopedia for PD on diagnosis, symptoms, medication, on/off fluctuations, multidisciplinary collaboration, etiology and disease progression	Address list of all ParkinsonNet professionals, a ParkinsonNet calendar with, eg, regional ParkinsonNet meetings, conferences and symposia	No wiki available	Eg, medication or treatment overview

^aRecent month means January 2013.

Methods

Implementation Strategy

Implementation of OHCs in clinical practice takes a collective effort of all health professionals involved. However, the community manager plays a vital role during the implementation and maintenance of all OHCs. The community manager of an outpatient Parkinson clinic is usually a local PD nurse specialist. PD nurse specialists are key practitioners when it comes to the coordination of care, patient education, and emotional support [21]. However, the community manager appointed in our clinic is a marketing and communication expert. The community manager distributes posters, information pamphlets, and "business cards" to patients and health professionals. The information pamphlet, which is available in every doctor's

office, contains information about the aim of the community, login procedures, and social media applications within the OHC. Other tasks of the community manager include management and maintenance of the community members' database, generation of content, motivation of health professionals and patients to participate, and monitoring of the expert forum. Recently, the first 10 PD patients received training on navigation through the online outpatient clinic.

The community manager of the ParkinsonNet community is a marketing and communication expert as well. She visited all 66 regional ParkinsonNet networks to educate health professionals about OHCs. ParkinsonNet professionals are urged to enroll in the ParkinsonNet community as part of their membership. Currently, the ParkinsonNet community is the main source of information for the professionals in the network.

Some information about new guidelines and state-of-art courses can be found only within this community.

In 2011, we introduced the Personal Health Community (PHC) in four Parkinson clinics in the Netherlands. During regular home visits, patients learn to navigate through and utilize the PHC by a local PD nurse specialist. ParkinsonNet organizes workshops to engage patients and health professionals in the pilot regions. Additionally, an information pamphlet, poster and a video to promote the PHC were introduced. Roughly, implementation of PHCs includes three phases: a pilot phase concerning patients and professionals from the Parkinson clinic only, a second phase in which primary care providers are invited, and a third phase in which new clinics are included. Our first experiences show that PHCs facilitate emotional support, health care accessibility, and improve relationships between professionals and patients.

Results

The Added Value of OHCs in Chronic Care

Based on our first experiences in PD care and the international literature, OHCs have four major advantages to improve the quality of chronic care. These include facilitation for the exchange of medical experience and knowledge, enhancing interdisciplinary collaboration across institutions and traditional echelons, providing a platform to support self-management, and the ability to improve patient-centered care.

OHCs Facilitate the Exchange of Medical Experience and Knowledge

Due to rapid advances in medical knowledge, many health professionals lack specific expertise and experience to address the complex health care needs of chronic patients [1,2]. Therefore, health care is increasingly organized within specialized networks, like ParkinsonNet [22,23]. Professional networks enhance information exchange, facilitate communication among participants and foster the adoption of new knowledge, such as revised guidelines [24,25]. Traditionally, these network processes occur largely offline during physical encounters, such as medical conferences. However, with the advent of modern communication technologies, professional networks can now be supported online. Within OHCs, professionals connect and communicate more easily, regardless of their working place within the network, and regardless of time. Moreover, OHCs can be used to develop disease-specific expertise among all community members, patients, and professionals, interested in a particular chronic condition [26].

OHCs Enhance Interdisciplinary Collaboration Across Institutions and Traditional Echelons

Health care delivery can become fragmented for chronic patients when they acquire relationships with multiple professionals and institutions. Increasingly, chronic care has evolved from individual consultations into multidisciplinary teamwork with care given by various physicians and therapists, who often work in different departments or organizations [27]. To manage complex patients with multiple co-morbidities, health

professionals must collaborate to make coordinated decisions and share responsibilities in health outcomes [28]. Yet, the collaboration and coordination of care should be improved considerably [29]. Given their synchronous and asynchronous communication capacity, OHCs offer a platform for supporting medical decision-making and interdisciplinary collaboration across professionals caring for complex patients [26,30,31]. OHCs enable communication between community members who are not able to have face-to-face interaction at any point in time. Moreover, OHCs bridge geographical distances and enable interaction across institutions and traditional echelons. An example is the Canadian Virtual Hospice, with information and support on palliative and end-of-life care [32]. Patients, close family members, and caregivers interact in several peer-to-peer discussion forums or private messages with a team of palliative care experts. Normally, these interactions would not have been possible due to physical limitations and geographical distances.

OHCs Provide a Platform to Support Self-Management

Typically, patients have a passive role and lack the tools to self-manage their condition [33]. However, modern patients search the Internet for medical information, wish to have open communication channels with their physicians, and prefer to participate in making treatment decisions [34]. Self-management refers to the efforts to enhance patient participation and assisting patients to gain control over their lives [35]. The concept is associated with improved communication between patients and clinicians, and it enhances quality of life [36]. Supporting patients with chronic diseases like type 2 diabetes, arthritis, and asthma to self-manage their condition helps to improve the quality and safety of care and reduces costly and inappropriate use of health care resources [37,38]. Increasingly, the Internet is used to support self-management and actively engage patients in treatment decisions [39]. Chronic patients using online communication tools become more knowledgeable, feel better socially supported and empowered, and have improved behavioral and clinical outcomes compared to nonusers [40,41]. Examples that include OHC principles are patient participation in online peer support groups and access to PHCs [15,42]. PHCs allow patients to have access to medical records, control their own online information, and enable individualized health communication [43].

OHCs Have the Ability to Improve Patient-Centered Care

Patient-centeredness is defined as “providing care that is responsive to individual patient preferences, needs, and values, and ensuring that patient values guide all clinical decisions” [44]. Contrary to some perceptions, patient-centeredness is not just about being nice to patients, but engaging them to become active participants in their care [45]. The concept is known for its advantages in terms of reduced health care utilization and improved efficiency, patient-doctor communication, treatment compliance, and health outcomes [46-48]. OHCs enhance patient-centered care by improved access to personalized information, emotional support, and patient participation [15-42,49]. PHCs are essentially patient-centered, while they engage patients in their care process and tailor care to their

individual needs. Professionals have the opportunity to benefit from patient peer-to-peer conversations that take place in OHCs by knowing that they have more effectively addressed their patients' needs [50]. Blog and forum items often involve aspects of patient-centered care, such as information and emotional support needs, patients' willingness to participate in treatment decisions, or an experienced lack of continuity of care.

Discussion

Barriers for Implementation in Clinical Practice

Why do only a few innovations become part of routine practice and why do most fail to survive beyond the pilot phase? To answer this question, full understanding of the clinical and technological barriers and incentives for achieving behavior change in practice is needed [51,52]. The following barriers to Internet technologies may be at play.

First, the implementation of Internet innovations can radically affect health care delivery and professionals' daily work processes, requiring considerable time and willingness to learn [53]. Doctors may be hesitant to adopt technologies that imply an interruption of their traditional practice patterns. The requirement of additional time is a prominent barrier to physician technology acceptance [54].

Second, the implementation of OHCs into clinical practice demands a paradigm shift in control and power, out of the hands of those who deliver care, into the hands of those who receive it [13]. Professionals should no longer regard patients as passive objects, but rather as equal, participatory partners who contribute to their own health. Thus, OHCs require a change in the mindset of both professionals and patients. Not surprisingly, in the age of Facebook, young clinicians may struggle to maintain professional distance on one hand and have close, meaningful relationships with their patients on the other [55].

Third, besides behavioral change, safety and financial issues have to be solved [56]. To ensure a safe and secure environment, the Dutch government authorizes PD patients to apply their personal verification code, normally used to complete and

submit a tax return form to the tax authority, while using our OHC platform. Health professionals are allowed to access closed communities only via their unique, electronic identity. Additionally, for OHCs to become integrated into everyday use, new and viable business models are needed. To utilize OHCs in daily practice is time consuming; however, they may also substitute normal ways of care delivery. Generally, health care is reimbursed by face-to-face interactions and offline medical services. Bearing this in mind, we would like to introduce the term "blended health", analogous to blended learning, which combines face-to-face contact with the possibilities of online tools. The intended result is a health care system not driven by technology, but using technology as a tool to facilitate patient-centered, collaborative care.

More and more, innovations in health care are based on Internet technologies and the willingness of PD patients to participate in such interventions is growing [57]. Generally, health related Internet use is associated with age and level of education [58]. The European Union investigated the level of Internet access within the 27 member states. Household Internet access ranged from 45% in Bulgaria to 94% in the Netherlands [59]. Therefore, Internet access is assumed to be a minor limitation in the adoption of OHCs in the Netherlands.

Conclusions

OHCs are a powerful tool to address some of the challenges chronic care faces today. A challenge now is to perform an in-depth evaluation of our platform, which is simultaneously being designed, developed, and deployed [60]. Further evaluation should address user needs, risks, benefits, and cost implications before OHCs can be fully adopted in daily practice [61,62]. We expect that innovations like OHCs can help to facilitate high-quality and affordable health care for future generations. Chronic care demands an integrated approach tailored to the needs of individual patients to optimize outcomes. In the absence of a formal team structure, OHCs help to facilitate communication among professionals and patients and support coordination of care across traditional echelons, which does not happen spontaneously in a busy practice.

Acknowledgments

An unrestricted grant towards this research was provided by the Dutch Ministry of Health, Welfare and Sport and is hereby acknowledged. There is no other financial or material support involved in this paper.

Conflicts of Interest

Professor Jan Kremer (Professor in reproductive medicine) and Professor BR Bloem (Professor in Neurology, with movement disorders as special interest) are the initiators of the online community platform MijnZorgnet [18] described in this paper.

References

1. Anderson G. Chronic Care: Making the Case for Ongoing Care. Princeton: Robert Wood Johnson Foundation; 2010. URL: <http://www.rwjf.org/content/rwjf/en/research-publications/find-rwjf-research/2010/01/chronic-care.html> [accessed 2012-12-05] [WebCite Cache ID 6CgH4LRi1]
2. Institute of Medicine Board on Healthcare. Retooling for an Aging America: Building the Health Care Workforce. 2008. URL: <http://www.iom.edu/Reports/2008/Retooling-for-an-Aging-America-Building-the-Health-Care-Workforce.aspx> [accessed 2012-12-05] [WebCite Cache ID 6CgH7dC9x]
3. Brownlee S. Overtreated: why too much medicine is making us sicker and poorer. New York: Bloomsbury; 2007.

4. Kowdley G, Ashbaker D. Health care costs in America—technology as a major driver. *J Surg Educ* 2011;68(3):231-238. [Medline: [21618817](#)]
5. Christensen C. The innovator's prescription: a disruptive solution for health care. New York: McGraw-Hill; 2009.
6. Munneke M, Nijkrake MJ, Keus SH, Kwakkel G, Berendse HW, Roos RA, ParkinsonNet Trial Study Group. Efficacy of community-based physiotherapy networks for patients with Parkinson's disease: a cluster-randomised trial. *Lancet Neurol* 2010 Jan;9(1):46-54. [doi: [10.1016/S1474-4422\(09\)70327-8](#)] [Medline: [19959398](#)]
7. Nijkrake MJ, Keus SH, Overeem S, Oostendorp RA, Vlieland TP, Mulleners W, et al. The ParkinsonNet concept: development, implementation and initial experience. *Mov Disord* 2010 May 15;25(7):823-829. [doi: [10.1002/mds.22813](#)] [Medline: [20461798](#)]
8. Shirky C. Here comes everybody: the power of organizing without organizations. New York: Penguin Press; 2008.
9. Nambisan P. Information seeking and social support in online health communities: impact on patients' perceived empathy. *J Am Med Inform Assoc* 2011 May 1;18(3):298-304 [FREE Full text] [doi: [10.1136/amiajnl-2010-000058](#)] [Medline: [21486888](#)]
10. Powell JA, Darvell M, Gray JA. The doctor, the patient and the world-wide web: how the internet is changing healthcare. *J R Soc Med* 2003 Feb;96(2):74-76 [FREE Full text] [Medline: [12562977](#)]
11. Swan M. Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. *Int J Environ Res Public Health* 2009 Feb;6(2):492-525 [FREE Full text] [doi: [10.3390/ijerph6020492](#)] [Medline: [19440396](#)]
12. Bjoernes CD, Laursen BS, Delmar C, Cummings E, Nøhr C. A dialogue-based Web application enhances personalized access to healthcare professionals--an intervention study. *BMC Med Inform Decis Mak* 2012;12:96 [FREE Full text] [doi: [10.1186/1472-6947-12-96](#)] [Medline: [22947231](#)]
13. Demiris G. The diffusion of virtual communities in health care: concepts and challenges. *Patient Educ Couns* 2006 Aug;62(2):178-188. [doi: [10.1016/j.pec.2005.10.003](#)] [Medline: [16406472](#)]
14. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. *J Med Internet Res* 2010;12(2):e19 [FREE Full text] [doi: [10.2196/jmir.1549](#)] [Medline: [20542858](#)]
15. Little M, Wicks P, Vaughan T, Pentland A. Quantifying short-term dynamics of Parkinson's disease using self-reported symptom data from an Internet social network. *J Med Internet Res* 2013;15(1):e20 [FREE Full text] [doi: [10.2196/jmir.2112](#)] [Medline: [23343503](#)]
16. Robben SH, Perry M, Huisjes M, van Nieuwenhuijzen L, Schers HJ, van Weel C, et al. Implementation of an innovative web-based conference table for community-dwelling frail older people, their informal caregivers and professionals: a process evaluation. *BMC Health Serv Res* 2012;12:251 [FREE Full text] [doi: [10.1186/1472-6963-12-251](#)] [Medline: [22894654](#)]
17. Tuil WS, ten Hoopen AJ, Braat DD, de Vries Robbé PF, Kremer JA. Patient-centred care: using online personal medical records in IVF practice. *Hum Reprod* 2006 Nov;21(11):2955-2959 [FREE Full text] [doi: [10.1093/humrep/del214](#)] [Medline: [16982658](#)]
18. MijnZorgnet. URL: <https://www.mijnzorgnet.nl/welkom/> [accessed 2013-02-26] [WebCite Cache ID 6Eihm2acB]
19. Esquivel A, Meric-Bernstam F, Bernstam EV. Accuracy and self correction of information received from an internet breast cancer list: content analysis. *BMJ* 2006 Apr 22;332(7547):939-942 [FREE Full text] [doi: [10.1136/bmj.38753.524201.7C](#)] [Medline: [16513686](#)]
20. Telligent software. URL: <http://telligent.com/> [accessed 2013-02-26] [WebCite Cache ID 6Eiht11Kh]
21. Axelrod L, Gage H, Kaye J, Bryan K, Trend P, Wade D. Workloads of Parkinson's specialist nurses: implications for implementing national service guidelines in England. *J Clin Nurs* 2010 Dec;19(23-24):3575-3580. [doi: [10.1111/j.1365-2702.2010.03279.x](#)] [Medline: [21083781](#)]
22. Simmons D, English P, Robins P, Craig A, Addicott R. Should diabetes be commissioned through multidisciplinary networks, rather than Practice Based Commissioning? *Prim Care Diabetes* 2011 Apr;5(1):39-44. [doi: [10.1016/j.pcd.2010.09.002](#)] [Medline: [20956096](#)]
23. Verhoef J, Oosterveld FG, Hoekman R, Munneke M, Boonman DC, Bakker M, et al. A system of networks and continuing education for physical therapists in rheumatology: a feasibility study. *Int J Integr Care* 2004;4:e19 [FREE Full text] [Medline: [16773150](#)]
24. Sales AE, Estabrooks CA, Valente TW. The impact of social networks on knowledge transfer in long-term care facilities: Protocol for a study. *Implement Sci* 2010;5:49 [FREE Full text] [doi: [10.1186/1748-5908-5-49](#)] [Medline: [20573254](#)]
25. Wensing M, van Lieshout J, Koetsenruter J, Reeves D. Information exchange networks for chronic illness care in primary care practices: an observational study. *Implement Sci* 2010;5:3 [FREE Full text] [doi: [10.1186/1748-5908-5-3](#)] [Medline: [20205758](#)]
26. Shachak A, Jadad AR. Electronic health records in the age of social networks and global telecommunications. *JAMA* 2010 Feb 3;303(5):452-453. [doi: [10.1001/jama.2010.63](#)] [Medline: [20124543](#)]
27. Ouwens M, Wollersheim H, Hermens R, Hulscher M, Grol R. Integrated care programmes for chronically ill patients: a review of systematic reviews. *Int J Qual Health Care* 2005 Apr;17(2):141-146 [FREE Full text] [doi: [10.1093/intqhc/mzi016](#)] [Medline: [15665066](#)]

28. Lemieux-Charles L. Understanding the conditions that lead to effective health services delivery networks. *Healthc Pap* 2006;7(2):40-5; discussion 68. [Medline: [17167318](#)]
29. Weinberg DB, Cooney-Miner D, Perloff JN, Babington L, Avgar AC. Building collaborative capacity: promoting interdisciplinary teamwork in the absence of formal teams. *Med Care* 2011 Aug;49(8):716-723. [doi: [10.1097/MLR.0b013e318215da3f](#)] [Medline: [21478768](#)]
30. Naik AD, Singh H. Electronic health records to coordinate decision making for complex patients: what can we learn from wiki? *Med Decis Making* 2010;30(6):722-731. [doi: [10.1177/0272989X10385846](#)] [Medline: [21183759](#)]
31. Wiecha J, Pollard T. The interdisciplinary eHealth team: chronic care for the future. *J Med Internet Res* 2004 Sep 3;6(3):e22 [FREE Full text] [doi: [10.2196/jmir.6.3.e22](#)] [Medline: [15471748](#)]
32. Chochinov HM, Stern A. The Canadian Virtual Hospice <[www.virtualhospice.ca](#)>. *J Palliat Care* 2004;20(1):5-6. [Medline: [15132069](#)]
33. Frosch DL, Elwyn G. I believe, therefore I do. *J Gen Intern Med* 2011 Jan;26(1):2-4 [FREE Full text] [doi: [10.1007/s11606-010-1560-1](#)] [Medline: [21061083](#)]
34. Kremer JA, Van Der Eijk M, Aarts JW, Bloem BR. The individual formerly known as patient, TIFKAP. *Minerva Med* 2011 Dec;102(6):505. [Medline: [22193382](#)]
35. Kralik D, Koch T, Price K, Howard N. Chronic illness self-management: taking action to create order. *J Clin Nurs* 2004 Feb;13(2):259-267. [Medline: [14723679](#)]
36. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. *Ann Behav Med* 2003 Aug;26(1):1-7. [Medline: [12867348](#)]
37. Newman S, Steed L, Mulligan K. Self-management interventions for chronic illness. *Lancet* 2004;364(9444):1523-1537. [doi: [10.1016/S0140-6736\(04\)17277-2](#)] [Medline: [15500899](#)]
38. Schillinger D. Supporting self management—a necessity in diabetes healthcare. *Patient Educ Couns* 2011 Nov;85(2):131-132. [doi: [10.1016/j.pec.2011.08.007](#)] [Medline: [21889866](#)]
39. Lorig KR, Ritter PL, Dost A, Plant K, Laurent DD, McNeil I. The Expert Patients Programme online, a 1-year study of an Internet-based self-management programme for people with long-term conditions. *Chronic Illn* 2008 Dec;4(4):247-256. [doi: [10.1177/1742395308098886](#)] [Medline: [19091933](#)]
40. Gibbons MC, Wilson RF, Samal L, Lehman CU, Dickersin K, Lehmann HP, et al. Impact of consumer health informatics applications. *Evid Rep Technol Assess (Full Rep)* 2009 Oct(188):1-546. [Medline: [20629477](#)]
41. 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](#)]
42. Krist AH, Woolf SH. A vision for patient-centered health information systems. *JAMA* 2011 Jan 19;305(3):300-301. [doi: [10.1001/jama.2010.2011](#)] [Medline: [21245186](#)]
43. Archer N, Fevrier-Thomas U, Lokker C, McKibbin KA, Straus SE. Personal health records: a scoping review. *J Am Med Inform Assoc* 2011;18(4):515-522 [FREE Full text] [doi: [10.1136/amiainl-2011-000105](#)] [Medline: [21672914](#)]
44. Corrigan JM, Donaldson MS, Kohn LT, Maguire S, Pike KC. A New Health System for the 21st Century. Crossing the Quality Chasm URL: [http://www.iom.edu/Reports/2001/Crossing-the-Quality-Chasm-A-New-Health-System-for-the-21st-Century.aspx](#) [accessed 2012-12-05] [WebCite Cache ID 6CgHCleDV]
45. DerGurahian J. Focusing on the patient. Planetree guide touts patient-centered care model. *Mod Healthc* 2008 Oct 27;38(43):7. [Medline: [18998450](#)]
46. Michie S, Miles J, Weinman J. Patient-centredness in chronic illness: what is it and does it matter? *Patient Educ Couns* 2003 Nov;51(3):197-206. [Medline: [14630376](#)]
47. Griffin SJ, Kinmonth AL, Veltman MW, Gillard S, Grant J, Stewart M. Effect on health-related outcomes of interventions to alter the interaction between patients and practitioners: a systematic review of trials. *Ann Fam Med* 2004;2(6):595-608 [FREE Full text] [doi: [10.1370/afm.142](#)] [Medline: [15576546](#)]
48. Rao JK, Anderson LA, Inui TS, Frankel RM. Communication interventions make a difference in conversations between physicians and patients: a systematic review of the evidence. *Med Care* 2007 Apr;45(4):340-349. [doi: [10.1097/01.mlr.0000254516.04961.d5](#)] [Medline: [17496718](#)]
49. Wasson JH, Forsberg HH, Lindblad S, Mazowita G, McQuillen K, Nelson EC. The medium is the (health) measure: patient engagement using personal technologies. *J Ambul Care Manage* 2012;35(2):109-117. [doi: [10.1097/JAC.0b013e31824a235e](#)] [Medline: [22415284](#)]
50. Sevin C, Moore G, Shepherd J, Jacobs T, Hupke C. Transforming care teams to provide the best possible patient-centered, collaborative care. *J Ambul Care Manage* 2009;32(1):24-31. [doi: [10.1097/01.JAC.0000343121.07844.e0](#)] [Medline: [19104291](#)]
51. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA* 1999 Oct 20;282(15):1458-1465. [Medline: [10535437](#)]
52. 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](#)]

53. Koivunen M, Hätönen H, Välimäki M. Barriers and facilitators influencing the implementation of an interactive Internet-portal application for patient education in psychiatric hospitals. *Patient Educ Couns* 2008 Mar;70(3):412-419. [doi: [10.1016/j.pec.2007.11.002](https://doi.org/10.1016/j.pec.2007.11.002)] [Medline: [18079085](#)]
54. Yarbrough AK, Smith TB. Technology acceptance among physicians: a new take on TAM. *Med Care Res Rev* 2007 Dec;64(6):650-672. [doi: [10.1177/1077558707305942](https://doi.org/10.1177/1077558707305942)] [Medline: [17717378](#)]
55. Jain SH. Practicing medicine in the age of Facebook. *N Engl J Med* 2009 Aug 13;361(7):649-651. [doi: [10.1056/NEJMp0901277](https://doi.org/10.1056/NEJMp0901277)] [Medline: [19675328](#)]
56. Grol R, Wensing M. What drives change? Barriers to and incentives for achieving evidence-based practice. *Med J Aust* 2004 Mar 15;180(6 Suppl):S57-S60. [Medline: [15012583](#)]
57. Shprecher D, Noyes K, Biglan K, Wang D, Dorsey ER, Kurlan R, et al. Willingness of Parkinson's disease patients to participate in research using internet-based technology. *Telemed J E Health* 2012 Nov;18(9):684-687. [doi: [10.1089/tmj.2011.0276](https://doi.org/10.1089/tmj.2011.0276)] [Medline: [22954069](#)]
58. van Uden-Kraan CF, Drossaert CH, Taal E, Smit WM, Moens HJ, Siesling S, et al. Health-related Internet use by patients with somatic diseases: frequency of use and characteristics of users. *Inform Health Soc Care* 2009 Jan;34(1):18-29. [doi: [10.1080/17538150902773272](https://doi.org/10.1080/17538150902773272)] [Medline: [19306196](#)]
59. Eurostat European Commission. Internet use in households and by individuals in 2011. 2011. URL: <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/> [accessed 2013-02-26] [WebCite Cache ID 6Eii3leoS]
60. Catwell L, Sheikh A. Evaluating eHealth interventions: the need for continuous systemic evaluation. *PLoS Med* 2009 Aug;6(8):e1000126 [FREE Full text] [doi: [10.1371/journal.pmed.1000126](https://doi.org/10.1371/journal.pmed.1000126)] [Medline: [19688038](#)]
61. Chiu TM, Eysenbach G. Stages of use: consideration, initiation, utilization, and outcomes of an internet-mediated intervention. *BMC Med Inform Decis Mak* 2010;10:73 [FREE Full text] [doi: [10.1186/1472-6947-10-73](https://doi.org/10.1186/1472-6947-10-73)] [Medline: [21092275](#)]
62. 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](https://doi.org/10.1136/bmj.328.7449.1150)] [Medline: [15142895](#)]

Abbreviations

OHC: online health community

PD: Parkinson disease

PHC: personal health record

Edited by G Eysenbach; submitted 11.12.12; peer-reviewed by B Meskó, C Young; comments to author 25.01.13; revised version received 05.03.13; accepted 08.03.13; published 25.06.13.

Please cite as:

van der Eijk M, Faber MJ, Aarts JWM, Kremer JAM, Munneke M, Bloem BR

Using Online Health Communities to Deliver Patient-Centered Care to People With Chronic Conditions

J Med Internet Res 2013;15(6):e115

URL: <http://www.jmir.org/2013/6/e115/>

doi: [10.2196/jmir.2476](https://doi.org/10.2196/jmir.2476)

PMID: [23803284](https://pubmed.ncbi.nlm.nih.gov/23803284/)

©Martijn van der Eijk, Marjan J Faber, Johanna WM Aarts, Jan AM Kremer, Marten Munneke, Bastiaan R Bloem. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 25.06.2013. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Video Consultation Use by Australian General Practitioners: Video Vignette Study

Moyez Jiwa¹, MA, MD, FRCP, MRCP, FRACGP, MMedSci; Xingqiong Meng², PhD, BMed, MAppE

¹Curtin University, Curtin Health Innovation Research Institute, Curtin University, Perth, Australia

²School of Public Health, Curtin University, Perth, Australia

Corresponding Author:

Moyez Jiwa, MA, MD, FRCP, MRCP, FRACGP, MMedSci

Curtin Health Innovation Research Institute

Curtin University

GPO Box U1987

Perth, 6845

Australia

Phone: 61 892661768 ext 1768

Fax: 61 8 92662608

Email: m.jiwa@curtin.edu.au

Abstract

Background: There is unequal access to health care in Australia, particularly for the one-third of the population living in remote and rural areas. Video consultations delivered via the Internet present an opportunity to provide medical services to those who are underserved, but this is not currently routine practice in Australia. There are advantages and shortcomings to using video consultations for diagnosis, and general practitioners (GPs) have varying opinions regarding their efficacy.

Objective: The aim of this Internet-based study was to explore the attitudes of Australian GPs toward video consultation by using a range of patient scenarios presenting different clinical problems.

Methods: Overall, 102 GPs were invited to view 6 video vignettes featuring patients presenting with acute and chronic illnesses. For each vignette, they were asked to offer a differential diagnosis and to complete a survey based on the theory of planned behavior documenting their views on the value of a video consultation.

Results: A total of 47 GPs participated in the study. The participants were younger than Australian GPs based on national data, and more likely to be working in a larger practice. Most participants (72%-100%) agreed on the differential diagnosis in all video scenarios. Approximately one-third of the study participants were positive about video consultations, one-third were ambivalent, and one-third were against them. In all, 91% opposed conducting a video consultation for the patient with symptoms of an acute myocardial infarction. Inability to examine the patient was most frequently cited as the reason for not conducting a video consultation. Australian GPs who were favorably inclined toward video consultations were more likely to work in larger practices, and were more established GPs, especially in rural areas. The survey results also suggest that the deployment of video technology will need to focus on follow-up consultations.

Conclusions: Patients with minor self-limiting illnesses and those with medical emergencies are unlikely to be offered access to a GP by video. The process of establishing video consultations as routine practice will need to be endorsed by senior members of the profession and funding organizations. Video consultation techniques will also need to be taught in medical schools.

(*J Med Internet Res* 2013;15(6):e117) doi:[10.2196/jmir.2638](https://doi.org/10.2196/jmir.2638)

KEYWORDS

videoconferencing; general practice; patient appointments; health care

Introduction

Australia is a geographically dispersed country in which one-third of the population lives in rural and remote locations. Inequity in health care is thought to be linked to the cost of

medical appointments and to the shortage of medical manpower in rural and remote areas [1,2]. In other sectors, access to services has been facilitated by information technology. For example, there is growing evidence for the role of information

technology to improve the customer experience in the retail and finance industries [3,4].

In theory, access to doctors can also be efficiently facilitated by online video technology [5]. However, video technology requires both practitioners and patients to be willing to consult via the Internet. To date, the deployment of online video technology in Australian primary care is not routine practice. The practice is limited to government-subsidized consultations involving specialist practitioners or to small numbers of privately funded schemes [6,7].

People who consult doctors in general practice are heterogeneous [8]. The reasons for seeking medical advice range from self-limiting conditions of recent onset to chronic and life-limiting problems [8]. The symptoms or problems presented may warrant information, education, reassurance, explanation, examination, prescription, referral, and/or investigations. The consultation provides an opportunity to address the patient's current problems, and also to consider and potentially prevent future problems [9].

In a face-to-face consultation, the doctor can use all 5 senses; however, in an Internet-based video consultation access to sensory information is limited, and the information that is available may be hampered by download speeds and/or the performance of computer hardware. Furthermore, there is no scope to intervene in person if the patient requires immediate resuscitation. Many of these limitations also apply to telephone consultations; yet, in parts of the world telephone consultations in primary care are considered routine and time saving [10,11].

This Internet-based pilot study aimed to explore Australian general practitioners' attitudes to video consultation with a range of patients who may not be known to them previously. The survey tool used in this study was based on the theory of planned behavior (TPB) [12].

Methods

This study was approved by the Curtin Human Research Ethics Committee (No: RD-61-12).

Participants were recruited from members of the Curtin Health Innovation Research Network (CHIReN), a virtual network of general practitioners (GPs) across Australia who have already consented to be invited to participate in studies with standardized patients. The study took place over approximately 12 weeks. Each participant was remunerated AUS \$50 as recompense for his/her time to participate in the study.

Participants answered questions after viewing video-recorded monologs by actor-patients. Video scenarios were produced and validated by a team of 6 GPs. Six videos were produced, each featuring an actor-patient presenting a range of clinical problems. Information on medical history, family history, and drug history were offered at the outset of each video. The range of scenarios is consistent with those reported in the GP activity reports [8,13]. Scenarios are described in Table 1. A screenshot from 1 of the videos is shown in Figure 1. The vignettes ranged from a self-limiting minor illness to a life-threatening medical emergency. Participating GPs provided their demographic details and answered questions about their impressions (see Table 2).

Figure 1. Video consultation vignette.



Table 1. Scenarios presented to participants in videos.

Video vignette	Patient information	Description of condition
1	Patient: Fay Connolly. Occupation: dental nurse. Age: 49. Nonsmoker. Alcohol: 2 units per week at most	History: Anxiety and depression for 3 months. Has refused antidepressants in the past, now struggling to cope. Having recurrent panic attacks, can't sleep at night. No energy, loss of libido, can't concentrate for any length of time. Tearful. Wants help. Not sure can keep going to work anymore. Last consultation June 2011: Bit anxious and depressed, referred to counselor-did not attend appointments. Husband is very worried
2	Patient: Lucy Jones. Occupation: unemployed. Age: 51. Smoker: 30/day. Alcohol: 6 units per week (as recorded in 2008)	History: Abnormal liver function tests-drinking half bottle of vodka every day. Started drinking after divorce 4 years ago, a few glasses of wine 3-4 times a week, now drinking steadily every day. Was stopped by police 1 week ago and facing court appearance for driving while intoxicated. Son is at university and now refusing to visit his mum. Was found sleeping in her own vomit in the bathroom 1 week ago. Spending most of her money on alcohol, rent hasn't been paid for 2 months. Feels depressed most of the time. Not eating well, tired all the time. Last consultation June 2010: upper respiratory tract infection
3	Patient: Adrian Marshall. Occupation: store room supervisor. Age: 49	History: Recurrent bouts of crushing chest pain. Started at 3 am this morning and now has been present for several hours. Nothing seems to help. Has vomited several times; feeling a bit breathless. Pain is worse when moving about, but now severe even when sitting quietly. Feeling a bit dizzy. Been sweating a bit. Feeling very worried. Past history: hypertension on Ace inhibitor since 2006. Last blood pressure recorded 6 months ago: 155/96 mmHg. Smoker 20/day. Wife is very worried. Seen on video with arm around patient
4	Patient: Nellie O'Reilly. Occupation: receptionist. Age: 59. Nonsmoker. Nondrinker	History: Recurrent bouts of central abdominal pain for 2 months, has had 6 attacks so far. No specific pattern to the symptoms, can occur at any time, pain is mainly on the right-hand side and can last 2 hours. Occasionally vomits during an attack. Left with dull ache after each episode. Sometimes feels like she is wearing a tight band around her upper abdomen and some aching of right shoulder. Last attack was 2 days ago in the middle of the night. Son offered to drive her to hospital but by the time he arrived at her house the pain was gone. Worried it might be something serious. Last consultation September 2011: upper respiratory tract infection. Advised over the counter analgesia and fluids
5	Patient: Richard Cunningham. Occupation: Truck driver. Age: 48. Smoker: 20/day	History: Diabetes mellitus on biguanide for 6 months. Blood pressure: 156/80 mmHg (3 weeks ago), previous readings: 150/85 mmHg (3 months ago). Diet: poor. Body mass index: 30 Cholesterol: 6.7 mmol/L, LDL: 2.5 mmol/L. HbA1C: 7.5 mmol/L. Needs repeat medication: Biguanide and calcium channel blocker. Last consultation November 2011: attended for driving license medical
6	Patient: Mary Smith. Occupation: librarian. Age: 60. Nonsmoker	History: Cough and sore throat for 3 days. Doesn't feel unwell otherwise, but can't manage at work because she needs to use her voice. Cough keeping her awake at night. Current medication: Ace inhibitor for hypertension for the past 2 years. No allergies. Wants an antibiotic-always seems to help. Last consult August 2011: upper respiratory tract infection. Prescribed: antibiotic

Table 2. Questions for participants after each video.

Domain	Question	Question type	Options/range
Intention	1. What is your differential diagnosis?	Free-text comment	
	2. Would you continue with this consultation online?	Choose 1 option	Yes/no/maybe
	3. How difficult was it for you to suggest a diagnosis for this scenario?	7-point scale	Not at all difficult to very difficult
Attitude	4. Managing a patient like this online is:	7-point scale	Harmful to beneficial
		7-point scale	Worthless to useful
		7-point scale	Not convenient (for me) to convenient (for me)
Subjective norm	5. Most patients/GPs/specialists think I ____ consult people like this online	7-point scale	Should to should not
	6. Medicare thinks I ____ consult people like this online	7-point scale	Should to should not
	7. It is expected of me that I should consult patients like this online	7-point scale	Strongly agree to strongly disagree
	8. I feel under social pressure to consult patients like this online	7-point scale	Strongly agree to strongly disagree
Perceived behavior control: self-efficacy	9. I am confident that I could consult patients like this online if I wanted to	7-point scale	Strongly agree to strongly disagree
	10. For me to consult patients like this online is:	7-point scale	Easy to difficult
Perceived behavior control: controllability	11. The decision to consult patients like this online is beyond my control	7-point scale	Strongly agree to strongly disagree
	12. Whether I consult online is entirely up to me	7-point scale	Strongly agree to strongly disagree

Theory of Planned Behavior

This theory postulates that a person's behavior is determined by his/her intention to perform the behavior. This intention is determined by their attitude toward the specific behavior, their subjective norms, and their perceived behavioral control. These 3 domains were explored in this study with reference to scenarios depicting clinical challenges regularly presented to GPs in Australia [13]. The items of the TPB, as measured in this study, were as follows, and each is taken from a guide to the development of such questionnaires [14].

Intention

The respondents were presented with a video scenario and asked whether they would continue with the video consultation. They were also asked to suggest the level of difficulty of making a diagnosis on a scale from 1 to 7, with 1 being not at all difficult and 7 being extremely difficult. The higher the number, the stronger the intention to perform the behavior. For diagnosis difficulty, we calculated the mean of responses for each participant (which may modify the relationship between intention and actual behavior) or the mean for all participants across each scenario (which may reflect differences between scenarios).

Attitude

Direct measurement of attitude involved the use of bipolar adjectives (ie, pairs of opposites), which are evaluative (eg, good–bad). We calculated the mean of the item scores to give an overall attitude score. The attitude items were also scored for internal consistency (Cronbach alpha=.88).

Subjective Norms

Direct measurement involves the use of questions referring to the opinions of important people in general. The subjective norm items were scored for internal consistency (Cronbach alpha=.87). We calculated the mean of the item scores to give an overall subjective norm score.

Perceived Behavior Control

This was achieved by assessing the respondent's self-efficacy and their beliefs about the controllability of the behavior. Self-efficacy was assessed by asking people to report how (1) difficult it was to perform the behavior, and (2) confident they were that they could do it. Controllability was assessed by asking people to report whether (1) performing the behavior was up to them, or (2) factors beyond their control determined their behavior.

Scoring

We checked that the subjective norm items had high internal consistency (self-efficacy: Cronbach alpha=.84; controllability: Cronbach alpha=.72). We calculated the mean of the item scores of the 2 questions related to self-efficacy (questions 9 and 10 in Table 2), and the mean of the item scores of the controllability-related questions (questions 11 and 12 in Table 2) to give mean scores for self-efficacy and controllability. The mean scores of the 2 self-efficacy questions and the 2 controllability questions were also calculated to give an overall perceived behavioral control score (Cronbach alpha=.58). Scores for questions 9, 10, and 12 were reversed before calculation, so that high scores consistently reflected a greater level of control over the target behavior.

Analysis

We confirmed that all internal consistency coefficients were acceptable (> 0.7); therefore, it was appropriate to include all the items in the composite variables. Using a multiple regression procedure, we entered intention as the dependent variable, and the direct measures of attitude, subjective norm, and perceived behavioral control (self-efficacy and controllability) as the predictor variables.

Sample Size

A sample of 47 to 62 GPs would give us 80% power to reject the null hypothesis; that is, 50% of GPs would proceed with the consultation in most cases if the true proportion of GPs who choose to proceed was 70%. Such proportions are consistent with previous research [15].

Data Collection and Analysis

Multivariable logistic regression was used to determine if any identifiable subgroups of GPs, according to demographic criteria, showed significant differences in their scores. *P* values less than .05 were considered statistically significant. Stata version 12.1 (StataCorp LP, College Station, TX, USA) was used to perform the analysis. Multiple regression models were adjusted for the lack of independence between individual participants by estimating the clustered standard errors to account for intragroup correlation (vce option in Stata).

Results

Forty-seven GPs were recruited from the 102 that CHiReN has on file, which is a response rate of 47%. One general practitioner

omitted all demographic questions and was excluded from the analysis, resulting in a total sample of 46. The demography of the sample is reported in Table 3, which also demonstrates a few significant differences between the participants and the known average profile of GPs in Australia.

The GPs offered a differential diagnosis for each video and the level of difficulty in making a diagnosis for that scenario. Data in Table 4 demonstrate that scenario 4 (patient with gall bladder disease) was the most difficult to diagnose, closely followed by scenario 2 (patient with alcoholism) and scenario 6 (patient with an acute cough).

Table 5 summarizes the participant's intention to continue with the consultation. Respondents were least likely to continue with the video consultation with the patient who appeared to be having chest pain, and most likely to continue with the patient seeking a repeat prescription for diabetes and hypertension.

Data relating to the TPB are presented in Table 6. A range of views were expressed with further comments presented subsequently. For the current study, the results show that GPs' self-efficacy and controllability toward online consultations are not in the same direction. When calculating the overall Cronbach alpha for the behavior control score, the 2 controllability scores are in the opposite direction (negative direction) of the 2 self-efficacy scores (positive direction). Therefore, we report self-efficacy and controllability separately. Although the overall behavior score was calculated and reported, it was not used in the regression model as a predictor because of the internal consistency of 4 items.

The relative risk (RR) ratio of difficulty of diagnosis and TPB scores associated with the GPs' intention to continue the consultation within each scenario is presented in Table 7. The ambivalent views and negative views are compared to positive views in multinomial logistic regression. Table 8 presents the results of regression analysis with intention to continue to consult as the dependent variable and scenarios and GP demographic factors as predictive variables. Results in the table are relative risks ratios for the groups who said that "maybe" or they "will not" continue the consultation compared with those who answered yes (RR=1). Results are derived from 1 multinomial logistic regression. Only significant variables, those with $P<.05$, were retained in the final model and reported. Table 9 presents the comments to each scenario as recorded by the participants after each scenario.

Table 3. Demography of participating general practitioners compared to nationally reported group data (where available).

Participants' details	Participant numbers n=46	National group
Age (years), mean (SD)	42 (11)	50.5 ^a
Gender (male), n (%)	26 (57)	56% ^b
Qualifications and experience		
Years since graduation, mean (SD)	18 (11)	No data
Years working as GP, mean (SD)	13 (11)	No data
Number of GPs in the clinic, mean (SD)	7 (4)	≥7 (29%) ^b
GP sessions/week, mean (SD)	7 (3)	No data
GP registrar/GP in training (yes), n (%)	8 (17)	1000 (3.8%) ^c
FRACGP (yes), n (%)	29 (63)	54% ^b
Accredited (yes), n (%)	45 (98)	91% ^b
Position, n (%)		
Principal	9 (20)	No data
Nonprincipal	30 (65)	No data
Other	7 (15)	No data
State, n (%)		
New South Wales	5 (11)	31.6% ^d
Queensland	4 (9)	17.7% ^d
Victoria	12 (26)	26.2% ^d
South Australia	2 (4)	9.2% ^d
Tasmania	1 (2)	2.4% ^d
Western Australia	21 (46)	10% ^d
Australian Capital Territory	1 (2)	1.8% ^d
Region of the clinic, n (%)		
Capital	21 (46)	No data
Other metropolitan	18 (39)	No data
Large rural	2 (4)	No data
Small rural	3 (7)	No data
Remote center	2 (4)	No data
Practice details, n (%)		
Major cities	32 (70)	71%
Inner regional	6 (13)	No data
Outer regional	3 (6)	No data
Remote	5 (11)	No data
Country where graduated university, n (%)		
Non-Australia	14 (30)	
Australia	32 (70)	67%
Patients seen/week, n (%)		
<100	20 (43)	No data
100-149	17 (37)	No data

Participants' details	Participant numbers n=46	National group
150-199	9 (20)	No data
Direct patient care hours/week (hours), n (%)		
<11	4 (9)	
11-20	5 (11)	11% ^b
21-40	30 (65)	56% ^b
41-60	7 (15)	33% ^b
Non-English consultations, n (%)		
No	35 (76)	
Yes, less than 25%	10 (22)	
Yes, more than 50%	1 (2)	24% ^b

^a[16]^b[13]^c[17]^d[18]**Table 4.** Diagnosis and rating for each video (N=46).

Video	Most common diagnosis		Number of differential diagnosis		Level of difficulty in making a diagnosis (scale 1 to 7)	
	Diagnosis	%	Mean (SD)	Median (IQR)	Mean (SD)	Median(IQR)
1	Anxiety and depression	71.7	2.5 (1.2)	2.0 (1)	2.2 (1.3)	2.0 (2)
2	Alcoholism	95.7	3.0 (1.4)	3.0 (2)	3.3 (1.9)	3.0 (3)
3	Myocardial infarction	82.6	2.6 (1.7)	2.0 (3)	1.6 (1.1)	1.0 (1)
4	Gall bladder disease	93.5	2.7 (2.0)	2.0 (2)	3.6 (2.0)	3.0 (4)
5	Diabetes mellitus plus hypertension	78.3	2.3 (1.1)	2.0 (1)	1.9 (1.1)	2.0 (1)
6	Upper respiratory tract infection	100	1.9 (1.1)	2.0 (1)	3.2 (1.8)	2 (3)

Table 5. Comments made by GPs regarding intention of continuing with each video consultation (N=46).

Video	Most common diagnosis	Intend to continue with consultation, %		
		Yes	Maybe	No
1	Anxiety and depression	31	41	28
2	Alcoholism	15	48	37
3	Myocardial infarction	2	7	91
4	Gall bladder disease	20	24	56
5	Diabetes, plus hypertension	41	35	24
6	Upper respiratory tract infection (URTI)	17	50	33

Table 6. Participants' response as per the domains of TPB per scenario (N=46).

Scenario	TPB, mean (SD)				
	Attitudes	Subjective norms	PBC control efficacy	PBC control controllability	Behavioral score ^a
Anxiety and depression	4.3 (1.1)	2.9 (1.0)	3.8 (1.7)	5.3 (1.8)	4.6 (1.1)
Alcoholism	3.8 (1.3)	2.7 (1.1)	3.3 (1.6)	5.5 (1.7)	4.4 (1.1)
Myocardial infarction	2.3 (1.6)	1.9 (1.3)	2.3 (1.7)	5.5 (1.6)	3.9 (1.1)
Gall bladder disease	3.4 (1.6)	2.7 (1.1)	3.3 (1.7)	5.4 (1.7)	4.4 (1.2)
Diabetes plus hypertension	4.5 (1.6)	3.6 (1.3)	4.6 (1.6)	5.5 (1.7)	5.0 (1.2)
URTI	3.9 (1.6)	3.6 (1.3)	4.0 (1.6)	5.5 (1.7)	4.8 (1.1)

^aMean score of the 2 PBC subcategories.**Table 7.** The relative risk (RR) ratio of difficulty of diagnosis and TPB scores associated with GPs' intention to continue the consultation within each scenario (N=46).

Risk and TPB	Scenario, RR (95% CI) ^a				
	Anxiety	Alcoholism	Gall bladder disease	Diabetes	URTI
Of saying "maybe" vs "yes"					
Difficulty	0.4 (0.2-1.0)	0.4 (0.2-1.0)	0.9 (0.3-2.8)	0.7 (0.3-1.8)	25 (0-100)
Attitudes	0.2 (0.1-0.7) ^b	0.3 (0.1-0.9) ^b	0.3 (0.1-0.9) ^b	0.4 (0.2-0.97) ^b	0.1 (0-0.7) ^b
Subjective norms	5.3 (1.1-25.7) ^b	1.8 (0.6-5.2)	1.1 (0.4-3.0)	0.6 (0.2-1.9)	0.3 (0-2.6)
Self-efficacy	0.8 (0.4-1.8)	0.9 (0.5-1.7)	2.3 (0.8-6.8)	0.9 (0.4-1.9)	1.3 (0.5-3.5)
Controllability	0.9 (0.6-1.4)	1.2 (0.7-2.1)	1.1 (0.5-2.4)	0.9 (0.5-1.4)	0 (0-100)
Of saying "no" vs "yes"					
Difficulty	0.4 (0.2-1.0)	0.4 (0.2-0.9) ^b	1.1 (0.4-3.4)	0.7 (0.2-2.0)	100 (0-100)
Attitudes	0.2 (0-1.1)	0.3 (0.1-0.8) ^b	0.1 (0-0.4) ^d	0.3 (0.1-0.9) ^b	0.1 (0-0.6) ^b
Subjective norms	6.0 (1.1-31.9) ^b	1.0 (0.4-3.0)	2.5 (0.7-8.8)	0.3 (0.1-1.3)	0.1 (0-1.3)
Self-efficacy	0.2 (0.1-0.6) ^c	0.5 (0.3-1.1)	1.0 (0.3-2.8)	0.8 (0.3-2.0)	1.0 (0.3-3.4)
Controllability	0.8 (0.4-1.6)	1.5 (0.8-2.5)	1.0 (0.4-2.5)	1.3 (0.7-2.6)	0 (0-17.8)

^aFor the groups who said that they "maybe" or "will not" continue the consultation compared with those who answered yes (RR=1). Results are derived from 6 multinomial logistic regressions according to the scenario; result values greater than 100 are truncated to 100. Myocardial infarction is not reported because only one 1 participant said yes and 3 participants said maybe. Due to such small numbers in some categories, it was not possible to include that scenario in the model.

^b $P < .05$ ^c $P < .01$ ^d $P < .001$

Table 8. Sociodemographic and scenarios as indicators associated with the intention to continue the consultation online.

Factors modeled	RR (95% CI)
Saying "maybe" compared to "yes"	
Scenario (anxiety, RR=1)	
Alcoholism	2.9 (1.1-7.7) ^a
Myocardial infarction	2.6 (0.3-26.0)
Gall bladder disease	0.9 (0.3-3.1)
Diabetes	0.5 (0.2-1.4)
URTI	2.6 (0.8-8.8)
Country of university (non-Australian, RR=1)	6.8 (1.8-25.2) ^b
Years after graduation	1.4 (1.2-1.7) ^c
Years as a GP	0.8 (0.7-0.9) ^c
GP registrar ("no," RR=1)	0.8 (0.2-3.1)
Clinic remoteness	0.6 (0.4-1.1)
Number of GPs	0.9 (0.7-1)a
Nonprincipal, principal/others (RR=1)	6.5 (1.8-22.8) ^b
Hours practiced/week	4.5 (2.0-9.9) ^c
Saying "no" compared to "yes"	
Scenario (anxiety, RR=1)	
Alcoholism	3.2 (1.2-8.5) ^a
Myocardial infarction	74.2 (7.9-695.7) ^c
Gall bladder disease	3.9 (1.0-13.0) ^a
Diabetes	0.5 (0.2-1.5)
URTI	2.4 (0.7-7.8)
Country of university (non-Australian, RR=1)	11.0 (3.1-39.8) ^c
Years after graduation	1.2 (1.0-1.5) ^a
Years as a GP	0.8 (0.7-0.98) ^a
GP registrar (no, RR=1)	0.3 (0.1-0.96) ^a
Clinic remoteness	0.5 (0.3-0.9) ^b
Number of GPs	0.9 (0.8-0.98) ^a
Nonprincipal, principal/others (RR=1)	1.9 (0.6-5.5)
Hours practiced/week	2.6 (1.3-5.2) ^b

^a $P < .05$ ^b $P < .01$ ^c $P < .001$

Table 9. Free-text comments per scenario.

Video (most common diagnosis)	Continue with video consultation?		
	Yes (n) ^a	No (n) ^a	Maybe (n) ^a
Anxiety and depression	"A lot could be sorted out for her online"(7)	"An online consultation would not be ideal as it may be harder to establish rapport" (3)	"Needs to be examined"(3)
Alcoholism	"This is another case where an initial treatment plan could be made online." (3)	"...the use of an online consultation in this case inhibits developing rapport particularly with a patient whom I have only seen occasionally" (2); "Needs physical examination and probably blood tests"(4)	"Needs physical examination" (11)
Myocardial infarction	"Depending on how far away he is from me, I would either go to him now after calling an ambulance to him, or if he is too far way, I keep talking to him after calling an ambulance to take him to hospital" (1)	"Needs to call an ambulance urgently" (32)	"Cardiac chest pain must come to surgery or to ED" (1)
Gall bladder disease	"Would like to see her for follow-up consultation for examination"(6)	"Needs examination" (5)	"Requires examination, cannot be achieved online" (23)
Diabetes plus hypertension	"Most of the issues in this consult could be managed online quite effectively" (10)	"Could be a convenient way of discussing results, however this would be variable upon the results. Still has issue of being unable to examine" (12)	"Would like to see him in person to reinforce the importance of the control of diabetes, quitting smoking, cholesterol and blood pressure" (3)
URTI	"I would not necessarily prescribe antibiotics. There are no symptoms that make me concerned about a chest problem" (2)	"It is very hard to manage this case online without physical exam" (10)	"You would have a lot of difficulty justifying to the patient why you have not prescribed antibiotics, when you have not examined the patient" (4)

^aNumber in brackets represents number of participants making similar comments.

Discussion

General practitioner participants in this study might conduct a video consultation with patients other than those presenting with what could be an acute life-threatening emergency. Participants formed 3 approximately equal groups: those who would continue with the video consultation, those who were ambivalent, and those who would not. The scenario involving the person with anxiety evoked this typical response. GPs who had qualified from an Australian university were more likely to be equivocal about video consultations. Similar opinions were expressed by those medical practitioners who had been qualified for longer. However, those who had been in general practice for longer and those who worked in group practices were more likely to favor video consultations.

Compared with the case of the patient presenting with anxiety and depression, GPs were more likely to reject continuing with a video consultation with the patient with alcohol dependence, myocardial infarction, or gall bladder disease. Their objections focused primarily around the inability to physically interact with the patient. Those who had been qualified as a medical practitioner for longer and those who worked longer hours were more likely to express negative views. On the other hand, participants who had been practicing as GPs for longer, GP registrars, those who worked in remote practice, and those from

larger group practices were less likely to be negative about video consultations.

Practitioners who were ambivalent about continuing with video consultations expressed the view that it was difficult to diagnose the patient presenting with symptoms of a cough, albeit an upper respiratory tract infection. By contrast, they had positive views about managing patients online for all but the chest pain scenario. They were also concerned about the attitude of significant others, such as patients, colleagues, and funders, about conducting video consultations particularly in the case of the patient with anxiety and depression. With regard to the low consistency between self-efficacy and controllability, we postulated that GPs did not feel they were able to conduct video consultations even if they wanted to because such consultations are not subsidized by government funding. However, they felt confident about their ability to conduct video consultations.

GPs are unlikely to offer video consultations for patients with a minor self-limiting illness of recent onset. The reticence could be ascribed to the perceived need for a physical examination. The importance of clinical examination to establish a diagnosis for acute cough has been emphasized in previous literature [19]. However, a recent review suggests that physical signs, if present, have poor predictive value to detect infections that may benefit from antibiotics [20]. It is also highly unlikely that practitioners would consider video consultations for patients who have a life-threatening medical emergency. This may stem from the

perception of an increased risk of failing to make an appropriate diagnosis in this context and the need for immediate resuscitation of a patient with cardiac chest pain [21,22]. Consultations for patients with chronic mental health issues may also be hampered unless there are clear indications that such consultations are approved by colleagues and funding agencies.

Limitations

The practitioners who participated in this study were generally younger than most Australian GPs. They were also more likely to be registrars and those working in a larger practice than average. Although the geographical distribution included a representative sample of practitioners from rural and remote areas, there were fewer from some states in Australia. We also acknowledge that the doctors had no opportunity to ask questions or seek clarification from the actor-patients. This was a significant issue, although it would have been difficult within the limitations of technology and the resources available to allow for such interactions in circumstances in which practitioners across Australia, living in different time zones, wished to participate in their own time and actors were only funded to perform a single vignette. Experience from previous studies with live consultations between actors and practitioner were that more limited numbers of practitioners were able and willing to participate. This introduces bias and less generalizable results [23]. Secondly, we did not interview the practitioners to explore their perspective on the consultations. Our comments are, therefore, limited to their responses to a questionnaire. In these circumstances, we can only draw limited conclusions that domains of the theory of learned behavior offer a recognized theoretical grounding to frame the conclusions. The free-text comments also provide further information on the impressions of the practitioners.

Strengths

The greatest strength of this study was the use of Internet-based video vignettes to gauge GP opinion. Video vignettes have significant advantages over other data collection methods, in particular, the advantage of realism, or something closely approximating it [23-25]. Video vignettes can present patient information to clinicians in a way that closely resembles their usual consultations. More understanding is generated in this way than by surveys on usual practice. In the context of this study, video consultations are not yet routine practice; therefore, it was important to present the practitioners with examples of scenarios. Video scenarios can simulate clinicians' usual working environments and generate a range of typical responses from them in a way that questions asked in the absence of an example scenario or in relation to a text-based scenario cannot. Data collected are likely to be valid. Such simulations could also be used to introduce video consultations to students at medical schools.

Conclusions

Australian GPs may adopt video consultations in their practice, but this is likely to be in larger practices with more established GPs, especially in rural areas. It is also likely that access to video consultations will need to focus on follow-up consultations, where the purpose of the consultation is not primarily to establish a diagnosis. Patients with minor self-limiting illnesses and those with medical emergencies are unlikely to be offered access to a GP by video. Medical practitioners appear confident about their ability to conduct video consultations; however, the process of establishing video consultations as routine practice will need to be endorsed by patients, members of the profession, and funding organizations. Video consultation techniques will also need to be taught in medical schools. Future research on this topic could follow a similar outline with vignettes, but include more interactive video consultations between practitioners and patients.

Acknowledgments

We would like to thank Dr Allison Rieck, Dr Shoreh Razmi, and Ms Jo Higgins for support in administering the survey. We also thank Dr Mike Civil, Dr Shohreh Razmi, Dr Marthe Smith, and Dr Devesh Oberoi for assistance in developing and administering the scenarios and Tammy McCausland for copyediting the final submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Video vignette alcohol dependence.

[[MOV File, 11MB - jmir_v15i6e117_app1.mov](#)]

Multimedia Appendix 2

Video vignette sore throat and cough.

[[MOV File, 11MB - jmir_v15i6e117_app2.mov](#)]

Multimedia Appendix 3

Video vignette diabetes.

[[MOV File, 6MB](#) - [jmir_v15i6e117_app3.mov](#)]

Multimedia Appendix 4

Video vignette chest pain.

[[MOV File, 8MB](#) - [jmir_v15i6e117_app4.mov](#)]

Multimedia Appendix 5

Video vignette anxiety.

[[MOV File, 11MB](#) - [jmir_v15i6e117_app5.mov](#)]

Multimedia Appendix 6

Video vignette gall bladder symptoms.

[[MOV File, 12MB](#) - [jmir_v15i6e117_app6.mov](#)]

References

1. Bourke L, Humphreys JS, Wakeman J, Taylor J. Understanding rural and remote health: a framework for analysis in Australia. *Health Place* 2012 May;18(3):496-503. [doi: [10.1016/j.healthplace.2012.02.009](#)] [Medline: [22418016](#)]
2. Young AF, Dobson AJ, Byles JE. Determinants of general practitioner use among women in Australia. *Soc Sci Med* 2001 Dec;53(12):1641-1651. [Medline: [11762890](#)]
3. Grewal D, Ailawadi KL, Gauri D, Hall K, Kopalle P, Robertson JR. Innovations in retail pricing and promotions. *Journal of Retailing* 2011 Jul;87:S43-S52. [doi: [10.1016/j.jretai.2011.04.008](#)]
4. von Hippel EA, Oliveira P. Users as service innovators: the case of banking services. *SSRN Journal* 2011:806-818. [doi: [10.2139/ssrn.1460751](#)]
5. Moffatt JJ, Eley DS. The reported benefits of telehealth for rural Australians. *Aust Health Rev* 2010 Aug;34(3):276-281. [doi: [10.1071/AH09794](#)] [Medline: [20797357](#)]
6. Australian Government Department of Human Services: Medicare. 2012. Telehealth URL: <http://www.medicareaustralia.gov.au/provider/incentives/telehealth/index.jsp> [accessed 2013-05-23] [WebCite Cache ID 6GpCLTxCS]
7. GP2U. 2012. Expert doctors available when you need them URL: <https://gp2u.com.au/> [accessed 2013-03-21] [WebCite Cache ID 6FIUOcJVV]
8. Britt H, Miller GC, Charles J, Henderson J, Bayram C, Pan Y, et al. General Practice Activity in Australia. Canberra, Australia: AIHW; 2008.
9. Deveugele M, Derese A, De Bacquer D, van den Brink-Muinen A, Bensing J, De Maeseneer J. Consultation in general practice: a standard operating procedure? *Patient Educ Couns* 2004 Aug;54(2):227-233. [doi: [10.1016/S0738-3991\(03\)00239-8](#)] [Medline: [15288919](#)]
10. Jiwa M, Mathers N, Campbell M. The effect of GP telephone triage on numbers seeking same-day appointments. *Br J Gen Pract* 2002 May;52(478):390-391 [FREE Full text] [Medline: [12014537](#)]
11. Alexandra D, Sá M, Navarro P, Santos S, Bernandes AC. P-547 "Hello?": telephone consultation in general practice – a Portuguese setting. In: *Swiss Medical Weekly Supplementum*. 2009 Presented at: 15th Wonca Europe Conference; September 16-19, 2009; Basel, Switzerland p. 201-202 URL: <http://www.smw.ch/docs/PdfContent/smw-12931.pdf>
12. Ajzen I, Madden TJ. Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. *Journal of Experimental Social Psychology* 1986 Sep;22(5):453-474. [doi: [10.1016/0022-1031\(86\)90045-4](#)]
13. Britt H, Miller GC, Charles J, Henderson J, Bayram C, Pan Y, et al. General Practice Activity in Australia. Canberra, Australia: AIHW; 2009.
14. Francis JJ, Eccles MP, Johnston M, Walker A, Grimshaw J, Foy R, et al. *Constructing Questionnaires Based on the Theory of Planned Behaviour: A Manual for Health Services Researchers*. Newcastle Upon Tyne, UK: Centre for Health Services Research University of Newcastle; 2004 May. URL: http://pages.bangor.ac.uk/~pes004/exercise_psych/downloads/tpb_manual.pdf [accessed 2012-03-26] [WebCite Cache ID 66Rg0YB0P]
15. Hanna L, May C, Fairhurst K. The place of information and communication technology-mediated consultations in primary care: GPs' perspectives. *Fam Pract* 2012 Jun;29(3):361-366. [doi: [10.1093/fampra/cmr087](#)] [Medline: [22006040](#)]
16. AIHW. Medical Workforce. Canberra, Australia: Australian Institute of Health and Welfare; 2011:69.
17. GPET Annual Report to 30 June 2012. Canberra, Australia: General Practice Education and Training Limited; 2012. URL: <http://agpt.com.au/PoliciesPublications/AnnualReport/> [accessed 2013-03-21] [WebCite Cache ID 6FIV7Qni8]

18. Primary Health Care Research & Information Service (PHC RIS). 2013. Fast Facts: GP numbers in Australia, 1999-2000 to 2010-11 URL: <http://www.phcris.org.au/fastfacts/fact.php?id=6775> [accessed 2013-03-22] [[WebCite Cache ID 6FIVRDEu8](#)]
19. Cals JW, Francis NA. Acute cough in adults. *BMJ* 2010 Feb;340:c574. [doi: [10.1136/bmj.c574](https://doi.org/10.1136/bmj.c574)] [Medline: [20154050](#)]
20. Aalbers J, O'Brien KK, Chan WS, Falk GA, Teljeur C, Dimitrov BD, et al. Predicting streptococcal pharyngitis in adults in primary care: a systematic review of the diagnostic accuracy of symptoms and signs and validation of the Centor score. *BMC Med* 2011;9:67 [[FREE Full text](#)] [doi: [10.1186/1741-7015-9-67](https://doi.org/10.1186/1741-7015-9-67)] [Medline: [21631919](#)]
21. Stevens RJ, Lasserson D. In primary care, when is chest pain due to coronary artery disease? *CMAJ* 2010 Sep 7;182(12):1281-1282 [[FREE Full text](#)] [doi: [10.1503/cmaj.100808](https://doi.org/10.1503/cmaj.100808)] [Medline: [20603344](#)]
22. National Heart Foundation of Australia, The Cardiac Society of Australia and New Zealand. Guidelines for the management of acute coronary syndromes 2006: acute management of chest pain. *MJA Suppl* 2006 Apr 17;184(8):S11-S12 [[FREE Full text](#)]
23. Jiwa M, McKinley RK, Spilsbury K, Arnet H, Smith M. Deploying a clinical innovation in the context of actor-patient consultations in general practice: a prelude to a formal clinical trial. *BMC Med Res Methodol* 2009;9:54 [[FREE Full text](#)] [doi: [10.1186/1471-2288-9-54](https://doi.org/10.1186/1471-2288-9-54)] [Medline: [19615058](#)]
24. Hughes R. Considering the vignette technique and its application to a study of drug injecting and HIV risk and safer behaviour. *Sociol Health & Illness* 1998 May;20(3):381-400. [doi: [10.1111/1467-9566.00107](https://doi.org/10.1111/1467-9566.00107)]
25. Hawley C, Adams A, Galbraith N. Video Vignette Study of Attitudes and Behaviour of Health Professionals in Consultations. Road Safety Research Report No. 91 Sub-report. London: Department of Transport; 2010:5.

Abbreviations

CHIReN: Curtin Health Innovation Research Network

GP: general practitioner

RR: relative risk

TPB: theory of planned behavior

Edited by G Eysenbach; submitted 28.03.13; peer-reviewed by B McKinsty, N Ozcakar; comments to author 21.04.13; revised version received 22.04.13; accepted 23.04.13; published 19.06.13.

Please cite as:

Jiwa M, Meng X

Video Consultation Use by Australian General Practitioners: Video Vignette Study

J Med Internet Res 2013;15(6):e117

URL: <http://www.jmir.org/2013/6/e117/>

doi: [10.2196/jmir.2638](https://doi.org/10.2196/jmir.2638)

PMID: [23782753](https://pubmed.ncbi.nlm.nih.gov/23782753/)

©Moyez Jiwa, Xingqiong Meng. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 19.06.2013. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Two New Meal- and Web-Based Interactive Food Frequency Questionnaires: Validation of Energy and Macronutrient Intake

Sara E Christensen¹, MSc; Elisabeth Möller¹, MSc; Stephanie E Bonn¹, MSc; Alexander Ploner¹, PhD; Antony Wright², BSC(Hons); Arvid Sjölander¹, PhD; Olle Bälter³, PhD; Lauren Lissner⁴, PhD; Katarina Bälter¹, PhD

¹Karolinska Institutet, Department of Medical Epidemiology and Biostatistics, Stockholm, Sweden

²UK Medical Research Council, MRC Collaborative Center for Human Nutrition Research, Cambridge, United Kingdom

³KTH - Royal Institute of Technology, School of Computer Science and Communication, Stockholm, Sweden

⁴University of Gothenburg, Department of Public Health and Community Medicine, Gothenburg, Sweden

Corresponding Author:

Sara E Christensen, MSc

Karolinska Institutet

Department of Medical Epidemiology and Biostatistics

Nobels väg 12a

Stockholm, SE-171 77

Sweden

Phone: 46 852482361

Fax: 46 8314975

Email: sara.christensen@ki.se

Abstract

Background: Meal-Q and its shorter version, MiniMeal-Q, are 2 new Web-based food frequency questionnaires. Their meal-based and interactive format was designed to promote ease of use and to minimize answering time, desirable improvements in large epidemiological studies.

Objective: We evaluated the validity of energy and macronutrient intake assessed with Meal-Q and MiniMeal-Q as well as the reproducibility of Meal-Q.

Methods: Healthy volunteers aged 20-63 years recruited from Stockholm County filled out the 174-item Meal-Q. The questionnaire was compared to 7-day weighed food records (WFR; n=163), for energy and macronutrient intake, and to doubly labeled water (DLW; n=39), for total energy expenditure. In addition, the 126-item MiniMeal-Q was evaluated in a simulated validation using truncated Meal-Q data. We also assessed the answering time and ease of use of both questionnaires.

Results: Bland-Altman plots showed a varying bias within the intake range for all validity comparisons. Cross-classification of quartiles placed 70%-86% in the same/adjacent quartile with WFR and 77% with DLW. Deattenuated and energy-adjusted Pearson correlation coefficients with the WFR ranged from $r=0.33$ - 0.74 for macronutrients and was $r=0.18$ for energy. Correlations with DLW were $r=0.42$ for Meal-Q and $r=0.38$ for MiniMeal-Q. Intraclass correlations for Meal-Q ranged from $r=0.57$ - 0.90 . Median answering time was 17 minutes for Meal-Q and 7 minutes for MiniMeal-Q, and participants rated both questionnaires as easy to use.

Conclusions: Meal-Q and MiniMeal-Q are easy to use and have short answering times. The ranking agreement is good for most of the nutrients for both questionnaires and Meal-Q shows fair reproducibility.

(*J Med Internet Res* 2013;15(6):e109) doi:[10.2196/jmir.2458](https://doi.org/10.2196/jmir.2458)

KEYWORDS

validity; reproducibility; food frequency questionnaire; Internet; weighed food record; doubly labeled water; adult

Introduction

The food frequency questionnaire (FFQ) is a commonly used method for assessing diet in large-scale epidemiological studies. The advantages of the FFQ include a low participant burden

compared to dietary records, and low cost because it is typically a self-administered method. However, there is a need for methodological improvement, including the FFQ layout and its ease of use.

Most FFQs list food items according to food groups (vegetables, meats, dairy, etc), yet people typically consume food grouped into meals. Moreover, meal-based questionnaire designs have been shown to facilitate recall of dietary intake in previous studies [1,2]. Therefore, we developed a meal- and Web-based FFQ, called Meal-Q, with a design that allows for individually adapted follow-up questions. Thus, participants only answer questions relevant to their own food habits. For example, a high consumer of bread and cheese will get follow-up questions about the number of slices of bread and cheese, whereas a low consumer will not. This feature reduces the answering time and improves the ease of use.

Approximately 90% of the Swedish adult population used the Internet in 2011 [3], justifying development of Web-based questionnaires for national population-based studies. Furthermore, the Web-based design makes the use of Meal-Q more cost-efficient than a paper-based FFQ and facilitates assessment of large samples. The ability to use built-in checks for missing answers and the immediate transfer of answers into digital format also assures complete data collection and improves data quality [4,5].

We evaluated the validity and reproducibility of energy and macronutrient intake assessed with Meal-Q by comparing it to a weighed food record (WFR) and doubly labeled water (DLW). By using truncated data from Meal-Q, we also validated a shorter version called MiniMeal-Q.

Methods

Background

The development of Meal-Q was based on results from a population-based study in which 700 randomly selected Swedish participants reported, through either face-to-face interviews or telephone 24-hour recalls, on the food products they consumed for breakfast, lunch, dinner, and snacks (E Möller and S Christensen, personal written communication, August 2008). This dietary information guided the design of a meal- and Web-based FFQ called MaxMeal-Q. After a pretest of MaxMeal-Q in a randomly selected group of individuals (N=216), the shorter version, Meal-Q, was formed by omitting less commonly consumed food items and dishes. Subsequently, Meal-Q was included in the Validation of Methods Assessing diet and physical activity (VALMA) study. The reference methods were a 7-day WFR on the Web and DLW for estimation of energy expenditure [6]. The Research Ethics committee at Karolinska Institutet approved the study.

After the validation study was completed, researchers from LifeGene, a large population-based cohort study [7], decided to use Meal-Q under the condition that the answering time be reduced. Therefore, we developed the shorter version, MiniMeal-Q, by omitting food items consumed on average with a low intake frequency and that contributed least to the total energy and nutrient intake. Yet, food items representing important food sources of certain nutrients were kept (eg, black pudding that contributes to iron intake). After a time test, LifeGene decided to use MiniMeal-Q. We validated MiniMeal-Q in the present study by truncating Meal-Q data to simulate MiniMeal-Q. The inherent dependence between Meal-Q and MiniMeal-Q should be taken into account when comparing their validity.

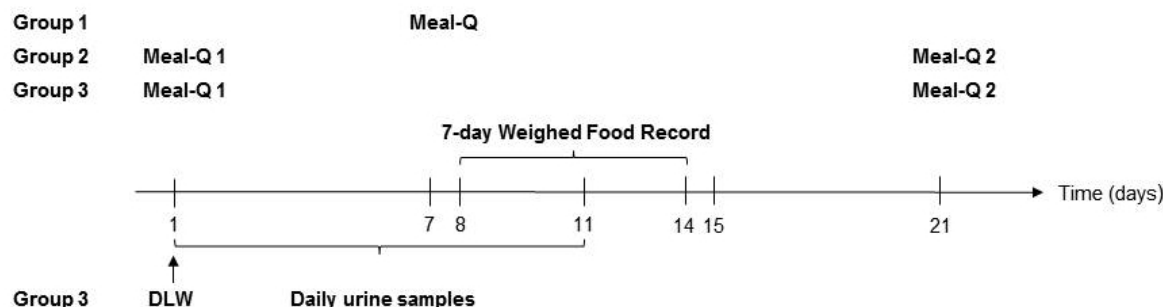
Recruitment

In April 2009, 180 healthy volunteer men and women aged 20 to 63 years were recruited to the VALMA study through public advertisement in Stockholm County, Sweden. Recruitment also took place at 3 universities including announcements among nutritionist students. Access to the Internet and an email address were prerequisites for eligibility, as well as not being on a weight-loss diet, not being pregnant, and not having given birth within the past 10 months. At an introductory meeting, participants were informed about the study and signed informed consent forms. Participants self-reported their height and weight, which was used to calculate body mass index (BMI).

Study Design

After recruitment, the participants were divided into 3 age- and gender-balanced groups: group 1 (n=87), group 2 (n=53), and group 3 (n=40). Each group followed a 3-week study scheme shown in Figure 1.

All groups filled out Meal-Q and the WFR on the Web at their own choice of location (eg, at home) and group 3 was also given DLW. Groups 2 and 3 filled out a second Meal-Q after 3 weeks. Data from the first administered Meal-Q assessment and the WFR from all groups were compared for validity evaluation. The data from the first Meal-Q assessment was truncated for simulated analysis of MiniMeal-Q. The first and second Meal-Q assessments from groups 2 and 3 were compared for reproducibility evaluation. Information about education, occupation, and tobacco use (smoking and Swedish snuff) was collected in the first questionnaire. Answering time was automatically recorded, and directly after completion of the first Meal-Q, a short Web survey followed to evaluate its ease of use.

Figure 1. The 3-week study scheme of the VALMA study.

Dietary Assessment

Meal-Q

The interactive Meal-Q included 102 to 174 food items (depending on the number of follow-up questions) and asked about dietary intake during the past few months. For an example of a questionnaire module, see [Figure 2](#). Meal-Q assessed intake of (1) food items, dishes, and beverages, (2) energy and nutrients, including alcohol, (3) supplements, (4) meal patterns, and (5) eating behavior, such as restaurant visits, intake of fast food, light products, probiotics, and the use of cooking fat and salt. Respondents chose from predefined food items and intake frequencies and only filled in what they ate at least once a month. For each of the following food groups, 5 photos of portion sizes were included: (1) rice, potatoes, and pasta, (2) meat, chicken, fish, and vegetarian substitutes, and (3) vegetables (raw or cooked). The photos were used to calculate portion sizes for cooked dishes and vegetables. For other food items, standard portion sizes were used based on information from the National Food Agency, the Swedish Consumer Agency, measured portion sizes developed by the research group, as well as standard portion sizes used in other FFQs at Karolinska Institutet.

MiniMeal-Q

MiniMeal-Q includes 75 to 126 food items—approximately 30% fewer items than Meal-Q—and has similar questions on meal patterns and eating behavior. After the VALMA study was finished, MiniMeal-Q was sent out to 79 volunteer VALMA participants to assess answering time and ease of use.

Weighed Food Records on the Web

At study start, participants were given oral instructions, a kitchen scale, and a handbook with instructions on how to complete the 7-day WFR by using a Web-based program. Participants were asked to weigh and report all consumed food products and beverages at the highest detail level possible (eg, each food item in a dish was encouraged to be reported in its individual weight). The participants could choose among over 2000 food items in the program's food database, and they also recorded which day they consumed the food item as well as for which meal (ie, breakfast, lunch, dinner, or between meals). Data collectors

checked all records for completeness. In the program, participants also provided a 7-day pedometer-based physical activity record. The participants were asked to report their total number of daily steps as well as other activities not captured by pedometers, such as bicycling or swimming. From this, the physical activity level (PAL) was calculated for each participant and the information was used for identification of potential energy underreporters in the WFR by using the Goldberg cut-off method [8].

Nutrient Database

Intake of food items and dishes from Meal-Q, MiniMeal-Q, and the WFR were converted into energy (kJ/day) and macronutrient (g/day) intake using the national database on nutrient content published by the Swedish National Food Agency [9]. The nutrient conversion for the questionnaires was done by computer programs developed and validated specifically for this study, whereas the conversion of the WFR was done with the Web-based WFR program. Dietary supplements were not included in the analyses.

Doubly Labeled Water Method

Total energy expenditure was determined in group 3 (n=40) using the DLW method [10] over 11 consecutive days ([Figure 1](#)). The details of this procedure have been described previously [11]. Briefly, on day 1 at the study site, each participant provided a 5-milliliter urine sample before receiving an oral DLW dose calculated according to body weight [12]. Subsequently, daily spot urine samples were collected for a total of 9 days. Participants were instructed to collect urine samples at a similar time each day (but not the first void of the day). All samples were kept refrigerated. On day 11, the 9 urine samples were returned to the study site and the eleventh urine sample was collected. All samples were shipped to the Medical Research Council, Human Nutrition Research, Cambridge, United Kingdom, for isotopic analysis, which has been previously described in detail [13]. Enrichments of $^2\text{H}/^1\text{H}$ and $^{18}\text{O}/^{16}\text{O}$ in urine samples were determined by mass spectrometry. Following conversion to the universally accepted Vienna Standard Mean Ocean Water (VSMOW) / Standard Light Arctic Precipitation (SLAP) scale, total energy expenditure (TEE) was calculated by using standard equations [14-16]. CO_2 production (mole/day)

was estimated using Schoeller et al's correction for fractionation [15] and a respiratory quotient of 0.85. The respiratory quotient is based on omnivores with 30% to 35% energy contribution from fat and suitable to the VALMA population. The results of the CO₂ production were used to calculate the TEE of each participant by using the modified Weir equation [17].

Statistical Analysis

Descriptive characteristics of study participants are presented as mean (SD) and as counts (%). Differences in BMI and age between study groups, between men and women, and between included and excluded participants were assessed using a 2-sample *t* test. Differences in education, nutrition background (studying or working in the nutrition field), and tobacco use were assessed using Fisher's exact test. The level of statistical significance was set to $\alpha = .05$.

Median and interquartile range (IQR) of crude energy and macronutrient intake was calculated and compared among Meal-Q, MiniMeal-Q, and the WFR. Energy intake from the questionnaires was also compared to TEE from DLW. Wilcoxon signed rank tests were used to determine differences between all methods. The median (IQR) answering time in minutes of each questionnaire was calculated and ease of use was evaluated from the Web survey. The between-person variance captured in the truncated MiniMeal-Q as compared to Meal-Q was calculated using linear regression.

For validity and reproducibility analyses, macronutrients were adjusted for total energy intake using the residual method [18]. Variables deviating from the normal distribution were

transformed using the square, square root, or log transformation, as appropriate. Absolute agreement and potential difference in bias within the intake range were evaluated by plotting the differences between questionnaires and WFR or DLW against the average of the 2 methods, according to the method of Bland and Altman [19]. The degree of variation was represented by the limits of agreement, ie, ± 2 SD of the mean difference. The ranking agreement and magnitude of misclassification when comparing questionnaires with the WFR and DLW was tested by dividing participants into quartile categories of energy and energy-adjusted macronutrient intake. Proportions of participants classified into the same, adjacent, and extreme quartiles were calculated. Because variables were normally distributed after energy adjustment and transformation, Pearson correlation coefficients were used to measure the degree of linear relationship between the questionnaires and the WFR and DLW. Deattenuated correlations corrected for within-person variation in the WFR were calculated using the formulas of Beaton et al [20] and Liu et al [21], and confidence intervals (CI) were produced using the method of Willett and Rosner [22]. Confidence intervals for correlations with DLW were obtained using the bootstrap method [23].

Reproducibility of Meal-Q was evaluated by comparing crude median energy and macronutrient intake between the first and second Meal-Q and by cross-classification of energy and energy-adjusted [18] quartiles of macronutrient intake. Intraclass correlation coefficients (ICCs) [24] were also computed using 1-way ANOVA with random effects. Statistical analyses were performed using STATA statistical software version 11.2 (StataCorp LP, College Station, TX, USA).

Figure 2. Screenshot of a Meal-Q module: breakfast and snack items and follow-up question on bread (translated from the Swedish questionnaire version in the VALMA study).

For the type of food you eat at least once a month, choose in the drop down menu how often you eat them.

Only fill out what you usually eat.

	Times per day	Times per week	Times per month
White bread (eg tin loaf, loaf, flatbread)	... ▼	... ▼	... ▼
Whole grain bread (eg rye bread, whole meal bread, Rusk)	... ▼	... ▼	... ▼
Crisp bread	... ▼	... ▼	... ▼
Processed sour milk, yoghurt, yoghurt drink, smoothie	... ▼	... ▼	... ▼
Muesli, cereals	... ▼	... ▼	... ▼
Oatmeal porridge	... ▼	... ▼	... ▼
Flaxseeds	... ▼	... ▼	... ▼
Hard cheese	... ▼	... ▼	... ▼
Bag cheese, dessert cheese (eg Philadelphia, brie)	... ▼	... ▼	... ▼
Marmalade, jam, apple purée, honey	... ▼	... ▼	... ▼
Liver pâté	... ▼	... ▼	... ▼
Cold cuts (eg ham, salami)	... ▼	... ▼	... ▼
Egg, omelete	... ▼	... ▼	... ▼

<< >>

You have mentioned that you eat bread. How many slices of bread do you usually eat each time?

☐ 1-2 slices
☐ 3-4 slices
☐ 5-6 slices
☐ 7 slices or more
☐ Don't know/Don't want to answer

<< >>

Results

One participant was excluded due to dropout (group 1) and 2 others due to illness (group 2 and 3). Eleven participants (4 in group 1, 6 in group 2, and 1 in group 3) were identified as energy underreporters by applying the Goldberg cut-off [8] on energy intake from the WFR together with data from each participant's calculated PAL. Additionally, 3 underreporters (group 3) in the WFR were identified using individual PAL values calculated from DLW data. Because of the implausible energy intake by the WFR, the 14 underreporters were excluded for the validity comparison between Meal-Q, MiniMeal-Q, and the WFR; therefore, 163 participants remained (group 1: n=82; group 2: n=46; and group 3: n=35). For the validity comparison between Meal-Q, MiniMeal-Q, and DLW in group 3, no

exclusion of energy underreporters was made; therefore, 39 participants remained. For the reproducibility analysis of Meal-Q, 4 participants had missing values in the second administered Meal-Q; therefore, 87 participants remained. We found no statistically significant differences between included and excluded participants in terms of age, BMI, education, nutrition background, or tobacco use.

Descriptive Statistics

As shown in Table 1, most of the study participants were students or highly educated. One-third were working full time, and almost as many had a nutrition background. Few participants used tobacco. There was no statistically significant difference between groups or sexes regarding age, BMI, education, nutrition background, or smoking (but more men than women used Swedish snuff).

Table 1. Characteristics of the participants in the validation study (n=167^a).

Characteristics	By group			By gender		All (n=167)
	Group 1 (n=82)	Group 2 (n=46)	Group 3 (n=39)	Men (n=35)	Women (n=132)	
Gender, n (%)						
Male	16 (19.5)	11 (23.9)	8 (20.5)			35 (21.0)
Female	66 (80.5)	35 (76.1)	31 (79.5)			132 (79.0)
Age (years), mean (SD)	34 (12)	31 (11)	33 (12)	33 (10)	33 (12)	33 (12)
BMI (kg/m ²), mean (SD)	23 (3.6)	23 (3.4)	23 (3.7)	24 (2.2)	23 (3.8)	23 (3.6)
Education >12 years, n (%)	64 (78.0)	38 (82.6)	32 (82.1)	28 (80.0)	106 (80.3)	134 (80.2)
Working full time, n (%)	33 (40.2)	12 (26.1)	10 (25.6)	12 (34.3)	43 (32.6)	55 (32.9)
Student, n (%)	41 (50.0)	31 (67.4)	26 (66.7)	19 (54.3)	79 (59.8)	98 (58.7)
Background in nutrition ^b , n (%)	21 (25.6)	15 (32.6)	13 (33.3)	6 (17.1)	43 (32.6)	49 (29.3)
Tobacco use ^c , n (%)	11 (13.4)	5 (10.9)	6 (15.4)	12 (34.3)	10 (7.6)	22 (13.2)

^aFrom this study sample, 4 underreporters were excluded for analysis with the WFR (n=163). There were no statistically significant differences in characteristics between groups or sexes, except for Swedish snuff between sexes (1.8% women and 4.2% men, $P=.001$) via 2-sample t test and Fisher's exact test.

^bStudying or working in the nutrition field.

^cTobacco use = smoking and/or Swedish snuff. Values are missing for 3 women in group 3.

The median time to answer the Meal-Q and the MiniMeal-Q was 17 (IQR 11) and 7 (IQR 4) minutes, respectively. Most (92%) participants perceived Meal-Q as easy to fill out, 91% thought the questions were relevant, and 93% reported that food items and dishes were presented in a logical order. For MiniMeal-Q, the figures were 95%, 88%, and 91%, respectively. The overall mean grade of Meal-Q and MiniMeal-Q's ease of use was 4.2 on a 5-point scale in which 5 was the best grade. The between-person variance captured by MiniMeal-Q as compared to Meal-Q ranged from 96% to 99% for energy and macronutrients.

Validity

Energy and macronutrient intake was higher in the WFR compared with both questionnaires, except for polyunsaturated fat assessed with Meal-Q (Table 2). In group 3 (n=39), the energy expenditure from DLW was higher than energy intake assessed by both questionnaires ($P<.001$, Wilcoxon's signed rank test). The energy expenditure from DLW was 11,423 kJ (IQR 2777) and the energy intake assessed with Meal-Q and MiniMeal-Q were 7954 kJ (IQR 2736) and 7358 kJ (IQR 2718), respectively.

As shown in Figure 3, the Bland-Altman plots with DLW indicate that the WFR and both questionnaires underestimated energy intake for most participants. Compared to the WFR, the questionnaires had a larger underestimation, larger variance, and a weak trend of decreasing accuracy with increasing intakes.

The Bland-Altman plots of Meal-Q and the WFR in Figure 4 showed a negative mean difference for energy and all macronutrients. There was a trend of decreasing accuracy with

increasing energy and polyunsaturated fat intake, and trends of increasing underestimation with increasing intakes for other macronutrients. Because of varying bias within the intake range, the proportion of participants outside the limits of agreement deviated somewhat from 5% for some plots. Bland-Altman plots for MiniMeal-Q and the WFR were very similar to those for Meal-Q and the WFR (see Multimedia Appendix 1).

Table 3 shows that the proportion of participants classified into the same or adjacent quartile for energy was 70% by Meal-Q and 67% by MiniMeal-Q as compared to the WFR. Correspondingly, the proportions for macronutrients ranged from 76% to 86% for both questionnaires. Quartile cross-classification of Meal-Q and DLW placed 77% into the same or adjacent quartile, and values were identical for MiniMeal-Q.

Pearson correlation coefficients (r) with the WFR and DLW were similar between Meal-Q and MiniMeal-Q (Table 4). Deattenuated and energy-adjusted correlations with the WFR ranged from $r=0.18$ - 0.73 for Meal-Q and from $r=0.18$ - 0.74 for MiniMeal-Q. Correlation with DLW was $r=0.42$ for Meal-Q and $r=0.38$ for MiniMeal-Q.

Reproducibility

Table 5 shows that there were no statistically significant differences in crude intakes between the first and second Meal-Q assessments. The proportion of participants classified into the same or adjacent quartile ranged from 85% to 96%. ICCs ranged from $r=0.43$ - 0.92 for crude intakes and from $r=0.57$ - 0.90 for energy-adjusted macronutrients.

Table 2. Daily crude energy and macronutrient intake assessed with the WFR, Meal-Q, and MiniMeal-Q (n=163).

Energy and macronutrients	WFR	Meal-Q ^a		MiniMeal-Q ^a	
	Median (IQR)	Median (IQR)	% of WFR	Median (IQR)	% of WFR
Energy (kJ)	9183 (2340)	7667 (3723)	83	7017 (3632)	76
Protein (g)	85 (37)	79 (40)	93	70 (34)	82
Carbohydrates (g)	243 (97)	211 (132)	87	190 (124)	78
Total fat (g)	86 (37)	65 (34)	76	62 (35)	72
Saturated fat (g)	33 (18)	22 (14)	67	20 (13)	61
Monounsaturated fat (g)	31 (16)	23 (13)	74	22 (11)	71
Polyunsaturated fat (g)	14 (8)	13 (8)	93	12 (9)	86
Alcohol (g)	6 (15)	5 (8)	85	5 (8)	83

^aIntakes assessed with Meal-Q and MiniMeal-Q were statistically significantly different from the WFR ($P=.01$), except for polyunsaturated fat assessed with Meal-Q ($P=.28$). Intakes assessed with Meal-Q and MiniMeal-Q were statistically significantly different from each other ($P<.001$) via Wilcoxon signed rank test.

Table 3. Quartile cross-classification of mean daily energy and energy-adjusted^a macronutrient intake assessed with Meal-Q, MiniMeal-Q, and the WFR (n=136) and cross-classification of mean daily energy intake and energy expenditure measured with DLW (n=39).

Energy and macronutrients	Same quartile, %		Adjacent quartile, %		Extreme quartile, %	
	Meal-Q	MiniMeal-Q	Meal-Q	MiniMeal-Q	Meal-Q	MiniMeal-Q
Energy	26	27	44	40	8.5	7.4
Protein	36	40	40	36	6.7	5.5
Carbohydrate	42	42	40	37	2.5	1.8
Total fat	37	33	41	46	8.0	9.2
Saturated fat	52	45	33	37	4.9	4.3
Monounsaturated fat	44	44	33	33	6.1	6.7
Polyunsaturated fat	33	31	47	49	5.5	4.9
Alcohol	50	49	36	37	4.3	3.7
DLW, energy (kJ)	33	33	44	44	2.6	2.6

^aAdjustments for total energy intake were made using the residual method [18].

Table 4. Pearson correlation coefficients between Meal-Q, MiniMeal-Q, and the WFR (n=163) and DLW (n=39).

Energy and macronutrients	Crude ^a		Energy-adjusted ^{a,b}		Deattenuated (95% CI) ^c	
	Meal-Q	MiniMeal-Q	Meal-Q	MiniMeal-Q	Meal-Q	MiniMeal-Q
Energy	0.16	0.16	—	—	0.18 (0.01-0.36)	0.18 (0.01-0.33)
Protein	0.22	0.21	0.30	0.31	0.33 (0.17-0.47)	0.34 (0.18-0.48)
Carbohydrates	0.54	0.54	0.62	0.57	0.65 (0.54-0.74)	0.60 (0.48-0.70)
Total fat	0.06	0.02	0.55	0.49	0.57 (0.45-0.67)	0.51 (0.37-0.62)
Saturated fat	0.15	0.11	0.57	0.54	0.60 (0.48-0.70)	0.57 (0.44-0.67)
Monounsaturated fat	0.13	0.08	0.52	0.46	0.56 (0.43-0.67)	0.50 (0.36-0.62)
Polyunsaturated fat	0.23	0.21	0.36	0.35	0.42 (0.25-0.56)	0.40 (0.23-0.54)
Alcohol	0.64	0.65	0.61	0.63	0.73 (0.59-0.82)	0.74 (0.60-0.83)
DLW, energy (CI) ^d	0.42 (0.16-0.68)	0.38 (0.10-0.66)	—	—	—	—

^aAll correlation coefficients were statistically significant ($P = <.001-.046$), except for crude total, saturated and monounsaturated fat for both questionnaires ($P = .06-.84$).

^bAdjustments for energy were made using the residual method [18].

^cDeattenuated values were obtained using the formulas suggested by Beaton et al [20] and Liu et al [21]. Confidence intervals were produced using the method suggested by Willett and Rosner [22].

^dConfidence intervals were obtained using the bootstrap method [23].

Table 5. Daily energy and macronutrient intake assessed with the 2 Meal-Q assessments in groups 2 and 3, quartile cross-classifications and crude and energy-adjusted^a intraclass correlation coefficients^b (ICC) (n=87)^c.

Energy and macronutrients	Median (IQR) intake			Quartile cross-classifications, %			ICC (95% CI)	
	Meal-Q 1	Meal-Q 2 ^c	Difference ^d	Same	Adjacent	Extreme	Crude	Energy-adjusted
Energy (kJ)	7720 (3567)	7383 (3205)	-125 (2497)	51	34	6.9	0.57 (0.42-0.71)	—
Protein (g)	79 (36)	78 (29)	-1.2 (24)	53	40	2.3	0.52 (0.37-0.67)	0.73 (0.63-0.83)
Carbohydrates (g)	209 (122)	206 (113)	0.7 (82)	52	41	2.3	0.64 (0.51-0.76)	0.67 (0.56-0.80)
Total fat (g)	62 (30)	62 (29)	-1.9 (23)	59	26	5.7	0.47 (0.30-0.63)	0.57 (0.43-0.71)
Saturated fat (g)	20 (11)	21 (13)	-0.9 (7.5)	61	25	3.4	0.43 (0.26-0.60)	0.58 (0.44-0.72)
Monounsaturated fat (g)	22 (12)	23 (10)	-0.4 (8.6)	56	32	3.4	0.50 (0.34-0.66)	0.60 (0.46-0.73)
Polyunsaturated fat (g)	13 (9.0)	13 (8.2)	-0.01 (4.84)	57	36	3.4	0.65 (0.53-0.77)	0.68 (0.56-0.79)
Alcohol (g)	4.6 (8.5)	4.3 (7.0)	-1.0 (2.0)	74	22	1.1	0.92 (0.89-0.95)	0.90 (0.87-0.94)

^aAdjustments for energy were made using the residual method [18].

^bIntraclass correlation coefficients [24] were computed using 1-way ANOVA with random effects.

^cMissing values on Meal-Q 2 for 4 participants.

^dMeal-Q 1–Meal-Q 2; $P = .27-.96$ via Wilcoxon signed rank test.

Figure 3. Bland-Altman plots showing the differences in energy intake assessed with the WFR, Meal-Q, and MiniMeal-Q and the energy expenditure measured with DLW plotted against the mean of the 2 methods (n=39). Each data point represents 1 participant. The long-dashed line shows the mean difference. The short-dashed lines show the 95% limits of agreement (mean difference ± 2 SD).

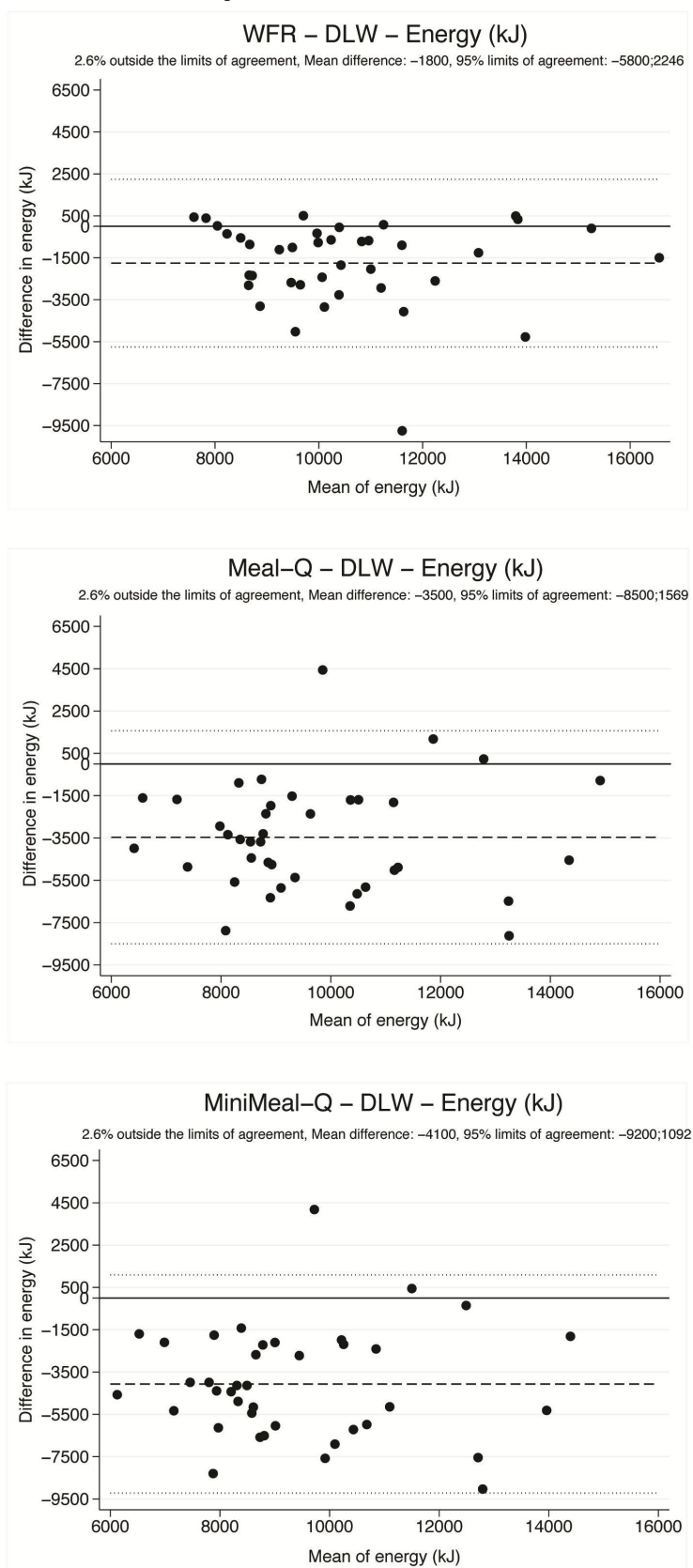
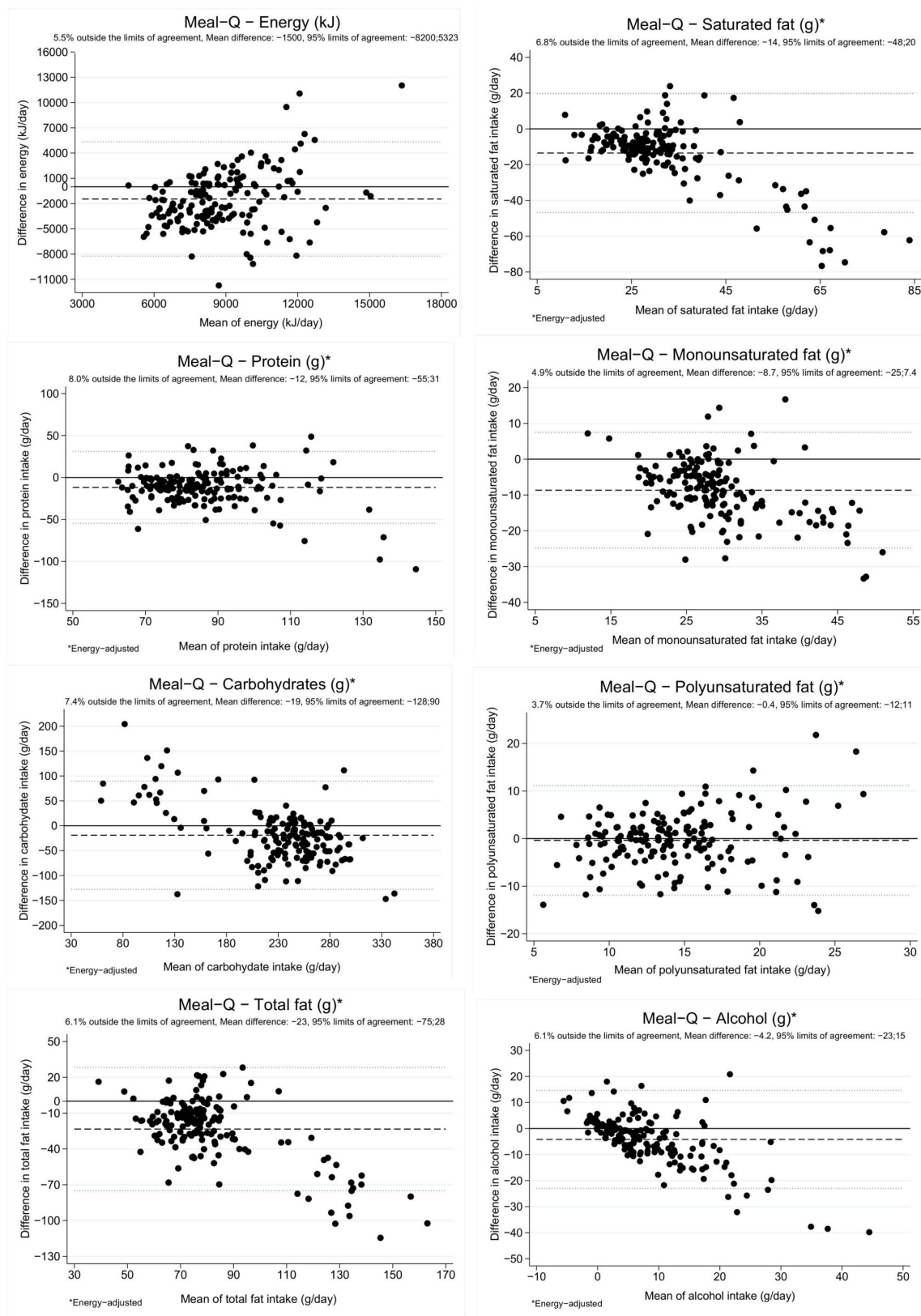


Figure 4. Bland-Altman plots showing the differences in energy, protein, carbohydrate, total fat, saturated fat, monounsaturated fat, polyunsaturated fat, and alcohol intake assessed with Meal-Q and intake assessed with the WFR plotted against the mean of the 2 methods (N=163). Macronutrients are energy-adjusted using the residual method [18].



Discussion

Principal Results

This study presents the validity and reproducibility of the new meal- and Web-based interactive Meal-Q, as well as a simulated validation of its shorter version, MiniMeal-Q. Both questionnaires were perceived as easy to use and had a short answering time. Trends of varying bias within the intake range were seen for energy and all macronutrients. Both questionnaires showed good ranking ability for carbohydrates, total fat, saturated fat, monounsaturated fat, and alcohol, whereas energy, protein, and polyunsaturated fat performed less well. Furthermore, Meal-Q showed fair reproducibility.

Comparison With Prior Work

The Bland-Altman plots of the questionnaires versus WFR and DLW and the plot on WFR versus DLW showed a varying bias within the intake range. Energy and polyunsaturated fat both seemed to be underestimated and overestimated for both questionnaires at higher intake levels. For other macronutrients, the plots indicated that the questionnaires had difficulty assessing higher intakes. In contrast, quartile cross-classification with the WFR showed a fair ranking agreement for most of the nutrients, although a lower agreement was seen for energy and polyunsaturated fat. Similar rankings have been seen in 3 other validation studies of FFQs against food records [25-27] including a Web-based method [25]. In nutritional epidemiology, the association between diet and disease is commonly studied by ranking the dietary intake; therefore, absolute intake is often less important than good ranking order [18]. Hence, despite an underestimation of absolute intake, the ranking agreement for Meal-Q and MiniMeal-Q suggests they are useful in epidemiologic studies regarding most nutrients.

The correlations between the questionnaires and the WFR ranged from 0.18 for energy to 0.74 for alcohol. High correlations between FFQs and diet records are in the order of 0.6-0.7 and it is unlikely that correlations above 0.8 can be obtained [28]. A review of FFQs concluded that the mean correlation with food records of ≥ 6 days was 0.42 for energy, 0.57 for total fat, 0.53 for protein, 0.58 for carbohydrates, and 0.76 for alcohol [29]. In light of this literature, energy and protein seemed to perform less well, whereas other macronutrients showed correlations within expected ranges. A limitation of the FFQ methodology is the predefined number of food items, frequencies, and portion sizes, which could lead to a “flattened slope” effect in scatter plots [30]. This is a result of respondents consuming little food to unintentionally overreport, and for those consuming a lot to underreport. Correlations from such data would be artificially low. However, a truly small between-person variance would also give similar results [31]. Therefore, the low to moderate correlations seen in this study could reflect a limitation of the questionnaire design, but may also reveal a true small between-person variance. Bland and Altman have discouraged the use of correlation coefficients to evaluate validity because they do not measure agreement [19]. However, because the use of correlations in validation studies is widespread, we have included them to enable comparisons with other studies.

The DLW measurements in group 3 showed that the WFR, Meal-Q, and MiniMeal-Q underestimated energy intake by 17%, 30%, and 36%, respectively. Similar figures for food records and FFQs have been seen in other studies using DLW [32-34]. The Bland-Altman plots showed that the underestimation of energy was considerable for both questionnaires and the large variance indicated difficulties in precision. The underestimation and variance was much smaller for the WFR. Correlations with DLW were moderate for both questionnaires, although the CIs were wide because of the large variance. The correlations were similar to a study by Andersen et al [35], but slightly lower than that of Kroke et al [36]. Despite the underestimation and the large variance, quartile cross-classification with DLW showed a fair ranking agreement, similar to that found by Kroke et al [36].

The moderate to strong quartile cross-classifications of the first and second Meal-Q assessments suggest the questionnaire has fair reproducibility. Correlations between repeated administrations of FFQs in other studies have ranged $r=0.5-0.8$ [31], and Meal-Q showed similar results. The reproducibility might have been affected by the short time period between the Meal-Q assessments, because participants are less likely to have true changes in intake or response after a short compared to a longer period [37]. In addition, it is important to keep in mind that the reproducibility cannot reveal systematic errors, which can be masked in a high correlation between 2 questionnaires.

The high between-person variance captured by MiniMeal-Q as compared to Meal-Q indicates that it is possible to use a shorter questionnaire while still assessing a similar intake range and keeping the ranking ability. Because MiniMeal-Q originates from Meal-Q data and is also compared to the same reference methods, their results become highly related. Therefore, caution must be taken when comparing their assessments and relative validity.

Regular use of the Internet in Sweden is higher among young people compared to older people. Among those aged 16 to 44 years, 88% to 94% use the Internet daily, whereas the proportion among the age groups 45 to 54 years and 65 to 74 years are 82% to 83% and 38% to 49%, respectively [3]. However, access to the Internet is high for all age groups—more than 90% for the young and middle-aged and 67% to 78% for the oldest age group. It is worth noticing that problems with cognition might be an issue in very old age groups, although this would also hold true for dietary assessment using a paper-based questionnaire. Concerns could be raised regarding whether Web-based questionnaires produce different kinds of bias as compared to paper-based questionnaires. However, bias associated with Web-based data collection does not seem to differ from that of paper-based questionnaires as seen in a large Swedish feasibility study of more than 45,000 participants [38].

Limitations and Strengths

To estimate the validity of a dietary assessment method, 2 statistical assumptions should be fulfilled. First, the assessed dietary intake should be linearly related to true intake. Second, the measurement errors should be independent between the test and the reference method. In this validation study, variables were linearly correlated to the WFR, although energy, protein,

and polyunsaturated fat had a weaker linear relationship. The questionnaires rely on memory and have predefined food items, frequencies, and portion sizes, whereas the WFR does not rely on memory, is open-ended, and has direct assessment of portion sizes. Nevertheless, the methods are linked to the same nutrient database and are likewise affected by social desirability, which could lead to an overestimated validity.

The strengths of this validation study include its large sample size and few dropouts. There was also high compliance to the questionnaires, the WFR and DLW. Using the DLW method is an additional strength that enabled an objective estimation of TEE for the evaluation of energy intake. The digital format of the questionnaires and the WFR also substantially reduced the risk of coding errors and missing data. The proportion of underreporters in this study (14/177, 8%) was notably lower compared with some other studies [39–42], even if studies have had proportions in the range of 2% to 85% [43]. The use of individual PAL values from each participant for the Goldberg cut-off is likely to have increased the sensitivity [44] and could be an explanation.

Due to time constraints, the study period had to be kept short. This could have given an overestimation of the validity because the questionnaires and the WFR assessments were performed within a short period of time. Furthermore, the WFR was only performed once. Ideally, several records with independent days

spread over a longer time period would have reflected the habitual dietary intake better. However, corrections for within-person variance in the WFR were made to minimize day-to-day variation and energy adjustment was made to avoid variations in intake related to total energy intake. The DLW measurement should also preferably have been done repeatedly over a longer time period to reflect habitual energy intake; however, this was not possible. Furthermore, most of the participants were women and many were students with a nutrition background. Also, the self-selection of participants could have biased the sample in favor of more motivated participants who are more inclined to give accurate answers compared to a sample from the general population [45]. Nevertheless, dietary intake from the WFR was in-line with a Swedish national dietary survey using food records (n=1214) [42], suggesting that the WFR intake would be comparable to an assessment within a more general and less-selected population. Acknowledging the highly educated study sample, the answering time might have been longer in a less-educated population.

Conclusions

Meal-Q and MiniMeal-Q are 2 Web-based FFQs shown to be highly user-friendly. Despite their short answering time, they had an ability to rank most macronutrient intakes well compared with the reference methods. In addition, Meal-Q showed fair reproducibility.

Acknowledgments

We would like to thank the participants in the validation study as well as Les Bluck at the Collaborative Center for Human Nutrition Research, UK Medical Research Council, for help with the DLW; Katarina Holm-Johansson at Energibalans.nu for providing the WFR program; Therese Andersson for statistical assistance; and Nina Lundqvist for assistance during the data collection. This study was supported by funds from Torsten and Ragnar Söderberg's Foundation, AFA Insurances, the Swedish Research Council, and the Swedish Council for Working Life and Social Research.

Authors' Contributions

The authors' responsibilities were as follows—SEC, EM, SEB, LL, OB, KB: questionnaire design; SEC, EM, SEB, LL, KB: validation study design; SEC, EM, SEB, KB: data collection; AW: analyses of DLW and interpretation of results; SEC, KB, OB: development and validation of the nutrient calculation programs MealCalc and MiniMealCalc; OB: calculation of nutrients; SEC, AS, AP: statistical analyses; SEC, AP, AS, LL, KB: interpretation of results; SEC: drafted the manuscript; and SEC, EM, SEB, AP, AW, AS, OB, LL, KB: review and revision of the manuscript including approval of final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Bland-Altman plots showing the differences in energy, protein, carbohydrate, total fat, saturated fat, monounsaturated fat, polyunsaturated fat and alcohol intake assessed with MiniMeal-Q and intake assessed with the WFR plotted against the mean of the two methods (n=163). Macronutrients are energy-adjusted using the residual method [19].

[JPG File, 1MB - [jmir_v15i6e109_app1.jpg](http://www.jmir.org/2013/6/e109_app1.jpg)]

References

1. Subar AF, Thompson FE, Potischman N, Forsyth BH, Buday R, Richards D, et al. Formative research of a quick list for an automated self-administered 24-hour dietary recall. *J Am Diet Assoc* 2007 Jun;107(6):1002-1007. [doi: [10.1016/j.jada.2007.03.007](https://doi.org/10.1016/j.jada.2007.03.007)] [Medline: [17524721](https://pubmed.ncbi.nlm.nih.gov/17524721/)]

2. Kohlmeier L. Gaps in dietary assessment methodology: meal- vs list-based methods. *Am J Clin Nutr* 1994 Jan;59(1 Suppl):175S-179S [FREE Full text] [Medline: 8279419]
3. Statistics Sweden. Private persons' use of computers and the Internet 2011 [Swedish]. 2012. URL: http://www.scb.se/Pages/PublishingCalendarViewInfo_259923.aspx?PublObjId=14444 [accessed 2012-11-27] [WebCite Cache ID 6CUuftyoy]
4. van Gelder MM, Bretveld RW, Roeleveld N. Web-based questionnaires: the future in epidemiology? *Am J Epidemiol* 2010 Dec 1;172(11):1292-1298 [FREE Full text] [doi: 10.1093/aje/kwq291] [Medline: 20880962]
5. Ekman A, Litton JE. New times, new needs; e-epidemiology. *Eur J Epidemiol* 2007;22(5):285-292. [doi: 10.1007/s10654-007-9119-0] [Medline: 17505896]
6. Ainslie P, Reilly T, Westerterp K. Estimating human energy expenditure: a review of techniques with particular reference to doubly labeled water. *Sports Med* 2003;33(9):683-698. [Medline: 12846591]
7. Almqvist C, Adami HO, Franks PW, Groop L, Ingelsson E, Kere J, et al. LifeGene--a large prospective population-based study of global relevance. *Eur J Epidemiol* 2011 Jan;26(1):67-77. [doi: 10.1007/s10654-010-9521-x] [Medline: 21104112]
8. Black AE. Critical evaluation of energy intake using the Goldberg cut-off for energy intake: basal metabolic rate. A practical guide to its calculation, use and limitations. *Int J Obes Relat Metab Disord* 2000 Sep;24(9):1119-1130. [Medline: 11033980]
9. National Food Agency. Food database [Swedish]. URL: <http://www7.slv.se/Naringssok/?epslanguage=sv> [accessed 2012-11-27] [WebCite Cache ID 6CU63QALA]
10. Davidsson L. Assessment of body composition and total energy expenditure in humans using stable isotope techniques, IAEA Human Health Series. Vienna: International Atomic Energy Agency; 2009. URL: http://www-pub.iaea.org/MTCD/publications/PDF/Pub1370_web.pdf [accessed 2013-05-06] [WebCite Cache ID 6GPh7wRHN]
11. Bonn SE, Trolle Lagerros Y, Christensen SE, Möller E, Wright A, Sjölander A, et al. Active-Q: validation of the web-based physical activity questionnaire using doubly labeled water. *J Med Internet Res* 2012;14(1):e29 [FREE Full text] [doi: 10.2196/jmir.1974] [Medline: 22356755]
12. Trabulsi J, Troiano RP, Subar AF, Sharbaugh C, Kipnis V, Schatzkin A, et al. Precision of the doubly labeled water method in a large-scale application: evaluation of a streamlined-dosing protocol in the Observing Protein and Energy Nutrition (OPEN) study. *Eur J Clin Nutr* 2003 Nov;57(11):1370-1377. [doi: 10.1038/sj.ejcn.1601698] [Medline: 14576749]
13. Assah FK, Ekelund U, Brage S, Wright A, Mbanya JC, Wareham NJ. Accuracy and validity of a combined heart rate and motion sensor for the measurement of free-living physical activity energy expenditure in adults in Cameroon. *Int J Epidemiol* 2011 Feb;40(1):112-120 [FREE Full text] [doi: 10.1093/ije/dyq098] [Medline: 20529884]
14. Lifson N, Gordon GB, McClintock R. Measurement of total carbon dioxide production by means of D2O18. *J Appl Physiol* 1955 May;7(6):704-710. [Medline: 14381351]
15. Schoeller DA, Ravussin E, Schutz Y, Acheson KJ, Baertschi P, Jéquier E. Energy expenditure by doubly labeled water: validation in humans and proposed calculation. *Am J Physiol* 1986 May;250(5 Pt 2):R823-R830. [Medline: 3085521]
16. Cole TJ, Coward WA. Precision and accuracy of doubly labeled water energy expenditure by multipoint and two-point methods. *Am J Physiol* 1992 Nov;263(5 Pt 1):E965-E973. [Medline: 1443129]
17. Weir JB. New methods for calculating metabolic rate with special reference to protein metabolism. *J Physiol* 1949 Aug;109(1-2):1-9 [FREE Full text] [Medline: 15394301]
18. Willett WC. Monographs in Epidemiology and Biostatistics: Nutritional Epidemiology, 2nd ed. New York: Oxford University Press; 1998.
19. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986 Feb 8;1(8476):307-310. [Medline: 2868172]
20. Beaton GH, Milner J, McGuire V, Feather TE, Little JA. Source of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. Carbohydrate sources, vitamins, and minerals. *Am J Clin Nutr* 1983 Jun;37(6):986-995 [FREE Full text] [Medline: 6846242]
21. Liu K, Stamler J, Dyer A, McKeever J, McKeever P. Statistical methods to assess and minimize the role of intra-individual variability in obscuring the relationship between dietary lipids and serum cholesterol. *J Chronic Dis* 1978;31(6-7):399-418. [Medline: 711832]
22. Rosner B, Willett WC. Interval estimates for correlation coefficients corrected for within-person variation: implications for study design and hypothesis testing. *Am J Epidemiol* 1988 Feb;127(2):377-386. [Medline: 3337089]
23. Wagstaff DA, Elek E, Kulis S, Marsiglia F. Using a nonparametric bootstrap to obtain a confidence interval for Pearson's r with cluster randomized data: a case study. *J Prim Prev* 2009 Sep;30(5):497-512 [FREE Full text] [doi: 10.1007/s10935-009-0191-y] [Medline: 19685290]
24. Lee J, Koh D, Ong CN. Statistical evaluation of agreement between two methods for measuring a quantitative variable. *Comput Biol Med* 1989;19(1):61-70. [Medline: 2917462]
25. Labonté M, Cyr A, Baril-Gravel L, Royer MM, Lamarche B. Validity and reproducibility of a web-based, self-administered food frequency questionnaire. *Eur J Clin Nutr* 2012 Feb;66(2):166-173. [doi: 10.1038/ejcn.2011.163] [Medline: 21934698]
26. George GC, Milani TJ, Hanss-Nuss H, Kim M, Freeland-Graves JH. Development and validation of a semi-quantitative food frequency questionnaire for young adult women in the southwestern United States. *Nutr Res* 2004;24(1):29-43. [doi: 10.1016/j.nutres.2003.09.006]

27. Shatenstein B, Nadon S, Godin C, Ferland G. Development and validation of a food frequency questionnaire. *Can J Diet Pract Res* 2005;66(2):67-75. [Medline: [15975195](#)]
28. Willett WC. Future directions in the development of food-frequency questionnaires. *Am J Clin Nutr* 1994 Jan;59(1 Suppl):171S-174S [FREE Full text] [Medline: [8279418](#)]
29. Molag ML, de Vries JH, Ocké MC, Dagnelie PC, van den Brandt PA, Jansen MC, et al. Design characteristics of food frequency questionnaires in relation to their validity. *Am J Epidemiol* 2007 Dec 15;166(12):1468-1478 [FREE Full text] [doi: [10.1093/aje/kwm236](#)] [Medline: [17881382](#)]
30. Freedman LS, Schatzkin A, Wax Y. The impact of dietary measurement error on planning sample size required in a cohort study. *Am J Epidemiol* 1990 Dec;132(6):1185-1195. [Medline: [2135637](#)]
31. Gibson RS. *Principles of Nutritional Assessment*, 2nd ed. New York: Oxford University Press; 2005.
32. Martin LJ, Su W, Jones PJ, Lockwood GA, Tritchler DL, Boyd NF. Comparison of energy intakes determined by food records and doubly labeled water in women participating in a dietary-intervention trial. *Am J Clin Nutr* 1996 Apr;63(4):483-490 [FREE Full text] [Medline: [8599310](#)]
33. Livingstone MB, Prentice AM, Strain JJ, Coward WA, Black AE, Barker ME, et al. Accuracy of weighed dietary records in studies of diet and health. *BMJ* 1990 Mar 17;300(6726):708-712 [FREE Full text] [Medline: [2386561](#)]
34. Trabulsi J, Schoeller DA. Evaluation of dietary assessment instruments against doubly labeled water, a biomarker of habitual energy intake. *Am J Physiol Endocrinol Metab* 2001 Nov;281(5):E891-E899 [FREE Full text] [Medline: [11595643](#)]
35. Andersen LF, Tomten H, Haggarty P, Løvø A, Hustvedt BE. Validation of energy intake estimated from a food frequency questionnaire: a doubly labeled water study. *Eur J Clin Nutr* 2003 Feb;57(2):279-284. [doi: [10.1038/sj.ejcn.1601519](#)] [Medline: [12571660](#)]
36. Kroke A, Klipstein-Grobusch K, Voss S, Möseneder J, Thielecke F, Noack R, et al. Validation of a self-administered food-frequency questionnaire administered in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study: comparison of energy, protein, and macronutrient intakes estimated with the doubly labeled water, urinary nitrogen, and repeated 24-h dietary recall methods. *Am J Clin Nutr* 1999 Oct;70(4):439-447 [FREE Full text] [Medline: [10500011](#)]
37. Tsubono Y, Nishino Y, Fukao A, Hisamichi S, Tsugane S. Temporal change in the reproducibility of a self-administered food frequency questionnaire. *Am J Epidemiol* 1995 Dec 1;142(11):1231-1235. [Medline: [7485070](#)]
38. Ekman A, Dickman PW, Klint A, Weiderpass E, Litton JE. Feasibility of using web-based questionnaires in large population-based epidemiological studies. *Eur J Epidemiol* 2006;21(2):103-111. [doi: [10.1007/s10654-005-6030-4](#)] [Medline: [16518678](#)]
39. Pryer JA, Vrijheid M, Nichols R, Kiggins M, Elliott P. Who are the 'low energy reporters' in the dietary and nutritional survey of British adults? *Int J Epidemiol* 1997 Feb;26(1):146-154 [FREE Full text] [Medline: [9126514](#)]
40. Hirvonen T, Männistö S, Roos E, Pietinen P. Increasing prevalence of underreporting does not necessarily distort dietary surveys. *Eur J Clin Nutr* 1997 May;51(5):297-301. [Medline: [9152679](#)]
41. Klesges RC, Eck LH, Ray JW. Who underreports dietary intake in a dietary recall? Evidence from the Second National Health and Nutrition Examination Survey. *J Consult Clin Psychol* 1995 Jun;63(3):438-444. [Medline: [7608356](#)]
42. Becker W, Pearson M. National Food Agency. 2002. The second national food consumption survey: dietary habits and nutrient intake in Sweden 1997-98 URL: <http://www.slv.se/en-gb/group1/food-and-nutrition/dietary-surveys1/riksmaten-1997-98-dietary-habits-and-nutrient-intake-in-sweden/> [accessed 2013-05-09] [WebCite Cache ID 6GUPp9LA8]
43. Maurer J, Taren DL, Teixeira PJ, Thomson CA, Lohman TG, Going SB, et al. The psychosocial and behavioral characteristics related to energy misreporting. *Nutr Rev* 2006 Feb;64(2 Pt 1):53-66. [Medline: [16536182](#)]
44. Black AE. The sensitivity and specificity of the Goldberg cut-off for EI:BMR for identifying diet reports of poor validity. *Eur J Clin Nutr* 2000 May;54(5):395-404. [Medline: [10822286](#)]
45. Cade J, Thompson R, Burley V, Warm D. Development, validation and utilisation of food-frequency questionnaires - a review. *Public Health Nutr* 2002 Aug;5(4):567-587. [doi: [10.1079/PHN2001318](#)] [Medline: [12186666](#)]

Abbreviations

BMI: body mass index
DLW: doubly labeled water
FFQ: food frequency questionnaire
ICC: intraclass correlation coefficient
PAL: physical activity level
SLAP: Standard Light Arctic Precipitation
TEE: total energy expenditure
VSMOW: Vienna Standard Mean Ocean Water
VALMA: validation of methods assessing diet and physical activity
WFR: weighed food record

Edited by G Eysenbach; submitted 27.11.12; peer-reviewed by J de Vries; comments to author 29.01.13; revised version received 12.03.13; accepted 12.04.13; published 05.06.13.

Please cite as:

Christensen SE, Möller E, Bonn SE, Ploner A, Wright A, Sjölander A, Bälter O, Lissner L, Bälter K
Two New Meal- and Web-Based Interactive Food Frequency Questionnaires: Validation of Energy and Macronutrient Intake
J Med Internet Res 2013;15(6):e109

URL: <http://www.jmir.org/2013/6/e109/>

doi: [10.2196/jmir.2458](https://doi.org/10.2196/jmir.2458)

PMID: [23739995](https://pubmed.ncbi.nlm.nih.gov/23739995/)

©Sara E Christensen, Elisabeth Möller, Stephanie E Bonn, Alexander Ploner, Antony Wright, Arvid Sjölander, Olle Bälter, Lauren Lissner, Katarina Bälter. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 05.06.2013. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Silence is Golden: Effect of Encouragement in Motivating the Weak Link in an Online Exercise Video Game

Brandon C Irwin¹, PhD; Deborah L Feltz², PhD; Norbert L Kerr^{3,4}, PhD

¹Department of Kinesiology, Kansas State University, Manhattan, KS, United States

²Michigan State University, Department of Kinesiology, East Lansing, MI, United States

³Michigan State University, Department of Psychology, East Lansing, MI, United States

⁴University of Kent, Canterbury, United Kingdom

Corresponding Author:

Brandon C Irwin, PhD

Department of Kinesiology

Kansas State University

1A Natatorium

Manhattan, KS, 66506

United States

Phone: 1 785 532 0660

Fax: 1 785 532 6486

Email: bcirwin@k-state.edu

Abstract

Background: Despite the physical and mental health benefits, few adults meet US Department of Health and Human Services physical activity guidelines for exercise frequency, intensity, and duration. One strategy that may increase physical activity duration is exercising with an Internet partner (ie, someone who is virtually present, as in video chat). Internet partners help people overcome many barriers associated with face-to-face exercise groups (eg, time, coordinating schedules, social physique anxiety). Past research examining individual performance in groups suggests that an increase in effort occurs when performing a task conjunctively, ie, when a participant is (1) less capable than fellow group members, and (2) participants efforts are particularly indispensable for group success (ie, where the group's potential productivity is equal to the productivity of its least capable member). This boost in effort is more commonly known as the Köhler effect, named after the German psychologist who first observed the effect. While encouragement between group members is common practice in face-to-face group exercise, the effect of encouragement between partners exercising conjunctively across the Internet is unknown.

Objective: To examine the impact of exercising alone, compared to exercising conjunctively with an Internet partner, both with and without encouragement, on exercise persistence (primary outcomes) and secondary psychosocial outcomes (self-efficacy, enjoyment, exercise intention).

Methods: Participants were recruited online and face-to-face from the campus of Michigan State University. With the assistance of the experimenter, participants (n=115) played an exercise video game in a laboratory, performing a series of five abdominal plank exercises where they were asked to hold the plank for as long as possible (Time 1). They were then randomized to a condition (Individual, Partner-without-encouragement, or Partner-with-encouragement), where they performed the exercises again (Time 2). The impact of condition on the primary outcome measures and secondary outcome measures were evaluated using a 2 (Gender) x 3 (Condition) ANOVA on change scores (Time 2-Time 1).

Results: Those who exercised in online teams (n=80) exercised significantly longer (time=78.8s, $P<.001$) than those who worked individually (n=35). However, exercise duration was shorter when one's more capable partner gave verbal encouragement (n=55) than when s/he did not (n=25) (a mean difference of 31.14s). These increases in effort were not accompanied by altered task self-efficacy, enjoyment of the task, or intention to exercise in the future.

Conclusions: Exercising conjunctively with an Internet partner can boost one's duration of exercise. However, encouragement from the stronger to the weaker member can mitigate these gains, especially if one perceives such comments being directed at someone other than themselves. To boost exercise duration, Internet-based physical activity interventions involving group interaction should make relative abilities of participants known and communication clear.

(*J Med Internet Res* 2013;15(6):e104) doi:[10.2196/jmir.2551](https://doi.org/10.2196/jmir.2551)

KEYWORDS

active video games; physical activity; exercise; Köhler effect; partner exercise; social influence; group dynamics; group exercise; virtual partner; intensity; duration; online video games

Introduction

Despite the links between regular physical activity and positive health benefits [1-7], less than 5% of US adults achieve recommended levels of intensity and duration [1-8]. Recent strategies to promote physical activity have harnessed the Internet, a medium through which the delivery of interventions can be highly cost-effective, with the potential to reach and impact a wide audience [9]. However, Internet-delivered interventions typically produce only small effect sizes [10] and, thus, may benefit from supplemental strategies to increase the intensity and duration of physical activity.

Recent research has highlighted the influential role of social factors in physical activity behavior, including exercising with a partner [9,11]. When exercising under the right conditions, exercise partners have been shown to increase the intensity and duration of exercise by up to 208% [12,13]. Importantly, exercising with a partner affords the opportunity to encourage one another, which may further increase the duration of exercise. However, few, if any, Internet-delivered interventions create opportunities for exercising in real-time with other exercise participants.

The purpose of the current study was to test the efficacy of an Internet partner in increasing exercise duration. Specifically, we tested whether being the “weak link” on a team with an Internet partner could motivate one to exercise longer than when exercising alone. Further, we tested whether encouragement from the partner could motivate one to exercise longer still.

Basic behavioral research in social psychology suggests that the presence of a superior partner may motivate one to exercise longer than they would when exercising alone and further still if one is on a team with that partner and is the team’s “weak link” [14]. This motivation-boosting phenomenon among the team’s weak link has been coined the Köhler effect, named after the German industrial psychologist who first observed the effect in his seminal work [15,16]. In light of this evidence, we chose to make participants the weak link between them and the Internet partner and hypothesized that this would lead to exercising longer than when exercising alone. Although no previous studies on the Köhler effect have examined the influence of encouragement on the duration of exercise in the weak link, we reasonably hypothesized that encouragement would further boost exercise duration.

Method

Design

The study used a randomized trial in a 3 (condition: individual control, partner with-encouragement, partner-without-encouragement) x 2 (performance block: Block 1 & Block 2) factorial design. Eligible participants were students who had no physical injuries that would prevent or obstruct their performance during an isometric plank exercise.

Setting

The study was conducted in a laboratory on the campus of Michigan State University in the Departments of Kinesiology and Psychology and was approved by the Institutional Review Board. Most of the data for two of the conditions (ie, 26 participants in the individual condition and 47 in the partner-without-encouragement condition) were collected as part of a study [17] completed about the time we decided to examine the effect of verbal encouragement (August 2009 to May 2010). In a new wave of data collection (August-December, 2010), we not only collected data for the partner-with-encouragement condition, but also some additional data in the two other conditions. By contrasting the latter with those collected for the aforementioned study [17], we could probe for possible history or cohort effects. We did not expect any systematic differences between the two waves of data collection because the lab settings, participant populations, and procedures for both data collection periods were identical, except for the introduction of the manipulation of verbal encouragement.

Exercise Task

The task for this study was an abdominal plank exercise. These are body-weight resistance exercises where participants suspend their own body weight using their abdominal muscles. These exercises are also isometric in nature, require very little coordination, and are highly effort-based. Each exercise targeted the abdominal muscles, but there were slight differences between each (eg, holding a push-up position on one’s forearms vs on each side; a detailed description of all exercises is provided elsewhere [12]).

The exercises were performed as part of an exercise video game designed for the PlayStation 2 (PS2) gaming system as used in a previous study [12]. The software used was EyeToy: Kinetic, a game that offers a variety of fitness activities (eg, yoga, strengthening exercises, combat exercises). This particular software operates in conjunction with an additional accessory called the Eye Toy, designed specifically for the PS2 system. The Eye Toy is a small camera that connects to the PS2 system via a USB cable and allows images of the user to be displayed on the TV monitor and interact with objects in a virtual environment supported by the software.

Procedure

A detailed description of experimental procedures is provided elsewhere [12]. In the current study, we simply describe the basic procedure and note how the encouragement manipulation was achieved.

After arriving at the laboratory, participant consent was obtained and all were ensured of the confidentiality and voluntary nature of their participation. Participants then watched a brief instructional video from the PS2-Eye Toy: Kinetic software in which a virtual trainer demonstrated the five exercises. A baseline measure of self-efficacy was then taken using an online

questionnaire completed in the laboratory. Participants were asked throughout the session if they understood all instructions and, if not, the experimenter addressed his/her questions. All participants then performed the first block of exercises, holding each of the five exercises for as long as they could and with 30s rest periods between each exercise. Immediately after each exercise, the participant announced his/her perceived exertion. All participants were given veridical feedback on their performance (ie, the average of the number of seconds they held each exercise).

The condition manipulation was introduced at this point. Participants in the individual control conditions simply rested for 10 minutes. Participants in the partner conditions were told that another participant was being run simultaneously at another lab on campus and that the 2 participants would be able to see one another over an Internet video connection during future trials. The participants then met briefly with that other same-sex participant in a controlled Skype-like interaction (we will refer to that other participant hereafter as “the partner”). In reality, the partner was an experimental confederate whose side of the interaction was prerecorded. After the interaction, participants were also given bogus feedback on how well the partner had done on the first trial. That feedback score was manipulated to be 1.4 times the participant’s own actual performance. This discrepancy was chosen based on previous research that suggests that this moderate discrepancy leads to the greatest increases in exercise duration (ie, the greatest Köhler effect) [18,19].

Participants were told that they and their partners would be a 2-person exercise team. In the encouragement condition, both teammates were told that they would have the opportunity to communicate with each other during the exercises but that, due to technical problems, this capability would be provided only for the partner. Thus, ostensibly both could speak, but only the partner would be heard by the subject. No mention was made of any audio link between partners in the no-encouragement condition. For both partner conditions, it was further explained that they were working towards a team score, where the team score would be the persistence score of the first teammate to quit an exercise (ie, as soon as either partner quit, the exercise was over). This task structure is more commonly known as a conjunctive task—when the group’s performance is defined by the performance of the least-capable member (ie, the “weak link”). Following these instructions, all participants responded again to the self-efficacy measure.

Block 2 then commenced. In the individual control condition, the participants could only see him/herself on the screen, as during Block 1. In the partner conditions, the participant could see the partner’s image (which was actually prerecorded) before and during the exercise; the participant believed that the partner could likewise see his/her (the participant’s) image. The images available to the participant suggested that s/he was always the first to quit each exercise. The video link was allegedly frozen as soon as either teammate quit an exercise and until just before the start of the next exercise; hence, the participant knew only that his/her partner had been able to persist longer, but not just how much longer. In the encouragement condition, a prerecorded series of phrases of encouragement was played through a set of computer speakers controlled by the

experimenter. The phrases were audible approximately every 15s (± 3 s) and followed the following fixed progression: “you can do it”, “you got this”, “keep it going”, “you’re doing good”, “stay strong here”, “give it your best”. After Block 2 was over, the participant completed a series of questionnaires (self-efficacy, intention to exercise, enjoyment of physical activity, and manipulation checks). S/he was then debriefed, thanked, and excused.

Measures

Duration of Exercise

Duration of exercise was measured via the total number of seconds that the exercise was held. Block scores were calculated by taking the summed total of the five exercises within each trial.

Self-Efficacy

Task self-efficacy (SE) was measured with a scale developed specifically for this program of research using an online questionnaire. The measures contained five items, each corresponding to one of the five exercises within each trial. All items were preceded by the stem “What is the number of seconds that you are completely confident you can hold” followed by “The first exercise”, “the second exercise”, and so on for each of the five exercises. Respondents wrote in the number of seconds in a blank box following each item. The questionnaire was administered at three time points: once before Block 1 (after the participant had watched a brief instructional video demonstrating the exercises), a second time after Block 1 but before performing the five exercises at Block 2, and a third time after Block 2. A total SE score for each trial was calculated by taking the sum of the five items within each trial.

Ratings of Perceived Exertion

Ratings of perceived exertion (RPE) was measured using the 6-20 version of the Borg RPE scale [20]. The scale ranges from 6-20 where 6 means “no exertion at all” and 20 means “maximal exertion”. Participants were asked to verbally rate their exertion at the end of each exercise, with particular reference to their perceived exertion at the moment right before the end of the exercise.

Task Enjoyment

Task enjoyment was measured using a short 8-item version of the Physical Activity Enjoyment Scale (PAES) using an online questionnaire [21]. Each item was rated on a 7-point bipolar scale beginning with the stem “Please rate how you feel at the moment about the physical activity you have been doing according to the following scales” (eg, 1=“I loved it”; 7=“I hated it”). Previous studies have shown high correlations with the longer, 18-item scale ($r=.94$) [22] and strong reliability (Cronbach alpha=.91) [23].

Intention to Exercise

Adapted from another measure [24], intention was assessed with a single item, “I intend to exercise tomorrow for at least 30 minutes” on a scale of -3 (“Not at all true for me”) to +3 (“Completely true for me”).

Postexperimental Questionnaire

Besides some questions checking participants' understanding of the instructions and procedures, there were questions probing their perceived task ability, a rating of task difficulty, and a rating of effort expended on the task, each made on 8-point scales. Participants in the partner conditions were also asked to rate their partner's relative ability on a 9-point scale (where 1="I am much more capable" and 9="my partner is much more capable").

Plan of Analyses and Participants

Following the methods of analyses done in earlier studies, exercise duration was taken as the sum of the time each participant held the five exercises within each block, producing a Block 1 and Block 2 score. There are, of course, individual differences in fitness, intrinsic task interest, and strength that we wished to control for. This can be done in different ways. In many prior studies [14,25], participants' Block 1 performance has been used as a baseline and difference scores (ie, Block 2-Block 1) were the primary dependent variable, one that expressed each participant's Block 2 score relative to his/her Block 1 score (the latter baseline score reflected individual differences in fitness and strength). An alternative (less vulnerable to certain problems that can arise from the use of difference scores, eg, [26]) is to use Block 1 scores as a covariate in the analysis of Block 2 scores. Here, we present the results using the former difference-score method because the mean values presented for such an analysis are more directly understood and interpreted than means adjusted for a covariate. But the reader should note that an identical pattern of results is obtained with either method.

The analyses of the exercise duration data was to proceed in two stages. The first was to check whether there were any history or cohort effects attributable to the interval between the two data collection waves. It employed a 2 (Condition: Individual vs partner-without-encouragement) x 2 (Sex) x 2 (Data collection wave: Early vs Late) ANOVA on exercise duration difference scores (ie, Trial 2 duration-Trial 1). Although overall sex differences in the magnitude of the Köhler effect are rare, some interesting sex effects have been reported, eg, [25]; hence, participant sex was included as an experimental factor. This stage 1 analysis would permit both confirming with the larger dataset that the Köhler effect originally reported in Kerr et al [17] and checking whether its magnitude differed between the participants drawn from Kerr et al [17] and those newly recruited in the same experimental conditions for the present study. The second stage was designed to examine the primary question addressed by this paper—would verbal

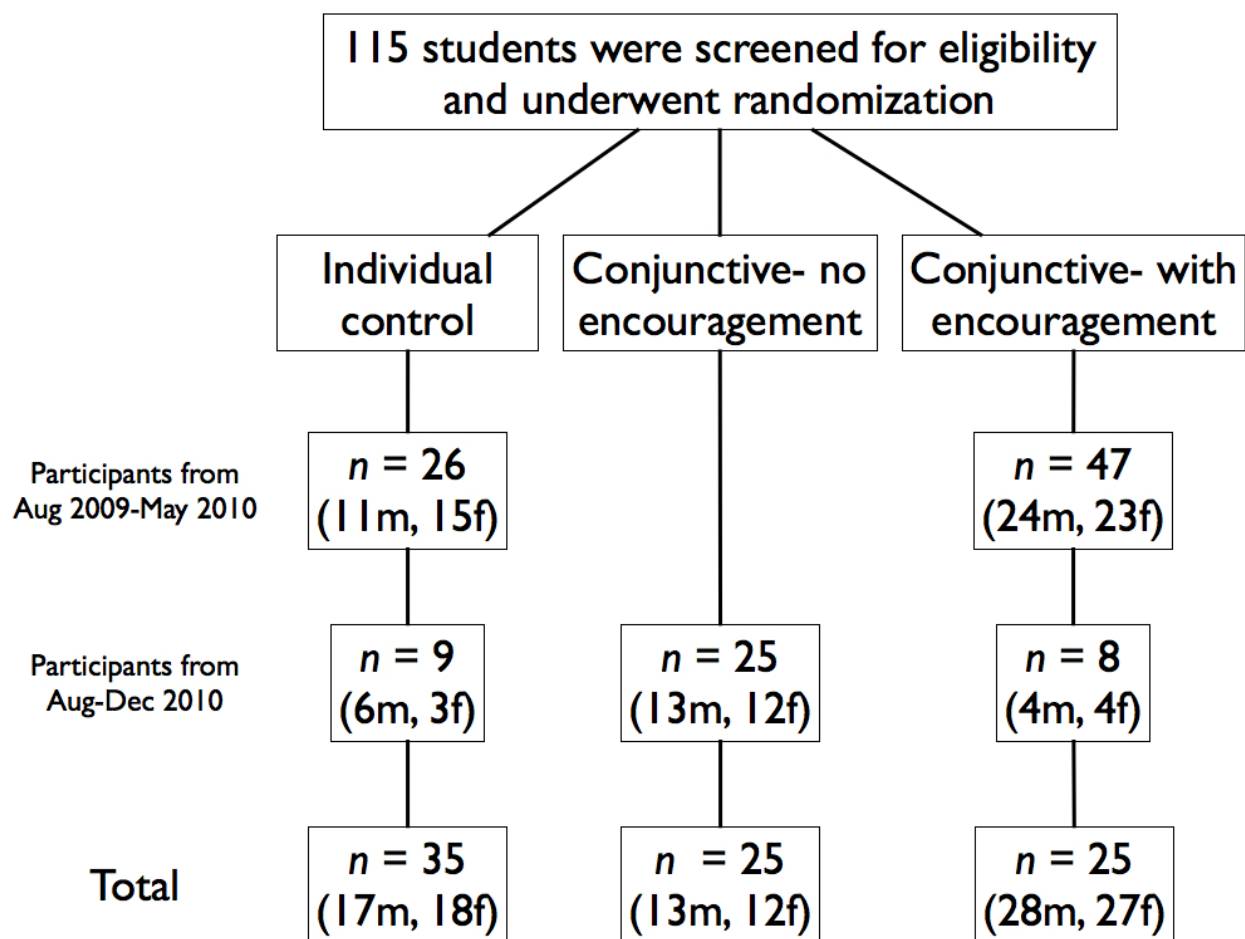
encouragement to the weaker partner alter the Köhler effect? If there were no data-wave moderation effects, as anticipated, we planned to drop the data-wave factor and add the no-encouragement condition in a 3 (Condition: Individuals, Partner-without-encouragement, Partner-with-encouragement) x 2 (Sex) ANOVA on exercise duration difference scores.

Although our primary focus was on exercise duration, we also examined several other variables, primarily to determine if encouragement altered any of the effects previously observed [12,17] when no encouragement was provided. For those variables collected at the end of the study (ie, task effort, intent to exercise, own task ability), the analyses would employ 3 (Condition) x 2 (Sex) analyses of variance; since the individual participants had no partner, this became a 2 (Condition: Encouragement vs No-encouragement) x 2 (Sex) ANOVA for ratings of own ability relative to one's partner. For those variables (ie, perceived effort, self-efficacy) collected during the exercise trials, a within-Ss Block factor was added to the ANOVA. Finally, a covariate (ie, the pre-exercise estimate of task self-efficacy) was included in the analysis of self-efficacy.

Results

Students were recruited from introductory psychology (online) and kinesiology courses (online and face-to-face) at a large Midwestern university and were given course credit for their participation. Students were recruited based on their interest in getting a good workout and were told that they would be playing an exercise video game and performing abdominal plank exercises for as long as they felt comfortable. The final total sample consisted of 115 participants (58 male, 57 female) college students (mean age 20.31, SD 3.26; see Figure 1). No participants dropped out of the study before completing their session.

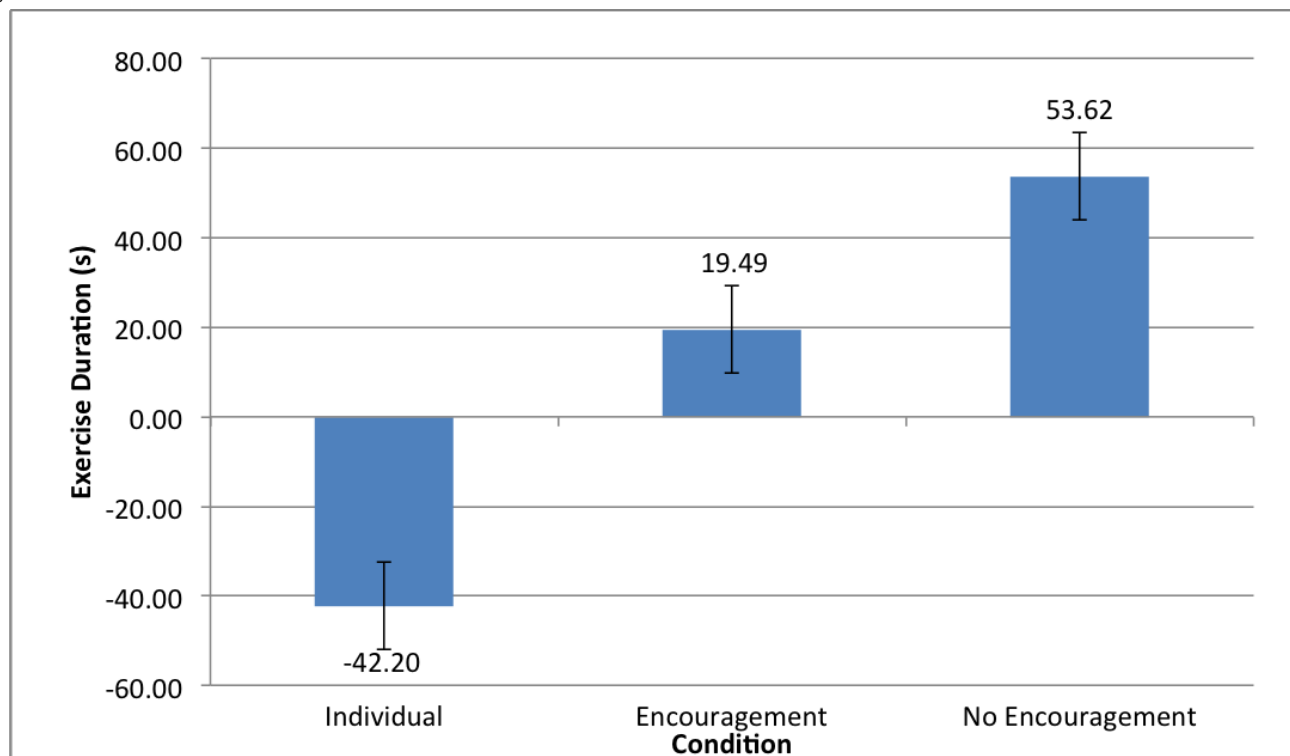
As noted earlier, the participant pool and methods of recruitment were identical for the two waves of data collection, and participants were randomly assigned to conditions (single blind) within each wave using a randomization function on Microsoft Excel generated by one of the primary investigators. Hence, we expected rough equivalence in participant characteristics across waves and conditions. This could be checked for the two characteristics collected in the study—age and year in school. As expected, when the five combinations of wave and condition were compared in analyses of variance, there was no hint of between wave/condition differences (all P s>.15). Overall, the average participant was a sophomore/junior (mean of 2.45 where 1=1st year, 2=2nd year, etc) aged 20.3 years (SD 3.3).

Figure 1. Participant flow.

Exercise Duration

The initial stage 1 analyses looked for possible history or cohort effects attributable to the interval between the two data collection waves. A boost in exercise duration (ie, the Köhler effect) was evident (Condition main effect $F_{1,79}=39.53$, $P<.001$, partial eta squared=.33); whereas due to fatigue, Individual condition participants persisted 39.4s less at Block 2 than at Block 1, and participants in the Partner-without-encouragement condition persisted 61.0s longer, a Köhler effect of 100.4s. More importantly, there were no significant effects involving the wave factor. Thus, the magnitude of the Köhler effect in the absence of partner encouragement was comparable within each wave of data collection. Hence, the data for these two waves were combined in all subsequent analyses.

In the primary stage two analyses, the data for the Partner-with-encouragement condition was included to see whether and how adding verbal encouragement might affect the duration-boosting effect. The 3 (Condition: Individuals, Partner-without-encouragement, Partner-with-encouragement) \times 2 (Sex) ANOVA on difference scores resulted in only one significant effect: the Condition main effect, $F_{2,105}=28.79$, $P<.001$, partial eta squared=.35. Again (see Figure 2), the Individuals showed a decline in exercise duration across blocks (estimated marginal mean of -42.2s, SD 54.2), whereas partners exercised longer in the second block, whether there was (19.49s, SD 59.30s) or was not encouragement (53.63s, SD 58.44). Post hoc Dunette tests indicated that the difference between Individuals and Partners was significant ($P<.001$) in both partner conditions. A Newman-Keuls post hoc test indicated that this effect was significantly smaller (34.1s) in the Partner-with-encouragement condition.

Figure 2. Condition Block 2 - Block 1 exercise duration means.

Ancillary Analyses

Exercise Self-Efficacy

PostBlock efficacy judgments (sum of number of seconds participants estimated they could persist at all 5 exercises) were examined in a 2 (Block: postBlock 1, postBlock 2) x 3 (condition) x 2 (sex) ANCOVA, which used the preBlock 1 SE score as a covariate (ie, the participant's estimated persistence at the five exercises collected prior to the first block of exercises). The latter should reflect and control for chronic individual differences in perceived self-efficacy at such exercises. As in previous studies, there was a Block main effect, $F_{1,98}=13.74$, $P<.001$, partial eta squared=.12. Participants were less sanguine about their prospects for persisting after Block 2 (adjusted mean=149.8s; SD 83.53) than after Block 1 (adjusted mean=188.4s; SD 104.1). Although condition affected actual persistence, it showed no effects on participant's expectations of what they could do in the future.

Subjective Effort

Two variables were relevant here—ratings of perceived exertion (RPE), averaged across exercises within blocks (for Block 1 alpha=.93, for Block 2 alpha=.96), and the postexperimental rating of effort expended at the task. A 2 (sex) x 3 (work condition) x 2 (block) analysis of the RPE data found that participants reported greater exertion at Block 2 (14.75; SD 1.97) than Block 1 (13.88, SD 1.89; $F_{1,104}=74.34$, $P<.001$, partial eta squared=.42), as one might expect given the fatiguing nature of the task. This effect was also stronger in the partner conditions (block difference mean of 1.33 in the encouragement condition and .93 in the no-encouragement condition; these did not differ significantly, ie, $P=.11$) than in the individual controls (.36; Block x Condition $F_{2,104}=6.76$, $P<.01$, partial eta

squared=.12), in line with the objective persistence results. A 3 (condition) x 2 (sex) ANOVA on ratings of how much effort had been expended, collected at the end of the experiment, produced no significant results.

Task Evaluation

An overall task enjoyment measure was computed based on the 8 items of the PAES scale (alpha=.86). The grand mean (4.68, SD .97, on the 8-point scale) was not significantly different from the scale midpoint (4.5), suggesting that participants were, at worst, indifferent toward the exercise task. There were no significant effects in a 3 (condition) x 2 (sex) ANOVA of this measure—working harder at the task with a partner did not appear to undercut participants' enjoyment of it. Participants' postexperimental rating of the difficulty of the task suggested that they viewed it as moderately difficult (grand mean 4.90, SD 1.65, significantly above the scale midpoint, $P<.05$). There was also a significant condition main effect on this measure, $F_{2,100}=3.63$, $P<.05$, partial eta squared=.07). Post hoc Scheffé contrasts indicated that the task was rated as significantly ($P<.05$) more difficult in the no-encouragement condition (5.27)—where actual effort had been maximal—than in the individual condition (4.34), with the encouragement condition falling in between (5.10) and differing from neither of the remaining conditions.

Intention to Exercise

Overall, at the end of the experimental session, participants expressed a positive intent to exercise for at least 30 minutes the following day. The grand mean was 1.61 (SD 1.70) on the 7-point scale anchored by -3 ("Not at all true for me") to +3 ("Completely true for me"); this was significantly ($P<.05$) above the midpoint of the scale. However, there were no significant effects in a 3 (condition) x 2 (sex) ANOVA of these intentions.

Perceptions of Task Ability

In the final questionnaire, participants were asked to rate their own ability (on an 8-point scale) and in the partner conditions, to rate their partner's relative ability (on a 9-point scale where 0="Not applicable. I didn't have a partner", 1="I was much more capable" and 9="My partner was much more capable"). A 3 (condition) x 2 (sex) ANOVA of own ability ratings produced a significant condition effect, $F_{2,102}=6.36$, $P<.01$, partial eta squared=.11. After receiving consistent feedback indicating that they were inferior to their partner, participants in the partner conditions felt less capable (encouragement=5.26, no-encouragement=5.16) than the individuals (6.52) who had no partner. A Newman-Keuls post hoc test indicated that the two partner conditions did not differ significantly, so receiving encouragement from a partner did not alter participants' sense of their own task ability.

However, the encouragement manipulation did affect participants' perceptions of their partner's ability. A 2 (encouragement vs no-encouragement) x 2 (sex) ANOVA of partner relative ability ratings revealed a strong condition effect, $F_{1,71}=10.18$, $P<.01$, partial eta squared=.13. Participants in the no-encouragement condition reported that their partner was significantly more capable (7.15) than those in the encouragement condition (5.60); the latter mean did not differ significantly from the scale midpoint of 5.0, a rating that signified equal capability between teammates. This condition effect was also qualified by participant sex, sex x condition $F_{1,71}=6.98$, $P<.02$, partial eta squared=.09; it was considerably stronger for males (condition difference of 2.87 for males vs .27 for females). Overall, females rated their partner's ability higher (7.14) than males did (5.60), with sex main effect $F_{1,71}=9.72$, $P<.01$, partial eta squared=.12.

Discussion

The purpose of this study was to test the efficacy of a superior Internet partner (with and without encouragement) as a strategy for increasing the duration of exercise. Consistent with previous studies [12,13], after controlling for individual differences in strength, participants who simply exercised with a partner as the weak link persisted longer (on average, 78.8s) than those who exercised by themselves, a gain of 33.6%. This is a considerable gain for those performing a strenuous, isometric exercise where the goal is to increase strength and for those striving to meet national physical activity recommendations [27]. Furthermore, and although one might reasonably expect encouragement to boost effort at the task, we found that such encouragement mitigated, but did not eliminate, the effort-boosting effect of being the weak link. These findings suggest that being the weak link with a superior Internet partner may be a useful strategy for increasing exercise duration, but that at least some forms of encouragement from a superior partner may only be minimally motivating compared to exercising alone and, compared to exercising with a moderately-superior partner, may actually be de-motivating.

The fact that encouragement attenuated the duration-boosting effect of being the weak link could be explained by a few

competing explanations. First, at the outset of the study, we reasoned that receiving positive encouragement could bolster one's self-efficacy [28] and/or perceived ability and that this increase could have one of two effects on exercise duration. First, a boost in self-efficacy could lead to increases in effort, as has been shown in a number of performance-based studies [29,30]. However, we found no boost in self-efficacy in either partner condition. Conversely, we reasoned that efficacy-boosting feedback could also undermine effort, either by being so inconsistent with actual relative performance that one gives up or by leading to an overly confident belief in one's capabilities and thereby undermining the desire to compare favorably with one's partner [31]. Indeed, we observed significant differences in effort between the two partner conditions. However, no such increases in self-efficacy or perceived ability were found between the two partner conditions. Thus, changes in self-efficacy/perceived ability did not likely account for the attenuation of the duration-boosting effect in the encouragement condition. That encouragement did not have an effect on one's self-efficacy does not come as a surprise, as feedback regarding one's own performance (ie, mastery experiences) is typically more influential than verbal encouragement in affecting one's judgments of his/her abilities [28]. Thus, it is likely that feedback indicating the partner's superiority on Block 1 and the constant and veridical performance feedback indicating that subjects were being outperformed by their partner on Block 2 overrode or diluted any potential efficacy-boosting effect of positive encouragement offered by the partner.

Second, it was also possible that receiving encouragement from a superior other might be interpreted as condescending or patronizing [32], which could result in negative judgments of one's partner or aversion to the task. Unfortunately, in the current study, we did not take any explicit measures of the subjects' interpretations of the partner's statements. However, subjects who received encouragement enjoyed the task equally as much as those who did not receive encouragement. Thus, the differences in effort between the two conditions was likely not due to the subjects' interpretations of the encouragement as unfavorable, but clearly future studies should measure judgments of one's partner to further and more explicitly explore this possibility.

One last possibility was that encouragement from a superior partner would be interpreted not as teammate support, but rather a method of self-encouragement. This could be expected especially if the intended target of the partner's message was somewhat ambiguous, as could be the case in the current study (eg, "You can do it"; where "you" could be taken to be directed at the partner or one's self). Interpreting the message as self-encouragement might suggest to the participants that the supposedly superior partner was in fact struggling with the task, thereby creating doubt in the degree of the partner's superiority. When one believes that his/her partner's ability is equal to or only slightly greater than one's own, there may be little to be learned or gained by trying to match the partner's performance [18]. In the present study, despite all the performance feedback suggesting that their partner was superior in ability, by the end of the experiment, those in the encouragement condition

perceived themselves as comparable to their partner in ability. This strongly suggests that participants in this condition took their partner's verbal comments (eg, "you can do it") not as encouragement directed from a more to a less capable partner, but rather as signs that the partner was him/herself struggling with the task and was engaging in self-encouragement.

There are a number of implications of this study for physical activity promotion. First, while many studies have shown the effectiveness of group-based approaches to physical activity promotion [33], few have systematically looked at the basic underlying mechanisms behind this strategy. In this study, we do so and identify with whom (a moderately superior partner) and under what conditions (when one is the "weak link"; exercising with an Internet partner; no encouragement) exercising in groups can boost one's effort and ultimately the duration of exercise. Second, as more physical activity promotion strategies are moving to digital modes of delivery, this study contributes to a body of basic research that provides fertile ground on which to base algorithms for electronically mediated, group-based physical activity interventions (eg, customized computer-generated partners, match-making applications). The current study suggests that when communication is part of such a partner- or group-based intervention, designers should facilitate clear and unambiguous communication between partners (eg, constrain communication to have the desired effort-boosting effect, perhaps by forcing partners to select from a fixed set of text messages).

Of course, this study has its limitations. The study was conducted in a highly controlled laboratory setting, and it may be premature to suggest that these findings generalize to free-living conditions and other populations who are more physically inactive. Second, subjects were recruited through both face-to-face and web-based systems, and because we did

not code participants for how they were recruited, we were unable to control for any individual differences between recruitment strategy (eg, being recruited via Internet may have selected for people who are more motivated to work out alone). Third, although we found that being the weak link can motivate participants to exercise longer, it is not clear if the same strategy could have the same positive impact on other dimensions of physical activity behavior (eg, frequency and intensity of exercise). Last, we tested participants in only one session of exercise, and repeatedly being the "weak link" over several sessions may actually be de-motivating [34]. We should note, however, that some of the potential limitations have been addressed in other studies. For instance, researchers have recently observed the effort-boosting effect of a superior partner in other physical activity tasks and conditions, such as duration of aerobic exercise [13] and performance in swimming relay competitions, respectively [35]. Further, there is evidence to suggest that being the weak link over several exercise sessions can actually strengthen the effect, leading to increasingly longer bouts of exercise with each successive bout [13].

Conclusion

The current study corroborates a growing body of research, which shows that exercising with a moderately superior Internet partner as the weak link can boost effort and lead to longer bouts of exercise [12,13], whether or not partners communicate with one another. However, such effort-boosting effects can be mitigated when superior partners try to encourage weaker group members, especially if this encouragement undermines the weaker member's perception of the superior ability of his/her partner. Future research should examine the effects of a wider range of messages, exercise tasks, and conditions to help inform the design of group-based, electronically mediated physical activity interventions.

Acknowledgments

The authors would like to acknowledge Nate Kenniston, Nick Thompson, Sam Forlenza, and Kaitlynn Sedabres for their help in conducting the study and David Dzewaltowski for his comments on an earlier draft of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V1.6.2 [36].

[PDF File (Adobe PDF File), 994KB - [jmir_v15i6e104_app1.pdf](http://www.jmir.org/2013/6/e104/app1.pdf)]

References

1. Lee IM, Skerrett PJ. Physical activity and all-cause mortality: what is the dose-response relation? *Med Sci Sports Exerc* 2001 Jun;33(6 Suppl):S459-S471. [Medline: [11427772](https://pubmed.ncbi.nlm.nih.gov/11427772/)]
2. Thompson PD, Buchner D, Pina IL, Balady GJ, Williams MA, Marcus BH, American Heart Association Council on Clinical Cardiology Subcommittee on Exercise, Rehabilitation, and Prevention, American Heart Association Council on Nutrition, Physical Activity, and Metabolism Subcommittee on Physical Activity. Exercise and physical activity in the prevention and treatment of atherosclerotic cardiovascular disease: a statement from the Council on Clinical Cardiology (Subcommittee on Exercise, Rehabilitation, and Prevention) and the Council on Nutrition, Physical Activity, and Metabolism (Subcommittee

- on Physical Activity). *Circulation* 2003 Jun 24;107(24):3109-3116 [FREE Full text] [doi: [10.1161/01.CIR.0000075572.40158.77](https://doi.org/10.1161/01.CIR.0000075572.40158.77)] [Medline: [12821592](https://pubmed.ncbi.nlm.nih.gov/12821592/)]
3. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 2002 Feb 7;346(6):393-403 [FREE Full text] [doi: [10.1056/NEJMoa012512](https://doi.org/10.1056/NEJMoa012512)] [Medline: [11832527](https://pubmed.ncbi.nlm.nih.gov/11832527/)]
 4. Vuori IM. Health benefits of physical activity with special reference to interaction with diet. *Public Health Nutr* 2001 Apr;4(2B):517-528. [Medline: [11683546](https://pubmed.ncbi.nlm.nih.gov/11683546/)]
 5. Brosse AL, Sheets ES, Lett HS, Blumenthal JA. Exercise and the treatment of clinical depression in adults: recent findings and future directions. *Sports Med* 2002;32(12):741-760. [Medline: [12238939](https://pubmed.ncbi.nlm.nih.gov/12238939/)]
 6. Pronk NP, Wing RR. Physical activity and long-term maintenance of weight loss. *Obes Res* 1994 Nov;2(6):587-599. [Medline: [16358405](https://pubmed.ncbi.nlm.nih.gov/16358405/)]
 7. Monninkhof EM, Elias SG, Vlems FA, van der Tweel I, Schuit AJ, Voskuil DW, TFPAC. Physical activity and breast cancer: a systematic review. *Epidemiology* 2007 Jan;18(1):137-157. [doi: [10.1097/01.ede.0000251167.75581.98](https://doi.org/10.1097/01.ede.0000251167.75581.98)] [Medline: [17130685](https://pubmed.ncbi.nlm.nih.gov/17130685/)]
 8. Metzger JS, Catellier DJ, Evenson KR, Treuth MS, Rosamond WD, Siega-Riz AM. Patterns of objectively measured physical activity in the United States. *Med Sci Sports Exerc* 2008 Apr;40(4):630-638. [doi: [10.1249/MSS.0b013e3181620ebc](https://doi.org/10.1249/MSS.0b013e3181620ebc)] [Medline: [18317384](https://pubmed.ncbi.nlm.nih.gov/18317384/)]
 9. Giles-Corti B, Donovan RJ. Relative influences of individual, social environmental, and physical environmental correlates of walking. *Am J Public Health* 2003 Sep;93(9):1583-1589. [Medline: [12948984](https://pubmed.ncbi.nlm.nih.gov/12948984/)]
 10. Davies CA, Spence JC, Vandelandotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. *Int J Behav Nutr Phys Act* 2012;9:52 [FREE Full text] [doi: [10.1186/1479-5868-9-52](https://doi.org/10.1186/1479-5868-9-52)] [Medline: [22546283](https://pubmed.ncbi.nlm.nih.gov/22546283/)]
 11. Carron AV, Hausenblas HA, Mack D. Social influence and exercise: A meta-analysis. *J Sport Exercise Psy* 1996;18:1-16.
 12. Feltz DL, Kerr NL, Irwin BC. Buddy up: the Köhler effect applied to health games. *J Sport Exerc Psychol* 2011 Aug;33(4):506-526. [Medline: [21808077](https://pubmed.ncbi.nlm.nih.gov/21808077/)]
 13. Irwin BC, Scorniaenchi J, Kerr NL, Eisenmann JC, Feltz DL. Aerobic exercise is promoted when individual performance affects the group: a test of the Kohler motivation gain effect. *Ann Behav Med* 2012 Oct;44(2):151-159. [doi: [10.1007/s12160-012-9367-4](https://doi.org/10.1007/s12160-012-9367-4)] [Medline: [22576339](https://pubmed.ncbi.nlm.nih.gov/22576339/)]
 14. Hertel G, Kerr NL, Messé LA. Motivation gains in performance groups: paradigmatic and theoretical developments on the Köhler effect. *J Pers Soc Psychol* 2000 Oct;79(4):580-601. [Medline: [11045740](https://pubmed.ncbi.nlm.nih.gov/11045740/)]
 15. Köhler O. Kraftleistungen bei Einzel- und Gruppenarbeit (Physical performance in individual and group situations). *Industrielle Psychotechnik* 1926;3:274-282.
 16. Köhler O. Über den Gruppenwirkungsgrad der menschlichen Körperarbeit und die Bedingung optimaler Kollektivkraftreaktion (On group efficiency of physical labor and the conditions of optimal collective performance). *Industrielle Psychotechnik* 1927;4:209-226.
 17. Kerr NL, Feltz DL, Irwin BC. To pay or not to pay? Do extrinsic incentives alter the Kohler group motivation gain? *Group Processes & Intergroup Relations* 2012 Aug 2012;16(2):257-268. [doi: [10.1177/1368430212453632](https://doi.org/10.1177/1368430212453632)]
 18. Feltz DL, Irwin B, Kerr N. Two-player partnered exergame for obesity prevention: using discrepancy in players' abilities as a strategy to motivate physical activity. *J Diabetes Sci Technol* 2012 Jul;6(4):820-827. [Medline: [22920808](https://pubmed.ncbi.nlm.nih.gov/22920808/)]
 19. Messé LA, Hertel G, Kerr NL, Lount RB, Park ES. Knowledge of partner's ability as a moderator of group motivation gains: an exploration of the Köhler discrepancy effect. *J Pers Soc Psychol* 2002 Jun;82(6):935-946. [Medline: [12051581](https://pubmed.ncbi.nlm.nih.gov/12051581/)]
 20. Borg GAV. Borg's Perceived exertion and pain scales. Champaign, IL: Human Kinetics; 1998.
 21. Kendzierski D, DeCarlo KJ. Physical Activity Enjoyment Scale: Two validation studies. *J Sport Exercise Psy* 1991;13(1):50-64.
 22. Raedeke TD. The Relationship Between Enjoyment and Affective Responses to Exercise. *Journal of Applied Sport Psychology* 2007 Feb 2007;19(1):105-115. [doi: [10.1080/10413200601113638](https://doi.org/10.1080/10413200601113638)]
 23. McArthur LH, Raedeke TD. Race and sex differences in college student physical activity correlates. *Am J Health Behav* 2009;33(1):80-90. [Medline: [18844523](https://pubmed.ncbi.nlm.nih.gov/18844523/)]
 24. Mohiyeddini C, Pauli R, Bauer S. The role of emotion in bridging the intention-behaviour gap: The case of sports participation. *Psychology of Sport and Exercise* 2009 Feb 2009;10(2):226-234. [doi: [10.1016/j.psychsport.2008.08.005](https://doi.org/10.1016/j.psychsport.2008.08.005)]
 25. Kerr NL, Messé LA, Seok DH, Sambolec EJ, Lount RB, Park ES. Psychological mechanisms underlying the Köhler motivation gain. *Pers Soc Psychol Bull* 2007 Jun;33(6):828-841. [doi: [10.1177/0146167207301020](https://doi.org/10.1177/0146167207301020)] [Medline: [17475617](https://pubmed.ncbi.nlm.nih.gov/17475617/)]
 26. Edwards JR, Parry ME. On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Acad Manag J* 1993;1577-1613 [FREE Full text]
 27. Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, Franklin BA, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Med Sci Sports Exerc* 2007 Aug;39(8):1423-1434. [doi: [10.1249/mss.0b013e3180616b27](https://doi.org/10.1249/mss.0b013e3180616b27)] [Medline: [17762377](https://pubmed.ncbi.nlm.nih.gov/17762377/)]
 28. Bandura A. Self-efficacy: the exercise of control. New York: W.H. Freeman; 1997.

29. Feltz DL, Riessinger CA. Effects of in vivo emotive imagery and performance feedback on self-efficacy and muscular endurance. *J Sport Exercise Psy* 1990;12:132-143.
30. Hutchinson JC, Sherman T, Martinovic N, Tenenbaum G. The Effect of Manipulated Self-Efficacy on Perceived and Sustained Effort. *Journal of Applied Sport Psychology* 2008 Oct 2008;20(4):457-472. [doi: [10.1080/10413200802351151](https://doi.org/10.1080/10413200802351151)]
31. Seok DH. Exploring self-efficacy as a possible moderator of the Köhler discrepancy effect. Master's thesis, Michigan State University 2004.
32. Vescio TK, Gervais SJ, Snyder M, Hoover A. Power and the creation of patronizing environments: the stereotype-based behaviors of the powerful and their effects on female performance in masculine domains. *J Pers Soc Psychol* 2005 Apr;88(4):658-672. [doi: [10.1037/0022-3514.88.4.658](https://doi.org/10.1037/0022-3514.88.4.658)] [Medline: [15796666](https://pubmed.ncbi.nlm.nih.gov/15796666/)]
33. Burke S, Carron A, Eys M, Ntoumanis N, Estabrooks P. Group versus Individual Approach? A Meta-Analysis of the Effectiveness of Interventions to promote physical activity. *Sport Exer Psychol Rev* 2006;2(1):19-35 [FREE Full text]
34. Lount RB, Kerr NL, Messé LA, Seok D, Park ES. An examination of the stability and persistence of the Köhler motivation gain effect. *Group Dynamics: Theory, Research, and Practice* 2008;12(4):279-289. [doi: [10.1037/1089-2699.12.4.279](https://doi.org/10.1037/1089-2699.12.4.279)]
35. Osborn KA, Irwin BC, Skogsberg NJ, Feltz DL. The Köhler effect: Motivation gains and losses in real sports groups. *Sport, Exercise, and Performance Psychology* 2012 Feb 2012;1(4):242-253 [FREE Full text] [doi: [10.1037/a0026887](https://doi.org/10.1037/a0026887)] [Medline: [17296997](https://pubmed.ncbi.nlm.nih.gov/17296997/)]
36. 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](https://doi.org/10.2196/jmir.1923)] [Medline: [22209829](https://pubmed.ncbi.nlm.nih.gov/22209829/)]

Abbreviations

PAES: Physical Activity Enjoyment Scale

RPE: ratings of perceived exertion

SE: self-efficacy

Edited by G Eysenbach; submitted 30.01.13; peer-reviewed by E Lyons, S Harden; comments to author 08.02.13; revised version received 01.04.13; accepted 07.04.13; published 04.06.13.

Please cite as:

Irwin BC, Feltz DL, Kerr NL

Silence is Golden: Effect of Encouragement in Motivating the Weak Link in an Online Exercise Video Game

J Med Internet Res 2013;15(6):e104

URL: <http://www.jmir.org/2013/6/e104/>

doi: [10.2196/jmir.2551](https://doi.org/10.2196/jmir.2551)

PMID: [23732514](https://pubmed.ncbi.nlm.nih.gov/23732514/)

©Brandon C Irwin, Deborah L Feltz, Norbert L Kerr. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 04.06.2013. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

The Conceptual and Practical Ethical Dilemmas of Using Health Discussion Board Posts as Research Data

Carol S Bond^{1*}, EdD, MSc, BA(Hons), RN; Osman Hassan Ahmed^{1*}, BSc (Hons), PGDip; Martin Hind¹, PhD, RN; Bronwen Thomas², PhD; Jaqui Hewitt-Taylor¹, PhD, RN

¹School of Health and Social Care, Bournemouth University, Bournemouth, United Kingdom

²Media School, Bournemouth University, Bournemouth, United Kingdom

*these authors contributed equally

Corresponding Author:

Carol S Bond, EdD, MSc, BA(Hons), RN

School of Health and Social Care

Bournemouth University

Royal London House

Christchurch Road

Bournemouth, BH1 3LT

United Kingdom

Phone: 44 1202 961748

Fax: 44 1202961748

Email: cbond@bournemouth.ac.uk

Abstract

Background: Increasing numbers of people living with a long-term health condition are putting personal health information online, including on discussion boards. Many discussion boards contain material of potential use to researchers; however, it is unclear how this information can and should be used by researchers. To date there has been no evaluation of the views of those individuals sharing health information online regarding the use of their shared information for research purposes.

Objective: To explore the views of contributors to online diabetes discussion boards with regards to if (and how) they feel their contributions to boards should be used by health researchers.

Methods: A qualitative approach was employed using online semistructured asynchronous (email) interviews. Interpretative description methodology was used to assess the interview transcripts, and quotations were extracted and anonymized to support each theme.

Results: 26 interviews were carried out. Participants agreed that forum posts are in the public domain and that aggregated information could be freely used by researchers. This was agreed to be a good way of ensuring that the view of people living with diabetes is being heard in research. There was no consensus on the need for permission to use individual information, such as quotations, with some people happy for this to be freely used and others feeling that permission is necessary.

Conclusions: Participants acknowledged the dichotomy of having placed information into the public domain in an unrestricted way, with some interviewees also wanting to retain control of its use. The Internet is a new research location, and rather than trying to apply traditional ethical norms to this new genre, a new *modus operandi* is required. The authors propose introducing new norms for presenting research carried out with online discussion boards.

(*J Med Internet Res* 2013;15(6):e112) doi:[10.2196/jmir.2435](https://doi.org/10.2196/jmir.2435)

KEYWORDS

social media; research ethics; health; research design

Introduction

The widespread use of the Internet throughout modern society has led to individuals being able to connect with others with the same health condition as them on a global scale [1]. Without

rehearsing a comprehensive history of the Internet, it has progressed from a relatively closed community of people who needed some technical knowledge or skills to use shared spaces such as Bulletin Boards and MUDs (multi-user domains), to an ubiquitous user-friendly network with social spaces

characterized by easy-to-use discussion boards and social media sites. This permeation has significantly affected the nature and extent of the user base, who now need little knowledge and few skills. One of the first attempts to measure Internet and Web use was carried out in 1996 [2], and although there was debate about the accuracy of extrapolating the survey data to the population of North America, final figures were agreed to be approximately 19.4 million Internet users and 14.6 million Web users. This represented approximately 7% of the population. By 2012, this had risen to around 88% of the population [3].

This change can also be seen in the health arena. Websites provide information and advice for most long-term health conditions [4-6], and many of these contain discussion boards where people can share experiences and support each other [7-9]. The evolution of the Internet from a limited, technical resource to today's dynamic "Web 2.0" where people are able to share information means that increasing numbers of people living with a long-term condition are now putting personal health information into the public domain, including on discussion boards [10].

Online discussion boards contain material that is potentially of use and of interest to researchers, especially to researchers interested in the experiences of those living with a long-term condition. Issues such as self-management, concordance with medicine and other regimes, and interactions with health care professionals are all discussed freely in these forums, with hundreds and sometimes thousands of individuals contributing their opinions on a certain topic. However, in keeping with studies conducted on any cohort, researchers must adhere to the appropriate ethical guidelines when conducting their research [11].

One of the early attempts to establish some ground rules for ethical online research was made in 2002 by the Association of Internet Researchers (AoIR) [12]. Since this research was undertaken, these recommendations have been updated [13]. The report authors acknowledge that much has changed in the field of Internet studies over the period between the versions, including a multitude of devices and the "interweaving of online and offline activities and experiences" (p. 2). These guidelines address the challenge of trying to apply the concept of a "human subject" to an online environment but are generic in discipline terms and do not specifically address health research.

The European Commission's Information Society Technologies (IST) Programme funded the RESPECT project [14], which drew up professional and ethical guidelines for carrying out socioeconomic research. The RESPECT guidelines reinforced the methodological challenges associated with online research (identifying dangers of conducting research in this manner) but stopped short of giving specific information tailored to the needs of online communities.

In the United Kingdom, the Department of Health's Research Governance Framework for Health and Social Care has set out the expectations for those studying NHS (National Health Service) patients, their carers, and significant others [15]. The overarching expectation is that "The dignity, rights, safety and wellbeing of participants must be the primary consideration in any research study" [15]. However, this framework does not

include specific mention of how these principles are best applied in the online environment.

Sharing health information online has brought about new opportunities for researchers and health professionals, providing a repository of information that has the potential to help clinicians better understand the needs associated with specific clinical populations. It has been suggested that health researchers have been slow to seize the opportunities of online research in comparison to disciplines such as media studies [16] and that ethical issues are partly responsible for this. This has been an ongoing debate. In 2000, two researchers [17] published an exploration of the dilemmas they faced in trying to carry out research within a list serve-based community. In 2007, Whitehead [18] found that the research community was divided about the correct approach to take when considering the ethical issues, and in 2009, Holmes [19] was of the opinion that the ethical standards for Internet research were not well developed. The discourse relating to the ethical use of information shared online has centered upon the views and experiences of researchers [20-22] and has not been focused exclusively on health information.

One of the first discussions about the ethics of health research online was in 2001 [23], and although this was before Web 2.0, the challenges identified are still relevant, as is the conclusion that best practice guidelines are needed. The issue of how ethical principles can be applied to online health research has provided a challenge to researchers [24]. Much of the work undertaken to date, such as the AoIR guidelines [13], has been developed in consultation with researchers.

To date, the views of those posting health information online in relation to how they anticipate the information they post being used has not featured significantly in the debate. Eysenbach and Till [23] reviewed comments posted on health-related discussion boards in 2001, concluding that members of Internet communities do not expect the posts they make to be used by researchers. This research identified the issue of the blurring of public and private spaces when using health-related discussion boards. Since that study, there has been a rapid growth in online interactions for health-related purposes [25], with recent figures suggesting up to 80% of Internet users have searched for health information online [26]. The emergence of social networking sites has had a huge impact on how individuals communicate and share health information online. The largest social networking site, Facebook, which was launched only in 2004, currently has more than 1 billion active users [27] and its use for health-related purposes has been reported for many conditions [28-30]. The extent to which individuals across the world are now using online spaces to share health information was unimaginable when questions relating to ethical issues in using online places for research were first identified.

In order for researchers to be able to understand what constitutes the ethical use of online health information, it is important to address the lack of evaluation of the views of those posting health information online. This study seeks to bring their voices into the knowledge base by examining the views of people living with diabetes who share health information on online discussion boards. The goal of this study was to undertake semistructured

interviews with users of online diabetes discussion boards, in order to better understand their views towards how the information that they share in these forums is used by researchers.

Methods

Ethical Approval

Ethical approval for this study was granted by the Ethics Committee of the School of Health and Social Care, Bournemouth University [31]. Data collection commenced in April 2012 and concluded in May 2012.

Study Design and Participants

The study used online semistructured asynchronous (email) interviews [32]. Both CSB, the primary investigator (PI) and the project researcher, OHA, had experience in online interviewing techniques. The target number of 24 participants was set prior to recruitment, with the ability to increase this should the need arise. Deciding how many interviews to conduct for qualitative research is an inexact science and very dependent on the context [33]. Drawing on previous experience, it was correctly anticipated that saturation of data would be reached by this stage.

Identification of Participants

Figure 1 outlines how participants were identified and recruited into the study. Four active diabetes forums were identified by the PI and chosen to recruit participants into the study. Four forums were selected to ensure that any particular character or interests on a board would have undue influence on the findings. None of these forums required membership in order to read the posts; however, two of the groups did require membership in order to post on the forum. Moderators and administrators of the forum were approached prior to posting on the forums in order to gain their approval for posting for research purposes. These moderators and administrators were also asked if they were willing to participate in the study themselves. This led to participants having a range of posting experience encompassing community leaders to occasional contributors.

Inclusion/Exclusion Criteria

Participants were eligible for inclusion in the study if they were an active member on one of the forums identified in the study. Active was defined as having made at least one post. There were no exclusions other than, as the interviews were conducted in English, people who were not sufficiently fluent in written English were unable to participate.

Recruitment

CSB and OHA posted in each of the forums on different threads within the forum in order to give the recruitment posts more publicity. The recruitment posts described the study and provided our contact details. Individuals interested in

participating were invited to contact us by email for more information. Once interested individuals had made contact with the research team, they were sent a brief overview of the study along with an information sheet. Following this, if the participants were happy to participate, they were sent the first interview questions by email and the interview was started.

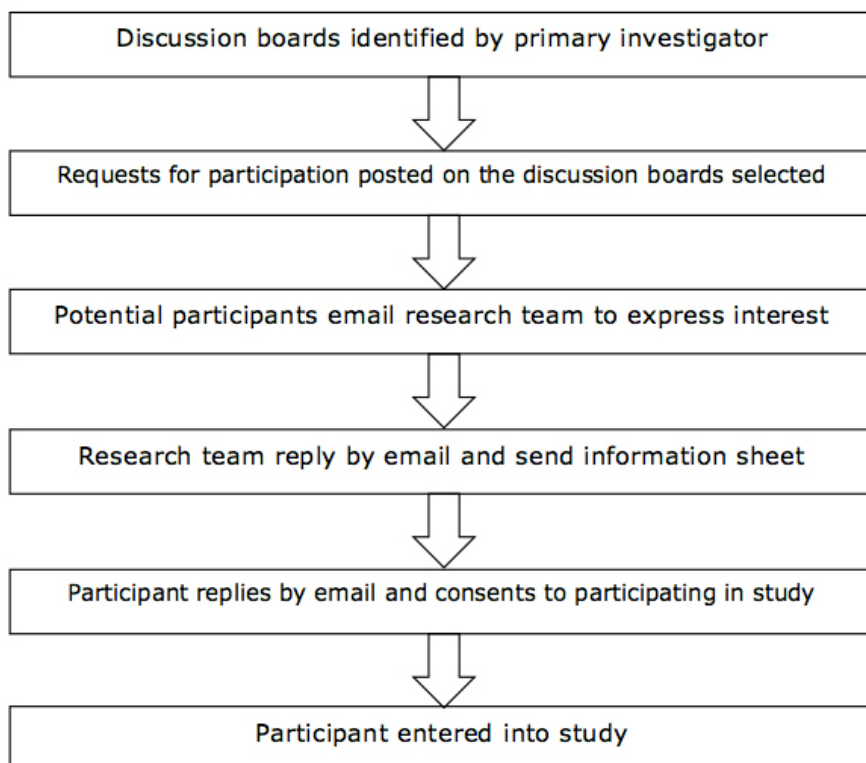
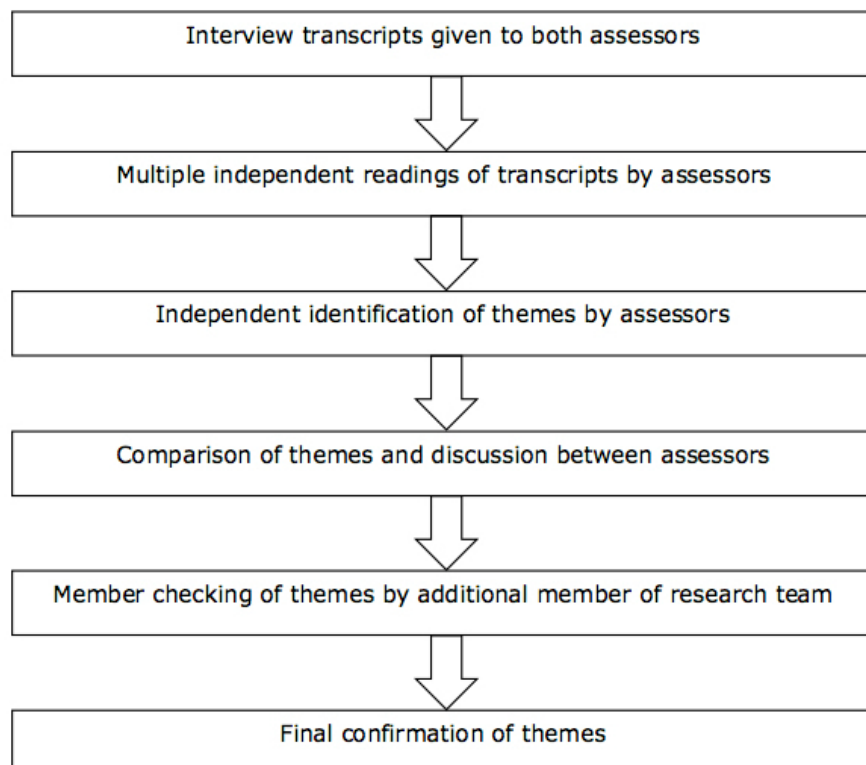
Data Collection

The data from participants were collected using asynchronous semistructured interviews [32]. The semistructured nature of these interviews meant that similarly themed information was collected from the participants, while the fact the interviews were asynchronous enabled participants to respond to questions in their own time (but within the time constraints of the research program).

Participants were sent several emails during the course of the interview with each email containing 1-2 questions. The questions (see [Multimedia Appendix 1](#)) aimed to ascertain whether participants felt it was acceptable for researchers to use information on health discussion boards, what permission should be sought prior to using this information, and whether the length of time since the post was made influenced the need to obtain permission. Once the participants had answered all of the questions, they were thanked for their contributions and their participation in the study was over.

Data Analysis

A qualitative approach using interpretative description [28] was employed to assess the responses to the semistructured interviews. An inductive approach was adopted, allowing the themes to emerge from the data rather than testing previously identified themes on the data. The rigor of this analysis was based on Guba and Lincoln's principles of credibility, confirmability and dependability [34,35]. This was achieved through prolonged engagement with sufficient depth of data; peer debriefing and analysis of materials; linking assertions, findings, and interpretations with the data and third party auditing of the data collection and analysis process. The process is set out in a flow chart (Figure 2) highlighting the stages in the process of analysis. CSB and OHA read each transcript several times independently and loosely attributed themes to the data with no consultation occurring between the assessors at this stage. Discussion was then undertaken to compare the themes identified and to resolve any areas of disagreement with theme allocation. Quotations were extracted to support themes, and each participant was assigned a number (eg, P04) when using the supporting quotations. Following this process and the identification of themes, their findings were sent to the other members of the research team (BT and MH) in order to verify the themes. In view of the lack of consensus on some issues among the study participants, a final check was carried out by someone not previously involved in the analysis (JHT) who reviewed the themes and conclusions to ensure consistency.

Figure 1. Participant recruitment and consent process.**Figure 2.** Data analysis process.

Results

Following the requests for recruitment posted on the discussion boards, 33 individuals expressed an interest in participation and were contacted by the research team. After this initial contact, 30 participants consented and responded to at least one question in the interview sequence, while four participants who consented did not complete the interview questions and did not respond to further follow-up prompts. The remaining 26 participants completed all of the semistructured interview questions and their answers were included for analysis.

It was anticipated that each of the interviews would be completed in a short space of time, and all of the interviews were completed in less than 10 days. Of the 26 completed interviews, 12 participants were identified (from their username/email correspondence) as being male, 9 as being female, while it was not possible to classify the remaining 6 participants from their username/email correspondence.

Following the analysis of the interview transcripts, several key themes emerged.

The Views of People Living With Diabetes Are Needed

There was general support for people using information posted on discussion boards in research. The need for people living with diabetes to have a voice was clearly felt by respondents: “it’s nice to see that someone takes note of what we, the diabetics have to say” (P23) and

I think that this type of information is vital to any type of research. Patients opinions are often not “heard” because of time constraints etc., these forums give an insight into what we feel we need from our HCPs and just as importantly...what we don’t need! [P26]

I have no reservations about your mining information from forums...it will provide much information about the human side of illness and how individuals singly and collectively approach and cope through sharing. Dare I say its importance cannot be understated. [P27]

There was also an altruistic view that this use should be for the “general good”: “presuming the research is for the general good and benefit of many” (P12) and

I think I’m probably happy for you to use direct quotes from my comments without permission, providing it is for research purposes and not for commercial gain. [P4]

Posts Made Are in the Public Domain

It was generally accepted by participants in this study that once information is posted on a forum, it is available to the public and thus could be accessed and used for research purposes: “Simple – it’s the Internet! Everything posted on the site is automatically in the public domain and available for anyone to read” (P15).

Some people acknowledged that the nature of online communities could create a different impression: “As a poster you really only think you’re saying things to the forum

community but of course you are saying it to a far wider audience” (P29).

Some respondents struggled with the dichotomy in their responses:

As someone who uses an Internet forum, on the one hand I’m happy for my information to be available publicly and for free to anyone who might need it, and on the other hand I wouldn’t appreciate that information contributing to something beyond “personal use” without my knowledge. Although at the same time, like any author, I have to be aware that that almost certainly will happen and I really have no control over what I’ve written once it’s out there for anyone to see. [P16]

Having acknowledged that they were putting information into the public domain, two more related themes emerged. Some people felt it could be used unreservedly, while others still felt there should be restrictions.

“The information is in the public domain, AND I have no problems with it being used” emerged as a theme: “When people post on the Internet, it is there for all to see. They should not complain if it can be harvested and used for the general good” (P12), “As they (forums) are open to the public and in the public domain why would it not be OK to use the information?” (P19), “I think information obtained from a discussion board when anonymised should be usable in your research, as it has been published on a public area” (P9), and “As the information is posted on the very public Internet, I don’t think there is a need for permission to use the posts”. (P10)

The difference between boards that are open to all to read, and those that require membership to access was mentioned by one respondent:

I feel if a forum is viewable to the public, IE you don’t have to be a member to view any of the forum threads, then you or anyone else can use any of the information you find on any forum. [P20]

“The information is in the public domain, but I’d be uncomfortable about it being freely used” also emerged. Some respondents were clearly struggling with views they felt to be contradictory, on the one hand acknowledging that their posts were in the public domain, but also feeling ongoing ownership of them:

I write a blog about my experience of diabetes and would feel very aggrieved if I found any of it quoted in a medical research paper without having been asked. I realise this is slightly contrary (since I am posting and effectively actively encouraging readership) but nevertheless it would feel like “theft” of my content. [P17]

Even though posts made on a public domain are, well, public I think it is a common courtesy to approach the poster by private message if necessary to ask their opinion. [P26]

Legally the comments are in the public domain and uncopyrighted so they are “free” for anyone to use.

Morally, I think that permission should probably be requested first to repeat comments verbatim, even rephrased one should probably state source. [P9]

Permission

Respondents were asked if they thought researchers need permission to use information in their posts. Responses grouped into two subthemes.

Permission Not Needed

Those who felt their information was in the public domain and therefore available to be used did not feel there was any need for permission to be sought:

If the forum is viewable to the public, then no permission is necessary. [P20]

If the forum is viewable to the public, then no permission is necessary. [P20]

Obviously u won't be using anyone's name so I don't think u need to obtain any permission. [P25]

When people post on the Internet, it is there for all to see. They should not complain if it can be harvested and used for the general good. [P12]

I think I'm probably happy for you to use direct quotes from my comments without permission, providing it is for research purposes and not for commercial gain. [P4]

Permission Required in Some Circumstances

Those who acknowledged the public nature of the forum but expressed reservations felt that some sort of permission for use should be obtained: "I think informed consent should be obtained from the person who posted the information" (P33) and "Using posts which were made in ignorance of them being used in the way you intend does not sit easily with me. I would hope that their use would at least be with the permission of the site administration" (P32).

There were however divergent views and no consensus over the role of site administrators in giving permission: "Absolutely not, what right do they have to share what is on the forum they admin" (P10) and "I think it would be a complete betrayal if (admin) were to give permission on behalf of the members" (P3).

The use of quotations was one area where this group people tended to agree:

If you took a quote direct from what someone has said "eg when I changed to a low carb diet I lost 2 stone and my HbA1c dropped to..." then I think you do need to gain permission from the poster. [P3]

If there were ever any intention to use or quote a specific post or case study then I would feel it would be absolutely essential to get approval from the individual before going ahead. [P17]

A different view was taken to "aggregated data": "if you're using the data in some kind of statistical analysis – and not quoting directly the posting then I'd say no permission is probably needed" (P10) and

If it's an "abstract" I don't think you need permission. For example if you discover from reading forums that many HCPs give dietary information to diabetics which is plainly making their condition worse, and that many diabetics find that eating a diet which is low in carbohydrates is more beneficial, then you could "generalise" about the situation. [P3]

If the information you collect ends up being essentially anonymous numbers...charts or that kind of thing then I don't think you need any individual's permission. If you want to use actual quotes from people that's a different matter as even if you make the quote anonymous in your research. [P29]

Provided no quotes are attributed I am OK with your extraction of any postings. [P27]

One would expect information to be depersonalised either by general aggregation or by use of pseudonyms in specific cases. [P12]

Few people would have a problem with generalised and anonymised references. [P17]

The time that has passed since the post was made was not considered to be relevant in the need to obtain permission:

A post I made two years ago, even if I've forgotten about it, is no less mine and I would be no less upset to find it had been used without my knowledge than a post I made yesterday. [P16]

Use of Citations

There were divergent views about attributing information to the poster. Some people felt that if quotations were to be used the author and/or forum should be cited:

I think in my mind it's a bit like any other written work, a book in a library for example. The text is there, for anybody to read and learn from, so in that sense it's public. On the other hand if we want to quote that author in writing a review or a paper, we have to include reference information. [P16]

If someone decided to republish my post in another forum or document, I would expect my comments to be kept in context and credited to me. [P04]

Other respondents valued anonymity over credit: "(permission isn't) As long as you don't identify the poster by more than sex, age, and type of diabetic" (P20) and "Obviously u won't be using anyone's name so I don't think u need to obtain any permission" (P25).

The challenge of using quotations and maintaining anonymity was raised by one respondent:

If you want to use actual quotes from people that's a different matter as even if you make the quote anonymous in your research it will be quite easy to find the author simply by typing in key phrases into Google which will then give links back to [the forum]. [P29]

Discussion

Principal Findings

The findings demonstrate a general agreement that the view of people living with diabetes should be heard by researchers and that the information on these forums contains information that is valuable to researchers. Participants had a range of experience including community leaders and occasional posters; it is possible that their experience of using discussion boards was a factor in how they saw the issues; however, this study did not attempt to explore this. It could be useful to carry out a quantitative study to test these findings with a large number of participants to explore if experience of using discussion board, or the Internet generally, is an influencing factor. Most participants appreciated that they had put their information into the public domain.

Concepts from mass media studies help to offer insight in order to better understand the process of sharing information online in this manner. The public sphere was initially proposed as an area where public opinion is expressed through rational debate and discussion by Habermas [36]. This view was formed before the Internet (with all of its powerful capabilities) was created and thus the extent to which this view can be applied to online spaces is not clear. Blogs have been likened [37] to the “letters to the Editor” that are common in traditional print media. The concept of “texts in the public domain” has also been discussed as being a social process of talk and text [38], with the authors concluding that where communication happens in ways other than face to face and has multiple audiences, a “pragmatic framework” for researchers to work with it needs to be developed.

Beyond the initial agreement about the public nature of their posts by our participants, two divergent subthemes emerged. One group was happy that they had put their information into the public domain and that it could therefore be reused by researchers. This approach and outlook could grow with the increased use of social media by society. Twitter for example has a “retweet” button, enabling the rapid dissemination of information. Retweeting is one of the core functions of Twitter [39], and it is considered completely acceptable to do so without seeking any prior permission. Basset [40] stated that the Internet is a cultural publication route, more akin to mass media, such as newspapers than any other medium. This “mass media” approach by researchers may be feasible for Twitter (and arguably blogs) where the author is writing to an unknown audience, although this is not supported by participant P17 from this study. Whether this approach can (or indeed should) also be applied to discussion boards is questionable, however, since participants are “talking” to other board users rather than intentionally addressing their comments to a wider, unknown audience.

The second subgroup in this study wanted to keep ownership of their words. While they were happy for aggregated data to be reused, they felt differently about quotations or the use of the information that could be traced back to them. Rather than applying the attribution norms of mass media communications, their views fit more comfortably into a traditional health research

ethics framework. Some people wanted to be able to give (or withhold) permission for the use of the information they had shared. Others wanted to know that their words would remain confidential and not be able to be traced back to them. Rather than focusing on the paradoxical nature of this stance, ways of meeting the wishes of this group (without negating the ability to use information from people who want to make it available) need to be explored.

Eysenbach and Till [23] recommended that consent should be obtained before using verbatim quotes, a view supported by the participants in this research. Seeking to gain permission to use posts from online discussion boards for research purposes is, however, likely to render the research difficult at best and unfeasible at worst. Posts often need to be considered in the context of the discussion, so if one poster was happy to give permission and wanted their views to be included, and another withheld permission, the conflict would need to be resolved and a decision made about whose wishes have primacy.

Reaching posters to seek permission is an additional dilemma. Board membership is often transient, contact details are not given or where they are they are out of date, and the way some boards are set up people have to have a posting history before they can be accessed or used. That creates another ethical dilemma to be resolved. Consent by proxy (through board administrators) generated mixed feelings from respondents (sometimes strongly expressed) so would not be an ethical resolution to this situation either. Failure to secure consent from sufficient posters could mean that the data available to use are not balanced nor representative of the boards. If the data are rendered unusable because of this then the wish of some posters that their contributions are used for greater good cannot be met. The outcome of this would be that a valuable research resource that could help develop concepts and care to ultimately benefit the wider community, which the boards seek to support, will be lost.

The AoIR [13] identifies that the concept of “the human subject” does not fit well with many online environments and that practical considerations such as harm, vulnerability and identification may be more important. Trying to apply rules that were created for a different context does not provide solutions to these dilemmas, leading us to the inevitable conclusion that new situations such as this require new rules. Having considered the dichotomy that both researchers and participants acknowledge, we propose that a new conceptual framework for online research is needed and that further work is required to ensure that all stakeholders in the process are included in the discussion. Key messages that have been drawn from our research are:

- Contributors appreciate the value of their boards to research that will ultimately benefit their communities and wish it to be used.
- Using aggregated data is acceptable to the community that created it.
- Using quotations ranges from being totally acceptable to totally unacceptable.

The Internet is not one entity, and different aspects may require different approaches. The information openly available on online

discussion boards does not fit into a human subject approach to research ethics, and while viewing it as more akin to text, mass media norms do not fit well either. It is a specific type of text—"personal health text" that requires sensitive handling and the generation of a new set of research rules.

We propose that the use of non-verbatim quotations should be considered as an alternative to verbatim quotations. At the simplest level, using nonverbatim quotations could be through editing the words of a single author to an extent by which it cannot be located by a search engine. As search engines become increasingly sophisticated, this solution is not likely to be easy to achieve, and indeed in time, this may become impossible. An alternative that we believe is preferable is to introduce a new concept to research—aggregated quotations. By this, we propose bringing several quotations around a topic together to maintain the essence of meaning, but also rendering it impossible to identify any individual poster.

The use of quotations is a way of presenting research that is generally linked with qualitative research [23] where the aim is to provide a richly textured and comprehensive set of data, without any gaps, and with the full breadth of interpretations included [41,42]. The value of verbatim quotations has been identified [43] to provide evidence of what was said; to present full or more clear explanation of participants' views; as a means of illustrating the codes and categories that have been developed from the data; to give depth to participant's meanings, "giving participants a voice", and to make research reports more readable. All of these aims can equally be achieved using thoughtfully constructed aggregated quotations.

The aim of the study was to identify what Internet users found acceptable rather than to explore the issue through a methodological lense. The findings, however, have implications for research methodology. It is beyond the scope of this paper to develop a detailed methodological exposition; however, rather than viewing this new approach as fitting into a qualitative framework, we see this proposal as being part of the formation of a new genre of research that focuses on the specific characteristics of Internet-based research. It may well be that what we are proposing is an approach that fits within mixed methodology, bringing together as it does the concept of rich data and quotations from qualitative research and the quantitative approach of aggregating data.

On a practical level, it is important that the discussion of the research methods details the approach taken to producing quotations, so that there is no risk of misrepresenting individuals. How many quotations and from how many individual contributors should be identified for each aggregated quotation used so that the quotations create a rich, comprehensive picture of what a range of participants feel while maintaining clarity and openness over the process.

This research was carried out with people contributing to online discussion boards, rather than any other type of online communication, such as chat rooms. The important consideration is that on discussion boards messages are typed and posted into a durable format. People posting are able to see a record of other posts and answers available to browse or be searched. Chat rooms and other similar forums suggest a more transient online presence (even if not a technically accurate perception), and previous research [44] has found that participants may have different responses. There is therefore no suggestion that these findings are transferable outside of discussion board-type communications.

Conclusion

The Internet is still a relatively young entity, and people are still developing their understanding of the potential conflicts between the public reality and private use of social media. It does however provide a source of rich research data about the experiences of people living with long-term conditions, among other things. This research found that people who contribute to discussion boards are doing so to share information and help others within that community and are sympathetic to researchers using the information they have shared to further that aim.

Acting ethically as researchers within this new genre is, however, a challenge. Trying to apply rules created for different situations does not provide a solution to how to use the information in health research. Therefore, the solution can be only to create a new set of norms that meet both the needs of researchers for rigor and openness in their research and the wish of posters to contribute, while protecting their personal information. The proposals in this paper, particularly the need to handle data from discussion boards in a new way, are offered as a practical way forward. They are also shared with the research and patient communities to offer a starting point for discussion.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questions asked.

[PDF File (Adobe PDF File), 178KB - [jmir_v15i6e112_app1.pdf](#)]

References

1. Vance K, Howe W, Dellavalle RP. Social internet sites as a source of public health information. *Dermatol Clin* 2009 Apr;27(2):133-136. [doi: [10.1016/j.det.2008.11.010](#)] [Medline: [19254656](#)]

2. Lewis PH. In a Recount, Cyber Census Still Confounds.: New York Times; 1996. URL: http://www.dartmouth.edu/~chance/course/Syllabi/Princeton96/internet_use.html [WebCite Cache ID 6EcveOOIt]
3. Nielsen Wire. 2012 Aug. Top US Web Brands and Technology News Websites URL: http://blog.nielsen.com/nielsenwire/online_mobile/august-2012-top-u-s-web-brands-and-technology-news-websites/ [WebCite Cache ID 6EcvujLDs]
4. Nguyen KD, Hara B, Chlebowski RT. Utility of two cancer organization websites for a multiethnic, public hospital oncology population: comparative cross-sectional survey. *J Med Internet Res* 2005 Jul 1;7(3):e28 [FREE Full text] [doi: [10.2196/jmir.7.3.e28](https://doi.org/10.2196/jmir.7.3.e28)] [Medline: [15998619](https://pubmed.ncbi.nlm.nih.gov/15998619/)]
5. Glasgow R. Engagement in a diabetes self-management website: Usage patterns and generalizability of program use. *J Med Internet Res* 2011;13(1):e9. [doi: [10.2196/jmir.1391](https://doi.org/10.2196/jmir.1391)]
6. Chiauzzi E, Pujol LA, Wood M, Bond K, Black R, Yiu E, et al. painACTION-back pain: a self-management website for people with chronic back pain. *Pain Med* 2010 Jul;11(7):1044-1058. [doi: [10.1111/j.1526-4637.2010.00879.x](https://doi.org/10.1111/j.1526-4637.2010.00879.x)] [Medline: [20545873](https://pubmed.ncbi.nlm.nih.gov/20545873/)]
7. Seçkn G. Internet Technology in Service of Personal Health Care Management: Patient Perspective. *Journal of Technology in Human Services* 2009;27(2):79-92.
8. Barney LJ, Griffiths KM, Banfield MA. Explicit and implicit information needs of people with depression: a qualitative investigation of problems reported on an online depression support forum. *BMC Psychiatry* 2011;11:88 [FREE Full text] [doi: [10.1186/1471-244X-11-88](https://doi.org/10.1186/1471-244X-11-88)] [Medline: [21592335](https://pubmed.ncbi.nlm.nih.gov/21592335/)]
9. Cowie GA, Hill S, Robinson P. Using an online service for breastfeeding support: what mothers want to discuss. *Health Promot J Austr* 2011 Aug;22(2):113-118. [Medline: [21819353](https://pubmed.ncbi.nlm.nih.gov/21819353/)]
10. Armstrong N, Powell J. Patient perspectives on health advice posted on Internet discussion boards: a qualitative study. *Health Expect* 2009 Sep;12(3):313-320. [doi: [10.1111/j.1369-7625.2009.00543.x](https://doi.org/10.1111/j.1369-7625.2009.00543.x)] [Medline: [19555377](https://pubmed.ncbi.nlm.nih.gov/19555377/)]
11. Williams SG. The Ethics of Internet Research. *Online Journal of Nursing Informatics* 2012;16(2).
12. Ess C, AoIR EW Committee. Ethical decision-making and Internet research: Recommendations from the aoir ethics working committee.: Association of Internet Researchers; 2002. URL: <http://aoir.org/reports/ethics.pdf> [accessed 2013-05-01] [WebCite Cache ID 6Ecw5VM3j]
13. Markham A, Buchanan E. Ethical Decision-Making and Internet Research: Version 2.: Association of Internet Researchers; 2012. URL: <http://aoir.org/reports/ethics2.pdf> [accessed 2013-05-01] [WebCite Cache ID 6Ecw5VM3j]
14. Huws U. Socio-Economic Research in the Information Society. Brighton; 2002. URL: <https://taw.bournemouth.ac.uk/owa/cbond@bournemouth.ac.uk/redir.aspx?C=77678d01ac7543af8079a555d2eb9351&URL=http%3a%2f%2fwww.webcitation.org%2f6EcwD4ys8>
15. Department of Health UK. Research governance framework for health and social care. 2005. URL: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4108962 [WebCite Cache ID 6EcwKQgWm]
16. Haigh C, Jones NA. An overview of the ethics of cyber-space research and the implication for nurse educators. *Nurse Educ Today* 2005 Jan;25(1):3-8. [doi: [10.1016/j.nedt.2004.09.003](https://doi.org/10.1016/j.nedt.2004.09.003)] [Medline: [15607241](https://pubmed.ncbi.nlm.nih.gov/15607241/)]
17. Bakardjieva M, Feenberg, A. Involving the Virtual Subject. *Ethics and Information Technology* 2000;2(4):233-240.
18. 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](https://doi.org/10.1016/j.socscimed.2007.03.005)] [Medline: [17512105](https://pubmed.ncbi.nlm.nih.gov/17512105/)]
19. Holmes S. Methodological and ethical considerations in designing an Internet study of quality of life: a discussion paper. *Int J Nurs Stud* 2009 Mar;46(3):394-405. [doi: [10.1016/j.ijnurstu.2008.08.004](https://doi.org/10.1016/j.ijnurstu.2008.08.004)] [Medline: [18838135](https://pubmed.ncbi.nlm.nih.gov/18838135/)]
20. McKee HA. Ethical and legal issues for writing researchers in an age of media convergence. *Computers and Composition* 2008;25(1):104-122.
21. Stewart K, Williams M. Researching online populations: the use of online focus groups for social research. *Qualitative Research* 2005;5(4):395-416.
22. Capurro R, Pingel C. Ethical issues of online communication research. *Ethics and Information Technology* 2002;4(3):189-194.
23. 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](https://pubmed.ncbi.nlm.nih.gov/11701577/)]
24. Flicker S, Haans D, Skinner H. Ethical dilemmas in research on Internet communities. *Qual Health Res* 2004 Jan;14(1):124-134. [doi: [10.1177/1049732303259842](https://doi.org/10.1177/1049732303259842)] [Medline: [14725180](https://pubmed.ncbi.nlm.nih.gov/14725180/)]
25. 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](https://doi.org/10.2196/jmir.1600)] [Medline: [21345783](https://pubmed.ncbi.nlm.nih.gov/21345783/)]
26. Fox S. Pew Internet Project. 2011. Peer‐to‐peer healthcare URL: <http://pewinternet.org/Reports/2011/P2PHealthcare/Summary-of-Findings.aspx> [accessed 2013-02-25] [WebCite Cache ID 6Eh0oB6SO]
27. Vanguard. 2012. Facebook Third Largest Country If URL: <http://www.vanguardngr.com/2012/05/facebook-third-largest-country-if/> [accessed 2013-02-22] [WebCite Cache ID 6EcwWBriW]
28. Ahmed OH, Sullivan SJ, Schneiders AG, McCrory P. iSupport: do social networking sites have a role to play in concussion awareness? *Disabil Rehabil* 2010;32(22):1877-1883. [doi: [10.3109/09638281003734409](https://doi.org/10.3109/09638281003734409)] [Medline: [20367328](https://pubmed.ncbi.nlm.nih.gov/20367328/)]
29. Bender JL, Jimenez-Marroquin MC, Jadad AR. Seeking support on facebook: a content analysis of breast cancer groups. *J Med Internet Res* 2011;13(1):e16 [FREE Full text] [doi: [10.2196/jmir.1560](https://doi.org/10.2196/jmir.1560)] [Medline: [21371990](https://pubmed.ncbi.nlm.nih.gov/21371990/)]

30. De la Torre-Díez I, Javier Diaz-Pernas F, Anton-Rodriguez M. A Content Analysis of Chronic Diseases Social Groups on Facebook and Twitter. *Telemedicine and e-Health* 2012;18(6):404-408.
31. Bournemouth University. Research Ethics Code of Practice. Poole, UK: Bournemouth University; Sep 2009.
32. James N, Busher H. Online interviewing. London: Sage Publications Limited; 2009.
33. Guest G, Bunce A, Johnson L. How Many Interviews Are Enough? *Field Methods* 2006 Feb;18(1):59.
34. Guba EG, Lincoln YS. Fourth generation evaluation. Newbury Park: Sage Publications; 1989.
35. Flick U. An introduction to qualitative research. London: Sage Publications; 2009.
36. Papacharissi Z. The virtual sphere 2.0: The Internet, the Public Sphere, and Beyond. In: Chadwick A, editor. *Routledge Handbook of Internet Politics*. London: Routledge; 2009:230-245.
37. Youngs G. Blogging and globalization: the blurring of the public/private spheres. *Aslib Proceedings* 2009;61(2):127-138. [doi: [10.1108/00012530910946884](https://doi.org/10.1108/00012530910946884)]
38. Sarangi S. Public discourse. In: Zienkowski J, Ostman J, Verschueren J, editors. *Handbook of Pragmatics Highlights (HoPH)*. Amsterdam: John Benjamins; 2011:248.
39. Liu Z, Liu L, Li H. Determinants of information retweeting in microblogging. *Internet Research* 2012;22(4):443-466.
40. Bassett EH, O'Riordan K. Ethics of Internet research: Contesting the human subjects research model. *Ethics and Information Technology* 2002;4(3):233-247.
41. Russell CK, Gregory DM. Evaluation of qualitative research studies. *Evid Based Nurs* 2003 Apr;6(2):36-40 [FREE Full text] [Medline: [12710415](https://pubmed.ncbi.nlm.nih.gov/12710415/)]
42. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse Educ Today* 2004 Feb;24(2):105-112. [doi: [10.1016/j.nedt.2003.10.001](https://doi.org/10.1016/j.nedt.2003.10.001)] [Medline: [14769454](https://pubmed.ncbi.nlm.nih.gov/14769454/)]
43. Corden A, Sainsbury R. Using verbatim quotations in reporting qualitative social research: researchers' views. Social Policy and Research Unit: University of York. York, UK: Social Policy and Research Unit: University of York; 2006.
44. Hudson J, Bruckman A. "Go Away": Participant Objections to Being Studied and the Ethics of Chatroom Research. *The Information Society* 2004 Apr 2004;20(2):127-139. [doi: [10.1080/01972240490423030](https://doi.org/10.1080/01972240490423030)]

Edited by G Eysenbach; submitted 13.11.12; peer-reviewed by S Lim Choi Keung, D Pullman; comments to author 31.01.13; revised version received 25.02.13; accepted 28.03.13; published 07.06.13.

Please cite as:

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

URL: <http://www.jmir.org/2013/6/e112/>

doi: [10.2196/jmir.2435](https://doi.org/10.2196/jmir.2435)

PMID: [23748147](https://pubmed.ncbi.nlm.nih.gov/23748147/)

©Carol S. Bond, Osman Hassan Ahmed, Martin Hind, Bronwen Thomas, Jaqui Hewitt-Taylor. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 07.06.2013. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Utilization and Perceived Problems of Online Medical Resources and Search Tools Among Different Groups of European Physicians

Marlene Kritz¹, MSc; Manfred Gschwandtner¹, MSc; Veronika Stefanov², PhD; Allan Hanbury², PhD; Matthias Samwald^{2,3}, PhD

¹Society of Physicians Vienna, Vienna, Austria

²Information & Software Engineering Group, Institute of Software Technology and Interactive Systems, Vienna University of Technology, Vienna, Austria

³Section for Medical Expert and Knowledge-Based Systems, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Vienna, Austria

Corresponding Author:

Matthias Samwald, PhD

Section for Medical Expert and Knowledge-Based Systems

Center for Medical Statistics, Informatics, and Intelligent Systems

Medical University of Vienna

Spitalgasse 23

Vienna, 1090

Austria

Phone: 43 1404006665

Fax: 43 1404006625

Email: matthias.samwald@meduniwien.ac.at

Abstract

Background: There is a large body of research suggesting that medical professionals have unmet information needs during their daily routines.

Objective: To investigate which online resources and tools different groups of European physicians use to gather medical information and to identify barriers that prevent the successful retrieval of medical information from the Internet.

Methods: A detailed Web-based questionnaire was sent out to approximately 15,000 physicians across Europe and disseminated through partner websites. 500 European physicians of different levels of academic qualification and medical specialization were included in the analysis. Self-reported frequency of use of different types of online resources, perceived importance of search tools, and perceived search barriers were measured. Comparisons were made across different levels of qualification (qualified physicians vs physicians in training, medical specialists without professorships vs medical professors) and specialization (general practitioners vs specialists).

Results: Most participants were Internet-savvy, came from Austria (43%, 190/440) and Switzerland (31%, 137/440), were above 50 years old (56%, 239/430), stated high levels of medical work experience, had regular patient contact and were employed in nonacademic health care settings (41%, 177/432). All groups reported frequent use of general search engines and cited “restricted accessibility to good quality information” as a dominant barrier to finding medical information on the Internet. Physicians in training reported the most frequent use of Wikipedia (56%, 31/55). Specialists were more likely than general practitioners to use medical research databases (68%, 185/274 vs 27%, 24/88; $\chi^2_2=44.905$, $P<.001$). General practitioners were more likely than specialists to report “lack of time” as a barrier towards finding information on the Internet (59%, 50/85 vs 43%, 111/260; $\chi^2_1=7.231$, $P=.007$) and to restrict their search by language (48%, 43/89 vs 35%, 97/278; $\chi^2_1=5.148$, $P=.023$). They frequently consult general health websites (36%, 31/87 vs 19%, 51/269; $\chi^2_2=12.813$, $P=.002$) and online physician network communities (17%, 15/86, $\chi^2_2=9.841$ vs 6%, 17/270, $P<.001$).

Conclusions: The reported inaccessibility of relevant, trustworthy resources on the Internet and frequent reliance on general search engines and social media among physicians require further attention. Possible solutions may be increased governmental support for the development and popularization of user-tailored medical search tools and open access to high-quality content for physicians. The potential role of collaborative tools in providing the psychological support and affirmation normally given by

medical colleagues needs further consideration. Tools that speed up quality evaluation and aid selection of relevant search results need to be identified. In order to develop an adequate search tool, a differentiated approach considering the differing needs of physician subgroups may be beneficial.

(*J Med Internet Res* 2013;15(6):e122) doi:[10.2196/jmir.2436](https://doi.org/10.2196/jmir.2436)

KEYWORDS

information seeking behavior; physicians; Internet; search engine; information quality; language barriers

Introduction

Background

Over the last decades, the World Wide Web has become an important source of information within the medical health domain [1]. Physicians' information needs, information-seeking behavior, and their use of online resources have been well studied [2]. Physicians primarily access the Internet at the point of care [3] to pursue medical updating and to communicate with colleagues [4]. The Internet gives medical professionals access to a vast amount of high-quality medical information, which could potentially aid medical decision making and patient care. It has been found that 51% of physicians claim that the Web has influenced treatment and assisted them in diagnostic procedures [5], and 30% of a sample of 411 physicians declared "often" changing medication and treatment plans as a consequence of obtaining information from the Internet [6]. The benefit and usability of medical information provided on the Internet increasingly relies on adequate content, quality evaluation, and the skilled selection of relevant websites. To improve on how computers can help physicians with medical information retrieval, an understanding of online resource and search requirements of physicians in different health care settings is required.

Use of Resources

There has been some research discussing what makes a source useful for medical professionals. Shaughnessy and colleagues [7] have measured the usefulness of information resources with the formula:

$$\text{Utility} = (\text{relevance} \times \text{validity} \times \text{interactivity}) / \text{work to access}$$

According to the formula, the ideal information source is directly relevant, contains valid information, and can be accessed with a minimal amount of work. Sources with low "work to access" (ie, easy accessibility) and high surface relevance—such as general search engines and colleagues—are often reported as popular among time-constrained medical professionals [3,8]. However, answers provided by such sources often lack scientific validity. In contrast, medical research databases such as PubMed can provide journal articles and systematic reviews backed up by good evidence but often fail to quickly provide relevant content, resulting in a considerable amount of mental "work to access" [9]. In addition, point-of-care databases such as UpToDate are associated with high levels of economic "work to access" as they provide relevant and valid information but require the payment of substantial subscription fees to be accessible. Consequently, "low work to access" resources such as general search engines and colleagues have remained

dominant information resources within the medical domain [6]. The role of the "psychological support and affirmation" provided by a medical colleague has remained dominant and irreplaceable [10]. However, while it has been postulated that general search engines such as Google can aid the diagnostic process [8] and can be efficient at answering quick questions within a critical care setting [9], it is questionable whether physicians always have the time and expertise to pursue the necessary data selection.

Barriers

It has been well documented that physicians face barriers that keep them from finding relevant information to their queries [11]. However, to date, there is a lack of research treating physicians as a heterogeneous group and comparing potential search barriers in different subgroups of physicians. A differentiated understanding of problems associated with current search systems could be beneficial towards the development of efficient user-tailored medical search systems.

Tools

Tools to overcome search barriers by simplifying content selection and quality evaluation of information are of growing interest within the medical search domain. There is a lack of knowledge of the tools physicians prefer to use when searching for online information. In particular, there is a lack of data on the acceptance of social, collaborative tools for aiding the selection of relevant search results.

Objective

This paper aims to provide insight on the professional Internet use among different subgroups of physicians: their use of resources, tools, and potential barriers in obtaining answers from the Internet. We are expanding current knowledge in this area by presenting a large quantitative survey among physicians located in several European countries and employed in a wide variety of health care settings. To date, no detailed, large-scale quantitative study has been published on the use of online resources, preferences of search tools, and search barriers of different subgroups of European physicians. Most studies in the field have been carried out in countries where English is the primary language of communication. Thus, the role of nonEnglish languages in online search behavior of medical professionals has remained unclear.

Our main objective was to analyze the use of online resources, search tools, and perceived search barriers among European physicians. We compared different subgroups of physicians to determine whether level of academic qualification, type of medical specialization, and medical experience could impact search behavior.

The outcomes of the present survey could assist the future development of effective medical search systems and provide orientation for the potential creation of institutional policies on using the Web in health care settings.

Methods

This survey is part of the project KHRESMOI, funded by the European Union. One of the goals of the project is to investigate the search requirements of physicians in order to guide the development of a medical search engine.

Questionnaire

An online questionnaire based on categorical assumptions of previous qualitative literature, on medical information needs [12], use of online resources [13], and physician search behavior and preferences [6] was developed. While previous research provided a good starting point for the questionnaire design, most items were modified, extended, or deleted to fit the requirements of European physicians after doing preliminary, structured pilot interviews with Austrian physicians and disseminating a pilot questionnaire to Austrian, Spanish, Swiss, and English physicians. We used informal semistructured interviews to develop the questionnaire. The questionnaire was validated via an “online test period” of 1 week, in which 12 physicians completed the survey and provided subsequent feedback.

The final version of the survey questionnaire (see [Multimedia Appendix 1](#)) consisted of 6 parts encompassing a total of 47 questions and was made available online in English, German, French, and Spanish on an open-access platform hosted by Health on the Net [14]. The questionnaire could be accessed by clicking on a link that was included in the emails sent out, or by clicking on a banner on the Health on the Net website. The number of items displayed per page varied from 1 to 14, and it was ensured that each part was displayed on one page. The whole questionnaire was spread over a total of 8 pages. All items contained the option of a nonresponse item (eg, “not applicable”) and/or an open-format “other item”. Participants were allowed to change responses using the back button. Computer Internet Protocol (IP) addresses were tracked and used to identify duplicate entries. In the case of duplicate entries within a time frame of 12 hours, the first entry was excluded from the analysis. The study was anonymous. Each participant was presented with an information sheet and asked to confirm being a physician/medical student prior to completing the questionnaire. The questionnaire was edited in simple HTML without Java applets or special scripts to allow a maximum number of the addressed users to access it with their browsers.

Most questions were either of multiple response/dichotomous questions or could be answered by selecting an option on a 5-point Likert scale rating. In our analysis, we included 15 sociodemographic variables and present the results of 14 selected items related to the use of resources, tools, and barriers ([Table 1](#)).

Table 1. Subgroups of physicians.

	Definition/inclusion criteria
Physician in training	Physicians that had completed a medical degree (MD) but were currently pursuing physician training to become a general practitioner or specialist.
Qualified physicians	Included all physicians that were working as qualified physicians (general practitioners and specialists)
General practitioners	Included all physicians that were working as qualified general practitioners
Specialists	Included all physicians who are were working as qualified specialists
Specialists without professorship	A subgroup of the specialists group, included physicians who are qualified specialists without having attained lectureship or professorship
Medical professors	A subgroup of the specialists group, included physicians who had completed a postgraduate lectureship qualification or were appointed university professors

Study Population

The target population consisted of European physicians of all specialties and was a convenience sample. The questionnaire was promoted in June and July 2011 by email at random to about 15,000 physicians around Europe through a banner on websites, medical newsletters, and an article in an Austrian public health journal with approximately 35,000 readers. The following institutions sent out newsletters with a link and information on the questionnaire: Society of Physicians Vienna (2800 members), Geneva Doctors Association (3100 members), Austrian Society of Internal Medicine (2100 members), Austrian Society for Gastroenterology and Hepatology (800 members), and Medical Media in Austria (5500 members). The following institutions placed a banner pointing to the questionnaire on their websites: Doctors.net.uk, Society of Physicians in Vienna,

Health on the Net Foundation, Professional Association of German Internists, and the European Academy of Allergy and Clinical Immunology.

Statistical Analysis

Statistical analyses were performed using SPSS 21 for Windows. Tables and graphs were constructed using Microsoft Office Excel 2007. Due to unsolved controversies in the literature of reliably treating Likert-items as interval-scaled, 5-point Likert items were treated as ordinal and other items as nominal or dichotomized [15]. Thus a nonparametric approach, believed to be the most reliable method for such data [16], was used for all statistical inferences.

Differences between independent variables—medical specialization (specialist vs general practitioner), level of

qualification (physician in training vs qualified physician), and academic qualification (medical professor vs medical specialist without professorship)—were initially explored using a Mann-Whitney U Test for all 5-point Likert scale items. However, since the distributions of most of the ordinal 5-point Likert-items were skewed, we decided to simplify all responses on 5-point Likert scales by re-categorizing responses into only three items (“always/often”, “sometimes”, “rarely/never”, or similarly “very important/important”, “neutral”, “unimportant”). Multiple response questions were coded and analyzed as dichotomous yes/no variables.

The questionnaire examined various sociodemographic variables, the use of online resources, available time, search strategy, barriers, search tools used, advanced search features used, and mobile accessibility. The sample population was split into subgroups of physicians, as defined in [Table 1](#).

Questionnaire results were analyzed to provide descriptive statistics, graphs, and cross-tabulations. A Chi-square analysis was performed to assess differences between ordinal dependent variables with two to four categories as well as all dichotomous and nominal data. A P value $\leq .05$ was considered as statistically significant.

Exclusion Criteria

Participants were excluded if they answered fewer than 30% of the questions (ie, at least 14 of the 46 questions had to be answered), did not complete their medical degree yet, or were not involved in patient treatment. The low threshold of inclusion was chosen to deliberately allow for partially filled out questionnaires. For example, it was important to include physicians who selectively failed to fill out the demographics section (about 23% of the total sample included in the analysis), to avoid creating a bias by excluding physicians who merely wanted to assure additional confidentiality.

Results

Of the initial 640 participants taking part in the study, 140 were excluded based on the selection criteria described earlier. This resulted in a sample of 500 participants that had completed all or a substantial part of the questionnaire being included in the analysis. Results are reported as a percentage of the total number of responses to each question (the denominator therefore varies according to the individual question response rate). No statistical corrections such as weighting were used. However,

nonresponders to individual questions were excluded from the analysis of those questions.

Demographics

In total, 63% (271/430) of the participants were male, and 56% (239/430) of the participants were older than 50 years. Most participants came from the main areas of questionnaire dissemination: the majority came from Austria (43%, 190/440), almost a third from Switzerland (31%, 137/440), followed by the United Kingdom (8%, 37/440), and Germany (4%, 17/440). Respectively, most participants spoke German (46%, 198/429) or French (31%, 133/429). The majority of the participants (81%, 345/427) reported living in urban areas. The reported level of education and medical work experience was exceptionally high; 13% of the physicians (54/432) reported having completed medical lectureship training. More than half of the respondents reported having obtained work experience of “20 years or more” (55%, 238/428). In fact, 96% (415/432) reported currently working as medical professionals. Most were self-employed (39%, 168/432) or worked in nonacademic health care settings (41%, 177/432), only 4% (17/432) identified as retired or unemployed, and 64% (278/432) of the physicians identified as working specialists, 21% (89/432) as general practitioners, and 13% (55/432) reported pursuing physician training. 90% (387/429) of the physicians reported seeing patients on a regular basis. However, the majority (87%, 337/387) of those with regular patient contact consulted less than 40 patients per working day.

Overall, levels of self-perceived competence of medical English were high. Among the nonnative English speakers (91%, 391/429), 89% (348/391) of the physicians reported that their level of medical English is above average, with only 1 participant reporting not understanding any English. While 99% (492/498) reported having regular Internet access, 92% (454/496) reported using the Internet on a daily basis. High levels of Internet experience were reported, with 82% (408/499) using the Internet for more than 10 years. In addition, of those physicians who reported having regular direct patient contact (90%, 387/429), as many as 24% (92/385) reported that they frequently (“often” or “always”) access the Internet during a patient consultation, and 44% (171/385) reported never accessing the Internet during a patient consultation.

[Table 2](#) illustrates how the physicians in the sample were distributed among the subgroups we defined based on medical qualification, medical specialization, and academic qualification.

Table 2. Comparisons between different groups of physicians.

	Medical qualification		Medical specialization		Academic qualification	
	Physician in training	Qualified physician	General practitioner	Specialist	Specialist without professorship	Medical professor
n/N ^a	55/422	367/422	89/367	278/367	224/278	54/278
% of N	13	87	24	76	81	19

^aN is based on the number of respondents that provided a response to the corresponding question in the questionnaire.

Use of Online Resources

The frequencies with which physicians reported using various types of online resources are shown in [Figure 1](#) and in [Table 1](#)

in [Multimedia Appendix 2](#). [Figure 1](#) illustrates the percentage of respondents per group that claimed to use a given resource “often” or “always” when searching for medical information on the Internet. Numerical values and statistical parameters are

shown in Table 1 in [Multimedia Appendix 2](#). Most physicians reported frequently using general search engines (78%, 372/476), medical research databases (59%, 277/469), Wikipedia (40%, 184/461), or medical society websites (38%, 176/467) to obtain medical information online. Currently available specialized medical search tools were reported as the least popular resources (6%, 25/458).

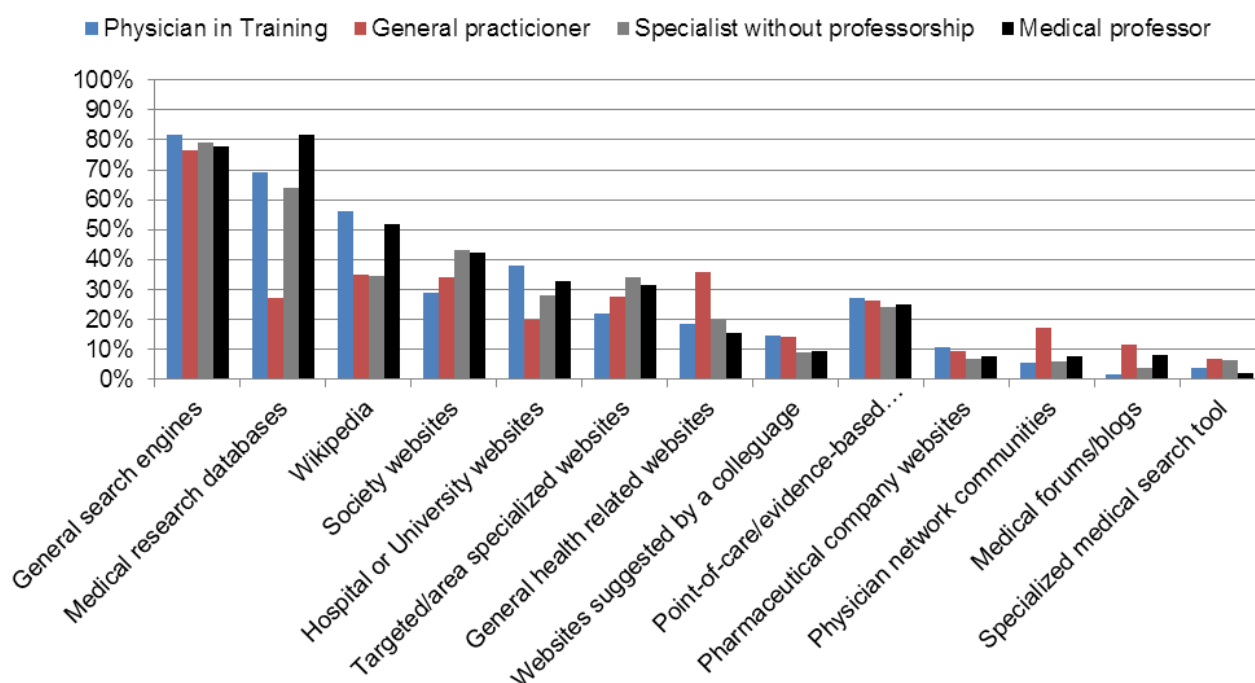
Both the level of medical specialization and level of academic experience were associated with how often physicians consulted medical research databases. Medical professors were the only group that cited using medical research databases (82%, 44/54) as often as general search engines (75%, 42/54) during their search for medical information. Specialists reported using medical research databases more frequently than general practitioners (68%, 185/274 vs 27%, 24/88; $\chi^2_2=44.905$, $P<.001$). On the other hand, general practitioners were more likely than specialists to consult general health websites (36%, 31/87 vs 19%, 51/269; $\chi^2_2=12.813$, $P=.002$), online physician communities (17%, 15/86 vs 6%, 17/270; $\chi^2_2=9.841$, $P<.001$), and medical forums/blogs (12%, 10/86 vs 5%, 12/266; $\chi^2_2=9.841$, $P<.01$). Medical professors were more likely than specialists without professorship to use medical research databases (82%, 44/54 vs 64%, 141/220; $\chi^2_2=7.461$, $P=.024$), and Wikipedia (52%, 27/52 vs 34%, 74/215; $\chi^2_2=7.461$, $P=.024$). Physicians in training were more likely to report using Wikipedia (56%, 31/55 vs 37%, 131/353; $\chi^2_2=8.997$, $P=.011$),

hospital/university websites (38%, 21/55 vs 27%, 95/357; $\chi^2_2=6.409$, $P=.041$), and websites suggested by colleagues (15%, 8/54 vs 10%, 37/355; $\chi^2_2=8.653$, $P=.001$) than qualified physicians. Qualified physicians were more likely than physicians in training to consult general medical society websites (41%, 147/361 vs 29%, 16/55; $\chi^2_2=11.622$, $P=.003$).

Overall, online physician communities were relatively unknown; 28% (116/419) of the total sample was unaware of their existence. Only 24% of the total sample (72/303) reported using such communities. Of those who did use them, the most popular physician communities reported were “Doctors.net” ($n=10$) and “doc2doc” ($n=8$). However, 33% of general practitioners (29/89) and 24% (65/276) of the specialists were unaware of the existence of physician society communities. However, among the physicians who were aware of their existence, general practitioners were more likely to access physician society community websites than specialists ($\chi^2_1=19.083$, $P<.001$).

We asked physicians which restriction criteria they used during their search for medical information. Overall, journals (48%, 150/333), source (35%, 136/391), and books (19%, 75/391) were the most popular restriction criteria reported. Specialists were more likely than general practitioners to restrict their search results to journals (50%, 130/258 vs 39%, 29/75; $\chi^2_1=5.951$, $P=.015$) while general practitioners were more likely to restrict their search by source (45%, 34/75 vs 33%, 86/258; $\chi^2_1=4.911$, $P=.0026$).

Figure 1. Online resources used for obtaining medical information.



How Much Time Do Physicians Spend on Searches?

The average amount of time that physicians were willing to devote to complex queries was between 10-20 minutes ([Table 3](#)). In comparison to other groups, medical professors and physicians in training devoted the largest amount of time to

complex queries. General practitioners reported devoting the least amount of time to complex queries. A substantial proportion (46%, 209/453) of physicians reported frequently checking second and third pages of results.

Physicians in training devoted more time to complex queries than qualified physicians (Mode: 20-30 minutes vs 10-20 minutes; $\chi^2_4=9.619$, $P=.047$). In total, 39% (21/54) of medical professors reported that they would be prepared to devote more than 30 minutes to a complex query. Medical professors were more likely to devote time to complex queries ($\chi^2_4=25.3028$, $P<.001$) and check the second and third page of search results ($\chi^2_2=10.9823$, $P=.004$) than specialists without professorship. Thus, it appears that they had more time to search more comprehensively.

How Successful Are Physicians at Searching the Web?

Table 4 illustrates how often physicians fail at retrieving information from the Internet, why they think they failed, and what they usually did if they could not find the required information.

Overall, 59% (266/453) of physicians reported “sometimes”, “often”, or “always” being confronted with situations where they fail to retrieve the medical information they require from the Internet. As many as 15% (67/453) reported that this happened on a frequent basis (“often”, “always”).

When asked about situations where answers to medical questions could not be found on the Internet, most participants (76%, 278/368) indicated that an excess number of search results made it too time consuming to select relevant information, and 24% (90/368) were not sure how to formulate their query. Qualified physicians were more likely (27%, 81/298 vs 10%, 4/41;

$\chi^2_1=6.121$, $P=.013$) to express difficulties in formulating search queries than physicians in training.

As a consequence of failing to find the answer to a medical question, most participants (61%, 261/427) reported doing another, more specific search, and 8% (34/427) reported consulting a colleague when failing to find the answer on the Internet.

What Are the Barriers to Finding Information?

Inaccessibility to relevant information, lack of time, and questionable trustworthiness were the most prominent barriers mentioned (Table 2 in [Multimedia Appendix 2](#)). Other important barriers were the absence of good quality filters/ratings as well as the questionable trustworthiness of search results.

In line with the findings illustrated in Table 3, general practitioners were more likely than specialists to perceive a lack of time to find relevant information as a barrier to obtaining medical information online (59%, 50/85 vs 43%, 111/260; $\chi^2_1=7.231$, $P=.007$). Physicians in training were more likely than qualified physicians to perceive inaccessibility of relevant information ($\chi^2_1=6.7742$, $P=.009$), too general search results ($\chi^2_1=4.884$, $P=.0027$), and questionable trustworthiness ($\chi^2_1=8.045$, $P=<.001$) as posing problems within the medical search domain. The finding that “search results appeared too general” is not surprising, since, as noted earlier, most participants reported the frequent reliance on generic search engines and social media when obtaining medical information.

Table 3. Time that physicians report having for answering complex questions (question: “How much time can you or are you generally willing to spend on trying to find the answer to an important, complex clinical question?”).

	Level of qualification		Level of medical specialization		Level of academic specialization		Overall % (n)
	Qualified Physician % (n)	Physician in Training % (n)	General practitioner % (n)	Specialist % (n)	Specialist without professorship % (n)	Medical professor % (n)	
< 10 min.	14 (50)	6 (3)	19 (17)	12 (33)	14 (30)	6 (3)	13 (53)
10-20 min.	33 (122)	28 (15)	38 (34)	32 (88)	33 (73)	28 (15)	33 (137)
20-30 min.	23 (85)	32 (17)	17 (15)	25 (70)	25 (55)	28 (15)	24 (102)
30-40 min.	17 (62)	11 (6)	18 (16)	17 (46)	19 (43)	6 (3)	16 (68)
> 40 min.	12 (46)	24 (13)	8 (7)	14 (39)	10 (21)	33 (18)	14 (59)
Total	100 (365)	100 (54)	100 (89)	100 (276)	100 (222)	100 (54)	100 (419)
Central ten- dency	Mode: 10-20 minutes	Mode: 20-30 minutes	Mode:	Mode: 10-20 minutes	Mode: 10-20 min- utes	Mode: More than 40 minutes	Mode: 10- 20 minutes
Statistical signifi- cance	$\chi^2_4=9.619$, $P=.047$		Not significant		$\chi^2_4=25.3028$, $P<.001$		

Table 4. Overall frequency, reason, and consequence of failure to retrieve medical information from the Internet.

Question	Categories	% (n/N) ^a
Frequency of failure: “How often do you face the situation where you cannot find the answer to a medical question on the Internet?”		
	Never, Rarely	41 (187/453)
	Sometimes	44 (199/453)
	Often, Always	15 (67/453)
Reason of failure: “What is the most common reason you failed to find an answer?”		
	Too many search results, too time-consuming to choose	76 (278/368)
	I was not sure how to formulate the query	24 (90/368)
Consequence of failure		
	Do another search using search terms that get MORE SPECIFIC	61 (261/427)
	Do another search using search terms that get LESS SPECIFIC	19 (81/427)
	Nothing, I stop searching on the Internet	9 (40/427)
	I send an email/Skype/chat with a colleague	8 (34/427)
	I post the question in a medical forum/physician community	3 (11/427)

^aN is based on the number of respondents that provided a response to the corresponding question in the questionnaire.

Search Tools and Advanced Search Options

As shown in Figure 2 and Table 3 in Multimedia Appendix 2, the search engine features physicians reported to be most important or desired were the possibility of “being able to quality rate medical information and perceive ratings of other physicians” (52%, 212/410), advanced search (45%, 188/422), being presented a list of popular websites (45%, 181/403), suggested relevant topics (39%, 159/413), search of images (33%, 133/409), and use from mobile platforms (33%, 137/412). The 15 most important tools are illustrated. Numerical values and statistical parameters are shown in Table 3 in Multimedia Appendix 2.

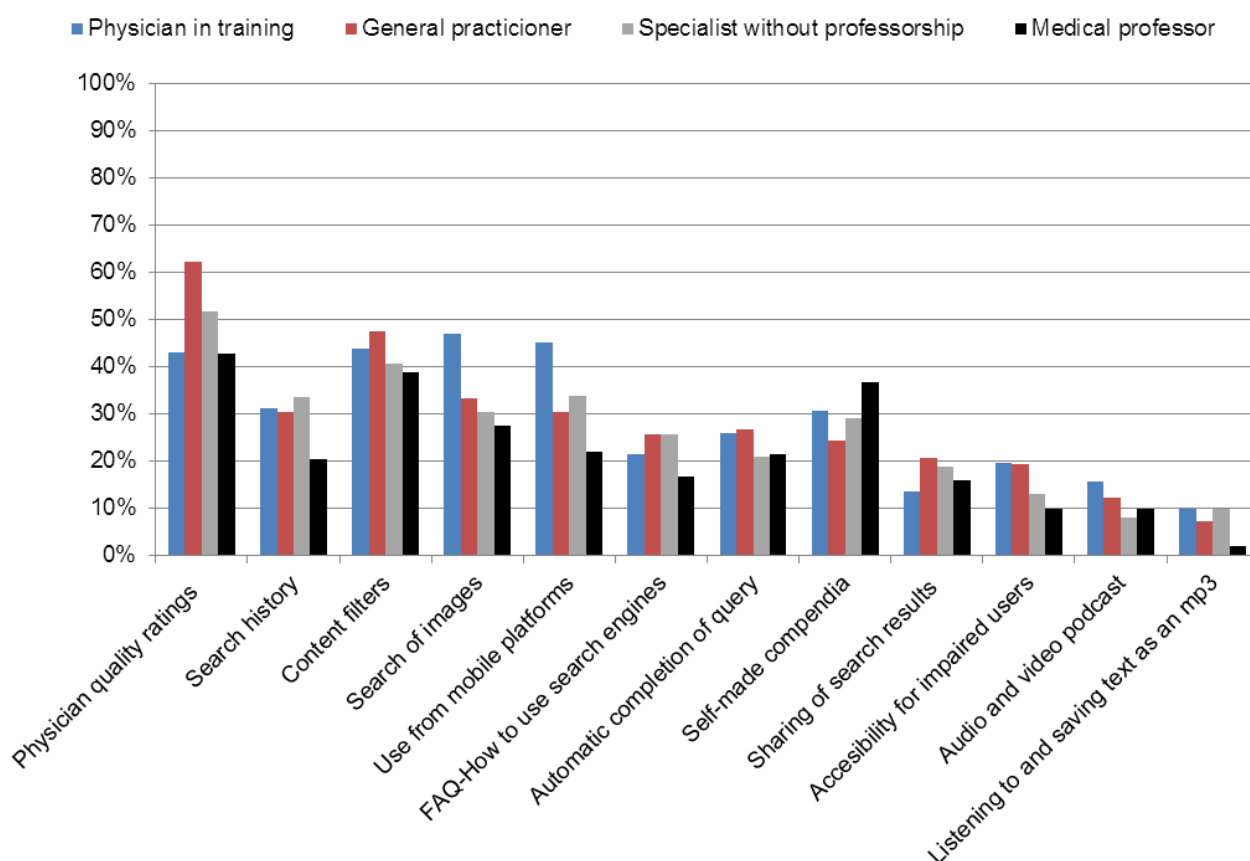
Physicians in training placed substantial importance on a wide variety of search tools, especially collaborative tools (physician quality ratings), media tools (mobile accessibility, image search), and data selection tools (advanced search). Congruent with these findings, it was found that physicians in training were more likely to report using mobile devices ($P<.01$) and placed higher importance on the use of video for the presentation of medical information ($P=.022$) than qualified physicians. Qualified physicians were more likely than physicians in training to place importance on the possibility to integrate patient data in the search process (36%, 124/343 vs 18%, 9/50; $\chi^2_2=6.617$, $P=.037$). On the other hand, physicians in training were more likely than qualified physicians to regard the existence of a spelling correction tool as important (31%, 16/51 vs 18%, 61/348;

$\chi^2_2=5.929$, $P=.052$). A possible explanation for this finding is that it could be assumed that they have less experience in using medical terminology.

General practitioners were more likely than specialists to perceive the availability of medical calculators as important (42%, 34/82 vs 30%, 77/258; $\chi^2_2=6.538$, $P=.0038$) and primarily sought collaborative and data selection tools. Furthermore, they were more likely than specialists to use the advanced search options of restricting results by language (48%, 43/89 vs 35%, 97/278; $\chi^2_1=5.148$, $P=.023$) and country (23%, 20/89 vs 13%, 36/278; $\chi^2_1=4.728$, $P=.003$). Some of this effect could be attributed to the fact that general practitioners reported lower levels of medical English competence than specialists.

Being able to restrict results by date (45%, 190/422) and language (36%, 153/422) were popular features, while filtering search results by format (13%, 56/422) was comparatively uncommon. Of the physicians whose mother tongue was not English, 26% (145/391) reported using language filters. Although language was not explicitly reported as a frequent barrier and levels of medical English competency were reported as high, a substantial proportion of physicians appear to prefer content being presented in their mother tongue. Furthermore, moderate importance was assigned to language tools such as automatic completion of queries (23%, 95/409), spelling correction (19%, 78/415), and automatic translation (15%, 65/421).

Figure 2. Responses to the question "How important do you perceive the following tools?", percentages of responses with the answer "Important" are illustrated.



Discussion

Principal Findings

In line with previous research [6], general-purpose search engines (eg, Google), medical research databases (eg, PubMed), and Wikipedia were popular resources, while specialized search engines were unpopular. The popularity of medical society websites (ie, websites representing organized groups of physicians) was somewhat surprising. This could be explained by the fact that these websites often contain large amounts of peer-reviewed articles and medical news.

General practitioners were primarily interested in secondary resources including general, "easy to use" health content and collaborative resources (eg, physician network communities, forums). In contrast, specialists, especially medical professors, expected to access primary resources (eg, scientific journals, PubMed abstracts). Qualified physicians liked to access medical society websites while physicians in training were drawn towards encyclopedic resources such as Wikipedia and reported consulting websites suggested by their colleagues.

Current medical search engines are often based solely around primary resources and may consequently fail to address the resource expectations of a substantial fraction of physicians. Even though point-of-care databases (eg, UpToDate) provide reliable, evidence-based clinical information, the reported use among physicians was shown to be limited. A possible explanation could be that most physicians are not willing to pay

high subscription fees to access medical information. In line with this notion, our study revealed that more than half of the physicians in our sample prefer advertisement-driven, free search services to paid services.

Our findings suggest that all groups of physicians, except general practitioners, are prepared to devote a considerable amount of time to important complex queries. However, physicians often fail to find the required information online. Known barriers to medical information retrieval such as inaccessibility of relevant information, questionable trustworthiness, and information overload [12] are confirmed by our findings. However, it appears that physicians either lack the time (general practitioner) or the skill (physician in training) to perform adequate data selection and evaluation when confronted with vast amounts of information.

The popularity of data selection tools (eg, suggested relevant topics, popular websites where most users found the answer) supports the notion that physicians seek help in finding relevant search results. The importance of the medical colleague in answering clinical questions has been verified over the last decade [6,17]. However, it has remained unclear to what extent the opinion of an anonymous colleague on the Internet can provide the needed "affirmation and support" [2]. Some support for the importance of "digital colleagues" was provided by the substantial fraction of physicians requesting collaborative tools (eg, functionality allowing them to share and perceive physician quality ratings of medical content). A general practitioner might,

for example, seek the feedback of a specialist to obtain help in verifying content from general health websites. A specialist may seek experienced digital colleague support to obtain help in answering a complicated specific medical question. Thus, our data suggest that physicians seek online colleague support and feedback to aid them in data selection and medical decision making.

The Internet provides physicians with the opportunity to actively communicate with colleagues all over the world and to share information in the context of open-access platforms. Our results confirm previous findings suggesting the growing dominance of social media in the medical health domain [18,19]. Social media have previously been defined to include both “professional physician platforms” as well as encyclopedic open-access resources such as Wikipedia [19]. While creation and usage of a specialized medical Wikipedia appears like an interesting solution [20], the success of such projects has been mixed [19]. A differentiated approach may be necessary. We found that physicians in training are the most likely subgroup to use Wikipedia while general practitioners are most likely to use physician communities.

Medical professors reported the least reliance on collaborative and data selection tools and primarily used tools such as self-stored compendia, which can aid the preparation of presentations or manuscripts. A possible explanation could be that medical professors are more proficient in online data selection due to high levels of research experience and consequently require less help from other physicians.

An interesting finding was that despite language not being mentioned as an explicit search barrier, and self-perceived understanding of medical English being high, many physicians, especially general practitioners, reported restricting their search results by language. Thus, it appears that many physicians residing outside the English-speaking domain are interested in local information in their mother tongue. Possible explanations could be a preference to read their mother tongue and increased relevance of local resources to their health care settings.

Strengths and Weaknesses of the Study

At the time of this writing and to our best knowledge, the current study is based on the largest sample size of all detailed European scientific studies on the professional use of the Internet by medical practitioners. With regard to the large number of questions answered by each participant, it is one of the most extensive studies in the field. A potential weakness of the study is that it is biased against medical practitioners who use the Internet rarely, since the questionnaires were primarily disseminated via email, and the online promotion and the questionnaire were available only through a website. Another

potential weakness is that the study is solely based on self-report, while physicians have been found to overreport their use of objective resources and underreport their reliance on subjective sources such as colleagues [17]. Third, physicians from certain European countries (Austria, Switzerland, United Kingdom, and France) and physicians living in urban areas are overrepresented in the sample.

Possible Explanations and Implications for Clinicians and Policy Makers

General search engines are popular but offer vast amounts of often irrelevant, invalid information that require physicians to pursue substantial data screening and filtering. Appropriate tools may help to overcome this problem. In terms of usability, simple search systems such as Wikipedia and Google appear to attract physicians, possibly due to time constraints but also due to their level of information need. A substantial proportion of physicians seek secondary resources, struggle with limited access to readable, high-quality content, and lack the time and skill to pursue relevant data selection. Since the quality of medical information accessed on the Internet is likely to increasingly impact medical care, it may be of importance that all physicians are trained in efficient online data selection and have easy accessibility to high-quality resources. A potential solution might also be governmental support of the development of openly available, effective medical search engines, open access to high-quality content for physicians and the improvement of the quality of popular existing Web resources such as Wikipedia.

Unanswered Questions and Future Research

Our research provides the investigative basis for questions that need to be answered in further experimental research. The finding that most clinicians use Wikipedia and general search engines to obtain medical information raises the question of the impact of such behavior on the quality of medical care. It also needs to be investigated why physicians are largely unwilling to pay for high-quality medical information services, and what alternative financing models for such services could look like. It is unknown to what extent different types of physicians have the expertise to evaluate the reliability of content provided on the Web. Level of initial knowledge or expertise may play a central role in determining data selection skills and the type of resources expected when searching for information. The extent to which initial knowledge of a medical topic has an impact on the type of medical resource used could be clarified in experimental research. Further quantitative research representing different health systems, including a higher proportion of younger physicians and comparing different areas of medical specialization, is needed.

Acknowledgments

We thank Celia Boyer for helpful discussions while devising this study.

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 257528 (KHRESMOI) and the Austrian Science Fund (FWF): [PP 25608-N15].

Conflicts of Interest

None declared.

Multimedia Appendix 1

Physician questionnaire (English).

[PDF File (Adobe PDF File), 295KB - [jmir_v15i6e122_app1.pdf](#)]

Multimedia Appendix 2

Tables.

[PDF File (Adobe PDF File), 394KB - [jmir_v15i6e122_app2.pdf](#)]

References

1. Masters K. For what purpose and reasons do doctors use the Internet: a systematic review. *Int J Med Inform* 2008 Jan;77(1):4-16. [doi: [10.1016/j.ijmedinf.2006.10.002](#)] [Medline: [17137833](#)]
2. Younger P. Internet-based information-seeking behaviour amongst doctors and nurses: a short review of the literature. *Health Info Libr J* 2010 Mar;27(1):2-10. [doi: [10.1111/j.1471-1842.2010.00883.x](#)] [Medline: [20402799](#)]
3. Magrabi F, Coiera EW, Westbrook JI, Gosling AS, Vickland V. General practitioners' use of online evidence during consultations. *Int J Med Inform* 2005 Jan;74(1):1-12. [doi: [10.1016/j.ijmedinf.2004.10.003](#)] [Medline: [15626631](#)]
4. Casebeer L, Bennett N, Kristofco R, Carillo A, Centor R. Physician Internet medical information seeking and on-line continuing education use patterns. *J Contin Educ Health Prof* 2002;22(1):33-42. [doi: [10.1002/chp.1340220105](#)] [Medline: [12004639](#)]
5. Podichetty VK, Booher J, Whitfield M, Biscup RS. Assessment of internet use and effects among healthcare professionals: a cross sectional survey. *Postgrad Med J* 2006 Apr;82(966):274-279 [FREE Full text] [doi: [10.1136/pgmj.2005.040675](#)] [Medline: [16597816](#)]
6. thinkhealth. Connecting with physicians online. 2009. URL: <http://www.fdashm.com/docs/Connecting%20with%20Physicians%20Online%20Webinar%20Deck--%20FINAL.PDF> [accessed 2012-10-31] [WebCite Cache ID 6BpEG1Srm]
7. Shaughnessy AF, Slawson DC, Bennett JH. Becoming an information master: a guidebook to the medical information jungle. *J Fam Pract* 1994 Nov;39(5):489-499. [Medline: [7964548](#)]
8. Tang H, Ng JH. Googling for a diagnosis--use of Google as a diagnostic aid: internet based study. *BMJ* 2006 Dec 2;333(7579):1143-1145 [FREE Full text] [doi: [10.1136/bmj.39003.640567.AE](#)] [Medline: [17098763](#)]
9. Hersh WR, Hickam DH. How well do physicians use electronic information retrieval systems? A framework for investigation and systematic review. *JAMA* 1998 Oct 21;280(15):1347-1352. [Medline: [9794316](#)]
10. Younger P. The effective use of search engines on the Internet. *Nurs Stand* 2005;19(48):56-64; quiz 66. [Medline: [16117270](#)]
11. Bennett NL, Casebeer LL, Kristofco RE, Strasser SM. Physicians' Internet information-seeking behaviors. *J Contin Educ Health Prof* 2004;24(1):31-38. [doi: [10.1002/chp.1340240106](#)] [Medline: [15069910](#)]
12. Ely JW, Osheroff JA, Chambliss ML, Ebell MH, Rosenbaum ME. Answering physicians' clinical questions: obstacles and potential solutions. *J Am Med Inform Assoc* 2005;12(2):217-224 [FREE Full text] [doi: [10.1197/jamia.M1608](#)] [Medline: [15561792](#)]
13. Anthes DL, Berry RE, Lanning A. Internet resources for family physicians. *Can Fam Physician* 1997 Jun;43:1104-1113 [FREE Full text] [Medline: [9189299](#)]
14. Health on the Net Foundation. URL: <http://www.hon.ch/> [accessed 2013-03-07] [WebCite Cache ID 6EwKzrSdA]
15. Norman G. Likert scales, levels of measurement and the "laws" of statistics. *Adv Health Sci Educ Theory Pract* 2010 Dec;15(5):625-632. [doi: [10.1007/s10459-010-9222-y](#)] [Medline: [20146096](#)]
16. Moge N. Learning technology dissemination initiative. So you want to use a Likert scale URL: <http://www.icbl.hw.ac.uk/ltidi/cookbook/cookbook.pdf> [accessed 2013-03-07] [WebCite Cache ID 6EwM6X1j1]
17. Covell DG, Uman GC, Manning PR. Information needs in office practice: are they being met? *Ann Intern Med* 1985 Oct;103(4):596-599. [Medline: [4037559](#)]
18. Johns Hopkins Medicine. ScienceDaily. Dec 7. One in four physicians uses social media daily URL: <http://www.sciencedaily.com/releases/2012/12/121207225028.htm> [accessed 2013-02-05] [WebCite Cache ID 6ED5sqaiM]
19. von Muhlen M, Ohno-Machado L. Reviewing social media use by clinicians. *J Am Med Inform Assoc* 2012;19(5):777-781 [FREE Full text] [doi: [10.1136/amiajnl-2012-000990](#)] [Medline: [22759618](#)]
20. Heilman JM, Kemmann E, Bonert M, Chatterjee A, Ragar B, Beards GM, et al. Wikipedia: a key tool for global public health promotion. *J Med Internet Res* 2011;13(1):e14 [FREE Full text] [doi: [10.2196/jmir.1589](#)] [Medline: [21282098](#)]

Edited by G Eysenbach; submitted 13.11.12; peer-reviewed by P Younger, A Roberts; comments to author 25.01.13; accepted 17.04.13; published 26.06.13.

Please cite as:

Kritz M, Gschwandtner M, Stefanov V, Hanbury A, Samwald M

Utilization and Perceived Problems of Online Medical Resources and Search Tools Among Different Groups of European Physicians
J Med Internet Res 2013;15(6):e122

URL: <http://www.jmir.org/2013/6/e122/>

doi: [10.2196/jmir.2436](https://doi.org/10.2196/jmir.2436)

PMID: [23803299](https://pubmed.ncbi.nlm.nih.gov/23803299/)

©Marlene Kritz, Manfred Gschwandtner, Veronika Stefanov, Allan Hanbury, Matthias Samwald. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 26.06.2013. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Viewpoint

Community Management That Works: How to Build and Sustain a Thriving Online Health Community

Colleen Young¹, BA

Canadian Virtual Hospice, Toronto, ON, Canada

Corresponding Author:

Colleen Young, BA

Canadian Virtual Hospice

30 Roseneath Gardens

Toronto, ON, M6C 3X6

Canada

Phone: 1 416 651 4121

Fax: 1 204 475 1497

Email: colleen@colleenyoung.com

Abstract

Health care professionals, patients, caregivers, family, friends, and other supporters are increasingly joining online health communities to share information and find support. But social Web (Web 2.0) technology alone does not create a successful online community. Building and sustaining a successful community requires an enabler and strategic community management. Community management is more than moderation. The developmental life cycle of a community has four stages: inception, establishment, maturity, and mitosis. Each stage presents distinct characteristics and management needs. This paper describes the community management strategies, resources, and expertise needed to build and maintain a thriving online health community; introduces some of the challenges; and provides a guide for health organizations considering this undertaking. The paper draws on insights from an ongoing study and observation of online communities as well as experience managing and consulting a variety of online health communities. Discussion includes effective community building practices relevant to each stage, such as outreach and relationship building, data collection, content creation, and other proven techniques that ensure the survival and steady growth of an online health community.

(*J Med Internet Res* 2013;15(6):e119) doi:[10.2196/jmir.2501](https://doi.org/10.2196/jmir.2501)

KEYWORDS

online community; virtual community; health; patient; health care professional, community management; community strategy; guidelines; tutorial; peer-to-peer health

Introduction

In her keynote address to participants of Medicine 2.0'11, Susannah Fox pointed out that “Patients and caregivers know things—about themselves, about each other, about treatments - and they want to share what they know to help other people” [1]. Thanks to the social Web, people facing a new diagnosis, undergoing treatment, or living with chronic illness can tap into larger networks. People are increasingly joining online health communities to share information and find support [2]. Fox went on to say, “Pew Internet research—and technology innovators—have found that if you can enable an environment in which people can share, they will and the benefits will entice others to join” [1].

Recognizing the potential benefits and increasing popularity of online patient support [2-7], more and more health care

institutions and organizations are adding social Web elements to their repertoire of patient and family support and care provider collaboration tools. But social Web (Web 2.0) technology alone does not create a *successful* online community. A successful community is one in which members participate actively and develop lasting relationships [8]. Building and sustaining a successful community requires an enabler [9] and strategic community management [8,10].

This paper describes the community management strategies, resources, and expertise needed to build and maintain a thriving online health community; introduces some of the challenges; and provides a guide for health organizations considering this undertaking. The tutorial draws on insights from an ongoing study and observation of online communities as well as experience managing and consulting a variety of online health communities (see [Table 1](#)).

Table 1. Online communities built and used in this research.

Community	Description	Date of inception	Number of members ^a	Stage
SharingStrength/ FortesEnsemble [11]	A Canadian online resource library and community for women with breast cancer	Mar. 2007	1050	Mitosis (archived & adopted)
Health Care Social Media Canada [12]	A community of practice for people interested in exploring social innovation in health care	Sept. 2010	6564	Maturity
CancerConnection/ ParlonsCancer [13]	Canadian Cancer Society's online community for people touched by cancer (H. Sinardo community manager)	Jan. 2011	2100	Establishment
Canadian Virtual Hospice/ Portail canadien en soins palliatifs [14]	Online resources and community for people living with limited time, losing someone, caring for someone, or working through grief	Dec. 2011	477	Late Inception

^aMembership numbers as of Feb. 28, 2013.

At the time of writing this paper, the author (CY) manages the online communities of Health Care Social Media Canada (hcsmda) [12] and Canadian Virtual Hospice/Portail canadien en soins palliatifs [14]. Hcsmda was founded September 2010 and has since grown to become a mature community of practice of over 6500 members from all sectors of health care (health care professionals and professional associations, patients, caregivers, patient organizations, public health institutions, health educators, researchers, policy makers, communicators, and many more). Unlike the other communities described in this paper, hcsmda relies on social media platforms such as Twitter and LinkedIn for its online interactions rather than online community software.

Virtual Hospice's online community was first introduced in 2004, but after the initial launch, the discussion forums languished with little to no activity. In 2011, the author joined Virtual Hospice's team to develop and implement a strategic online community management plan and establish a successful community. Today, the fledgling, but active, community continues to grow, providing peer-to-peer support and information for a very specific point in the health continuum, namely for people living with life-threatening disease, for friends and family who care about and for them, and for people dealing with grief and loss (see Figure 1).

SharingStrength/FortesEnsemble was a Canadian online resource library and community for women with breast cancer. The

website was archived in 2011 [11] and the thriving English and French communities were successfully adopted by the Canadian Cancer Society's then new online communities CancerConnection and ParlonsCancer [13] for people touched by cancer. During the adoption phase, the CancerConnection/ParlonsCancer communities were a welcoming new home and the presence of SharingStrength/FortesEnsemble's moderator (the author) helped ease the transition of SharingStrength/FortesEnsemble members. Additionally, SharingStrength/FortesEnsemble's community management practices and experience were transferred through collaborative moderator team training, the development of a moderation manual, and the creation of content for and with the community. Strategic development of CancerConnection/ParlonsCancer continues. Sustained collaboration between the community managers of CancerConnection/ParlonsCancer and Virtual Hospice helps members who are entering palliative care, or are caring for someone at the end of life, and who may wish to access the resources and services at Virtual Hospice or join that community.

Throughout the paper, the term "online community" is used; some researchers use "virtual community". While the terms are interchangeable, online community is more widely used [15].

Figure 1. Virtual Hospice forum activity (captured April 2, 2013).

CANADIAN Virtual Hospice
Information and support on palliative and end-of-life care, loss and grief.

French | Glossary | Contact Us | eNews | Login | Text Size: A A A A

Search

Home | About Us | Partners | Publications

Topics | **Support** | **For Professionals**

Ask a Professional
Asked and Answered

Discussion Forums
The Gallery

Your Stories
Meet the Team

You are In: Home | Support | Discussion Forums | Just Want to Talk

Discussion Forums

Search Discussions

Start New Topic

« 1 2 3 4 5 6 7 8 9 10 11 12 »

Topics	Messages	Last Message
Going from WE TO ME..... Started by nanalovesu On 26 Jun 2012, 2:01 PM	82	02 Apr 2013, 1:55 AM by: marstin
I am getting so burnt out....help!? Started by lilbear On 21 Mar 2013, 10:36 PM	48	01 Apr 2013, 8:36 PM by: lilbear
Feeling Overwhelmed by Dad's diagnosis Started by Yankeegirl79 On 01 Apr 2013, 1:46 AM	3	01 Apr 2013, 8:09 PM by: marstin
Dad won't let us help Started by NiteLad On 24 Feb 2013, 12:10 AM	33	01 Apr 2013, 4:09 PM by: moderator modératrice
How to deal? - my mother is dying Started by Sad1 On 23 Mar 2013, 5:34 AM	8	01 Apr 2013, 12:39 PM by: mzmajo
Husband with Cancer - Nurturer vs. Independent Started by debbied2007 On 17 Jan 2013, 2:10 PM	31	31 Mar 2013, 3:48 PM by: marstin
How do all of you deal with it? Started by frustrated On 27 Feb 2013, 1:43 PM	17	27 Mar 2013, 1:46 PM by: NatR
Losing my Mom - Taking this so hard Started by andreadzel On 12 Jan 2013, 1:36 AM	28	23 Mar 2013, 8:35 PM by: mzmajo
Remembering CarolynMarie Started by moderator modératrice On 16 Nov 2012, 5:31 PM	22	23 Mar 2013, 2:26 PM by: moderator modératrice
My mom and dear friend are both dying! Started by Cindie 46 On 12 Jan 2013, 1:24 AM	16	23 Mar 2013, 12:57 PM by: moderator modératrice

« 1 2 3 4 5 6 7 8 9 10 11 12 »

[Back to Top](#)

Topics | Support | For Professionals

Copyright 2003-2013 Canadian Virtual Hospice. All Rights Reserved.

Sitemap | Privacy Policy | Legal Notices

Login

Email:

Password:

Forgot your password? [Join today.](#)

Ask a Professional

Our team of experts answers your questions about life-threatening illness and loss.

Meet the Team

Meet the experts who answer your questions at Ask a Professional.

Books, Links, and More

Recommended by our team

Programs and Services

Find local, regional, and national services

Most Popular Articles

When Death is Near - Learn more about changes people may experience in the final days of life.

Health Care Directives - Having a health care directive can ensure treatment decisions are respected.

Considerations for a Home Death - How you can prepare to provide care at home.

Lack of Appetite - What you can do when someone loses interest in food.

Feet aren't ugly

A remarkable story of how one hospice volunteer connects with patients and families through their feet.

What Is an Online Community?

As yet, there is no universally accepted definition of online community. Howard Rheingold, who coined the phrase virtual community, describes them as “cultural aggregations that

emerge when enough people bump into each other often enough in cyberspace” [16]. A variety of disciplines have studied online communities, each one providing its own definition. Reviews by Preece and Maloney-Krichmar [17] and Iriberri and Leroy [8] illustrate how broadly defined online community is in this

rapidly growing field. While Preece defines online community as “any virtual social space where people come together to get and give information or support, to learn, or to find company” [15], in this paper, an online community is a group of people who share a strong common interest, form relationships, and interact online. This definition encompasses three of the elements required for a successful online community discussed later in this paper:

- share strong common interest = domain
- form relationships = sense of community
- interact online = activity

There are different types of online communities [18]. People may form a community around a common interest, place, action, practice, or circumstance. While many of the strategies and tactics outlined in this discussion apply to all types of communities, specific examples focus on health communities. (Multimedia Appendix 1 lists all the online health communities mentioned in this paper.) Most health- and disease-based communities can be described as communities of circumstance. Rather than seeking online social interaction about a chosen interest, such as gardening (GardenWeb [19]) or citizenship (BritishExpats [20]), patients and caregivers are motivated to seek online interaction when a circumstance is imposed upon them, such as a new diagnosis or a change in health or well-being. They want to learn about their disease, get support, be less afraid of the unknown, and help others in similar circumstances [18]. Community members who have had similar experiences and can respond empathetically [21] may encourage strong relationships to develop, making patient and caregiver communities some of the most important on the Internet [18].

Why Some Communities Succeed and Others Fail

Some online health communities thrive, sustaining activity for many years. Others languish, resembling ghost towns [22]. Their asynchronous tools, such as discussion forums (also known as message boards), often have multiple threads with few messages and last-post dates that are long past. Synchronous online chats and events are not well attended. Why do some communities succeed where others fail?

The success or failure of an online community depends, in part, on an organization’s commitment to sustained organizational and financial support for dedicated community management. Many health organizations are most concerned about the proliferation of misinformation [18]. While monitoring for misinformation through moderation and community watch mechanisms is important [3,23], developing and sustaining a successful online health community requires more than reactionary observation.

To build a thriving online community, organizations must ensure they have organizational commitment and the financial and human resources to not only start an online community, but also to support its growth and to evolve with the community throughout its life cycle. There are many flourishing online health communities. Before deciding whether to start an online community, resource-strapped health organizations should

perform an environmental scan and consider whether their proposed community differs from those that exist or whether it makes more sense to seek collaborative opportunities with an established community. Investment in community management is imperative to an online community’s success [24]. Furthermore, an organization needs to

- establish and understand the domain of the proposed community [9]
- develop and sustain a community management strategy according to the community’s life cycle stage [8]
- foster a sense of community [25]

The Domain

The *domain* is the common ground and sense of common identity upon which a community is built. Clearly defining the domain—whom the community is for, and why—affirms its purpose and value to both the community members and other stakeholders. The domain inspires people to participate in and contribute to the growth of the community [9]. In the case of health communities, domain may simply be a disease, such as cancer (CancerConnection [13]) or Crohn’s disease (Crohnology [26]), or life stage, such as birth (BabyCenter [27]) or death (VirtualHospice [14]).

An online community’s purpose and membership motivation also help determine which management strategies and platform tools will most likely contribute to its success [8,28]. In the case of online health communities, the purpose is usually to provide support and information to help people manage their health and well-being. Given that members are motivated by their need to receive emotional support, to reciprocate support, to learn, and to offer practical information, discussion forums are often the first platform tool introduced to a community platform. As the community evolves, the community manager may add tools that help members build a knowledge base or aggregate the community’s collective knowledge, such as a wiki to develop an e-book, group areas focused on a particular topic, or blogs led by specific community members.

The Community Life Cycle

Reviewing online community research, Iriberry and Leroy recognized that online communities evolve through distinctive life cycle stages: inception, creation, growth, maturity, and, sometimes, death [8]. Iriberry and Leroy developed a framework of recommendations for success relevant to the developmental stage of the online community. This framework has been widely adapted and refined by community management practitioners to build successful online communities and will be discussed in more detail later in this paper.

Sense of Community

Successful communities possess a strong sense of community, which comprises four elements [29]:

- Membership: Individuals have a feeling of belonging to, and identify with, the community.
- Integration and fulfilment of needs: The goals of individuals match those of the membership as a whole. As members satisfy their own needs, they also meet the community’s needs.

- **Influence:** Members feel they matter within the community and that they can influence and be influenced by the community.
- **Attachment:** Members share an emotional connection. Members believe they share or will share common history, places, or experiences.

Historically, measures of the sense of community in online communities have been adapted from McMillan and Chavis's widely used measure of sense of community for face-to-face communities [30]. Evaluating these four elements in an online community offers insight into whether the online interaction can be defined as merely an online "settlement" [25], such as comments on a Facebook fan page, or a true online community, where members display actions reflecting feelings of belonging, influence, and attachment, and both their own as well as the community's needs are being fulfilled. The Association of Cancer Online Resources (ACOR) [31] is a prime example of a true online community. ACOR is one of the oldest (established in 1995) and arguably the best known of online cancer communities, with a long history of sustained activity and a large body of reputable collective intelligence of hundreds of patients and caregivers.

While similarities exist between the sense of community of face-to-face and online communities, there are significant differences [30]. Given the increasing popularity of online communities and social networking, more research in this area

would be useful. Blanchard's Sense of Virtual Community measure represents a significant first step [30].

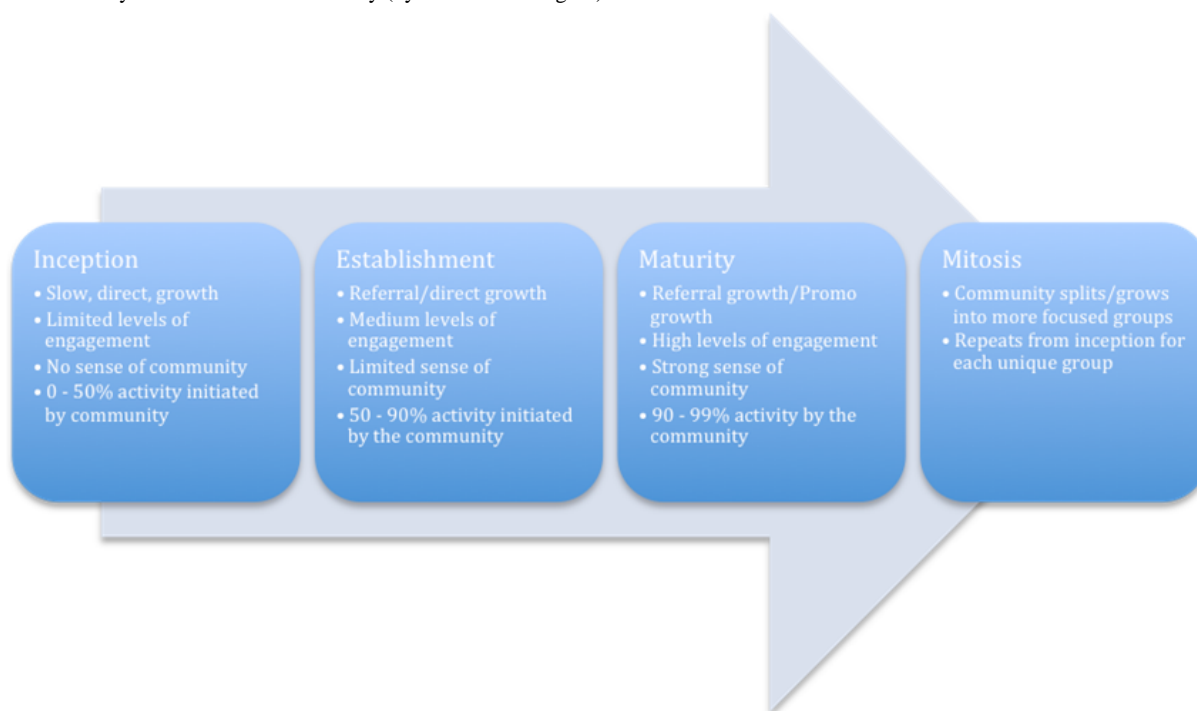
Guidelines for Building and Sustaining a Thriving Online Community

The Community Life Cycle

Depending on a community's developmental phase, its community management goals, strategies, priorities, and basic tasks will evolve. Figure 2 illustrates Richard Millington's adapted version of the life cycle of an online community based on Iriberri and Leroy's research [8,10]. This section introduces Millington's four stages and explains how they can be used to build an online health community. It also expands upon the mitosis stage, examining the possibility of arranging for the adoption of a community and how to avoid the death of a community.

Understanding a community's life cycle stage by monitoring its growth and activity (ie, posts, chats, events, and private messages) will help community managers decide when to adjust their community tactics and what changes the community would welcome and will enable them to substantiate the proposed evolutions for stakeholders. A community's life cycle is not linear and managing its growth, activity, and design is an iterative process that must adapt to the needs of the members and the community's purpose.

Figure 2. Life cycle of an online community (by Richard Millington).



Inception

The inception stage of an online community begins as soon as an organization begins to engage potential community members [32]. During the inception phase, the community manager

- develops relationships with potential members

- invites them to join and participate and encourages them to remain active
- helps establish tone and style
- recruits and nurtures an active core group to be community ambassadors

At the beginning of the inception stage, most of a community manager's time is dedicated to making connections and building a core group of active members. This work should start even before the community platform is available. Community managers seek potential members by tapping into their personal and professional networks and their organization's networks, monitoring discussions on social media, and recruiting volunteer participants. In other words, online community building starts with many one-on-one interactions, many of which occur offline. As Millington points out, "Getting your first 50 members is really hard work. It's much harder than you probably imagine. In fact, earning those first 50 people is a full-time job" [32].

Many health organizations have a volunteer corps. Volunteers can be recruited and offered training on how to use the community platform and how to foster supportive conversation online (see [Figure 3](#) for an example of a volunteer group, the "Care Team" from Tudiabetes [33]). Volunteers can seed the community—engage in conversations among themselves—to ensure there is activity before new potential members are invited to the online community. It is vital to have activity in the community, especially in health communities, because no one wants to seek support from, or ask a vulnerable question into, a void [34].

Engaged volunteers welcome new members and ensure they receive timely—ideally immediate—replies. Reducing the time between posting a message and receiving a response encourages new people to stay engaged with the community members and to become committed members [35]. Receiving a prompt reply to a comment (within minutes) is far more memorable than receiving a reply hours or days later [36]. Implementing an opt-out notification system that alerts members by email when there is a new response to their contributions promotes quick response by both new and core members and can increase participation and activity [36].

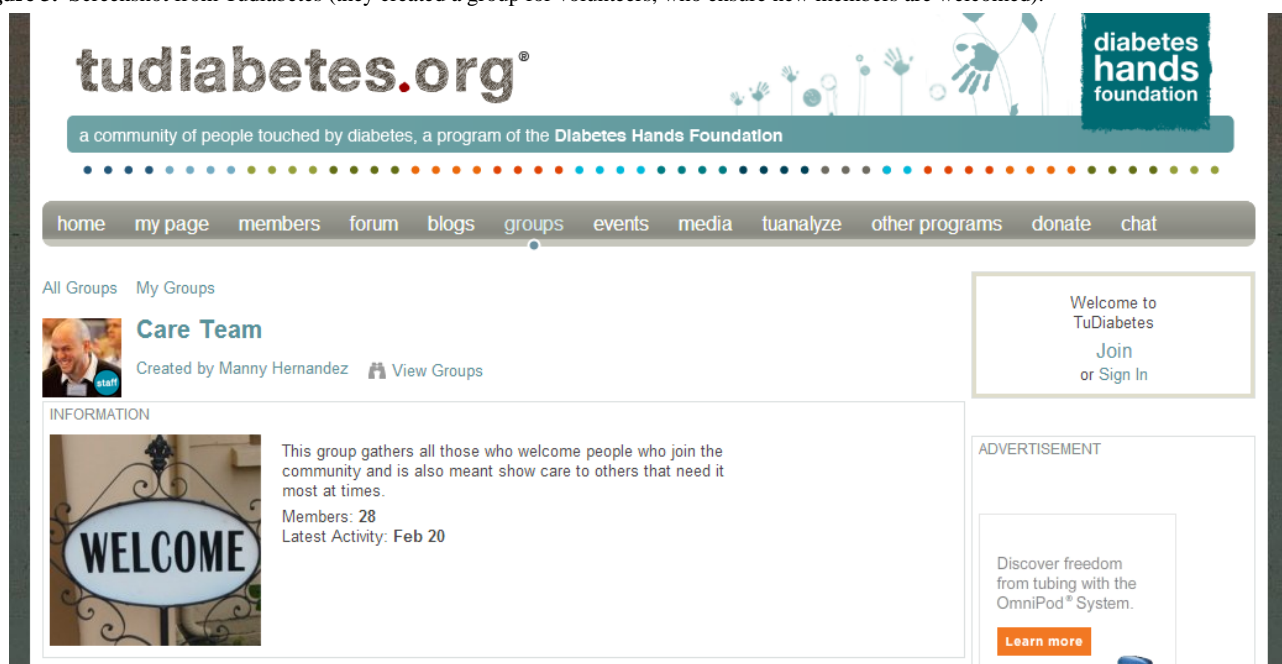
Core members, including volunteers, play a vital role in the success of a community. They provide activity, but they also establish the tone of the conversation, welcome newcomers, connect people in the community, give lurkers (people who read but do not post) the confidence to join the conversation, and invite people to the community. Core members also help make improvements to the community, such as identifying possible barriers to participation and usability problems. In the inception stage, all new members are potential core members. Community managers have the opportunity to demonstrate to early adopters that they matter and that they have influence; managers should also ensure that there are channels for members to give their feedback. In the beginning, many members join because of a direct invitation. Beyond asking them to join, community managers should guide new members as to what they can do, for example, respond to a particular thread, start a blog on a given topic, add to the profile page, etc.

During the inception stage, slow, steady growth is best. Successful community managers integrate a few members at a time and work at converting visitors into active members. Thus the goal is not to accrue a large number of registrants, but rather to steadily grow the number of members who are committed to participating [32]. As activity and membership grow, the community manager will

- observe behaviors and make adjustments to usability, design, and strategy according to member feedback and behavior
- look for on- and offline event opportunities, such as live chats or in-person meetings
- increase activity and nurture a sense of community
- moderate posts and remove or correct misinformation

As membership and activity grow, community management strategies will gradually shift to establishment-stage activities.

Figure 3. Screenshot from Tudiabetes (they created a group for volunteers, who ensure new members are welcomed).



Establishment

The establishment phase of the community's life cycle begins when members generate more than 50% of the activity and ends when they generate most (90%) of the growth and activity and when the sense of community starts to develop [32]. At this stage, the community manager focuses on modified versions of the inception-phase tasks, such as

- adding, nurturing, and supporting core members
- increasing activity with an eye to deepening the sense of community
- continuing with the growth strategy and broadening outreach
- expanding community tools

Activity remains the prime focus of the establishment phase. Supporting, observing, and investing time in the community's core members helps achieve the remaining goals of the establishment period, namely growth and a deepening sense of community [30]. As the community matures, trust and lasting relationships begin to emerge and the strengths of core members reveal themselves [37]. Community managers should capitalize on the roles that core members start to assume, assigning them ownership of these roles and rewarding them for their activity, such as by featuring them in a newsletter article or blog post or by developing a system of recognition awarding status, for example, highlighting the most helpful members or most viewed posts. In the community PatientsLikeMe [38], members receive profile stars to indicate the amount of health information they've shared (see Figure 4). The stars show other members who has shared how much. Once a member receives the maximum three stars, they can apply for mentor status. Note that the PatientsLikeMe model assumes that the more members share, the more they can understand about their own experience and presumably the more they help others to learn. However, quantity does not trump quality. A single quality post from a regular member can also have a significant positive effect on a community. Community managers may wish to highlight the best comments of the week or month to recognize influential members.

As their commitment deepens, core members become more invested in the community's success. The community manager can encourage further organic growth by developing a word-of-mouth referral plan, leveraging the commitment of these community ambassadors [32]. As Millington points out,

it is first important to understand what motivates people to recommend a community (give a referral). They may be motivated by the potential to

- increase their status within the community
- increase their status outside the community
- help others in a similar situation
- help build the community

The first two motivations can be satisfied by recognizing key members and their contributions as mentioned above. For the latter two, the community manager can implement mechanisms that cater to the "pay it forward" sentiment that is prevalent among patients and caregivers and is a key motivator for health communities. Online community members often have offline interactions with other patients and caregivers, health care providers, and support organizations. These interactions can be fertile ground for promoting the community and its benefits and for extending invitations to join. Community managers should encourage and facilitate these opportunities.

While direct invitations to join the community continue, the community manager's outreach efforts can shift to include broader awareness tactics at this stage, such as writing about the community (submitting articles to relevant newsletters and posting to social media channels and related blogs) and making presentations at relevant conferences and events.

Creating content for and about the community intensifies the sense of community. Content examples include producing a community newsletter or community section in the organization's existing newsletter, writing newsletter articles or blog posts about individual core members or inviting them to make a newsletter or blog contribution, or producing a video featuring the community and its membership.

New members and new activity bring new ideas for discussions, for new roles and responsibilities, and for the development of new tools, features, and technologies. Thus, during the establishment phase, the community manager also continues to

- moderate posts
- monitor growth and activity
- make usability, design, and strategy improvements according to member feedback and user behavior data
- create event opportunities

Gradually, the community will evolve to the maturity phase.

Figure 4. Profile of a member who has achieved mentor status on PatientsLikeMe site.

patientslikeme

Patients | Treatments | Symptoms | Research

Home > Find Patients w/ Treatments, Symptoms and Side Effects Like You > Rick N's Summary

What's New
Rick N's Journal

Charts
Rick N's Charts

About
Rick N's Summary

Rick N
Male, 58 years
Jefferson City, MO

Primary Condition: MS and 2 more ▾
Type: Primary Progressive
First symptom: Sep 1988 • **Diagnosis:** Jun 1996
Interests: Advocacy, and Faith

About Rick N
For so very long I have lived with Multiple Sclerosis, its burdens, , confusion , losing my mind , fear. The journey has had it's peaks and valleys , but there is a hidden realization that "all is well". A divine relationship with God comforts me in a way that did not exist before. It is a blessing that chisels character. I am a better person in spite of all the trials - Something that words can't express. After so many years with this disease, the values in life that I once thought were important are not. I live on Faith and Hope. And Win!

Interests
Advocacy, and Faith

Mentor
See areas of experience ▾

- Managing medications
- Managing injection medications
- Managing symptoms
- Working with MS
- Choosing a doctor or other health care provider
- Relationships, marriage, and sexuality
- Understanding a recent diagnosis
- Coping without a diagnosis

Profile Activity
24,948 Views

My CareTeam

Member since: Nov 27, 2008 **Last Login** Mar 05, 2013

Other Conditions
Generalized Anxiety Disorder
First symptom: Jan 2011
Diagnosis: Sep 2011
Pancreatitis
First symptom: Feb 2012
Diagnosis: Dec 2012

Maturity

An online community's maturity stage begins when more than 90% of activity and growth are generated by members [32]. Many successful communities oscillate between the establishment and maturity phases as members retire from the community and community management strategies circle back to inception and establishment phase tactics to spawn new activity with new members. In the maturity stage, the size of the community reaches its critical mass, activity continues, and a sense of community is well established. While mature communities are often considered self-sustaining, the need for community management strategy and activity remains. The community manager's attention turns to

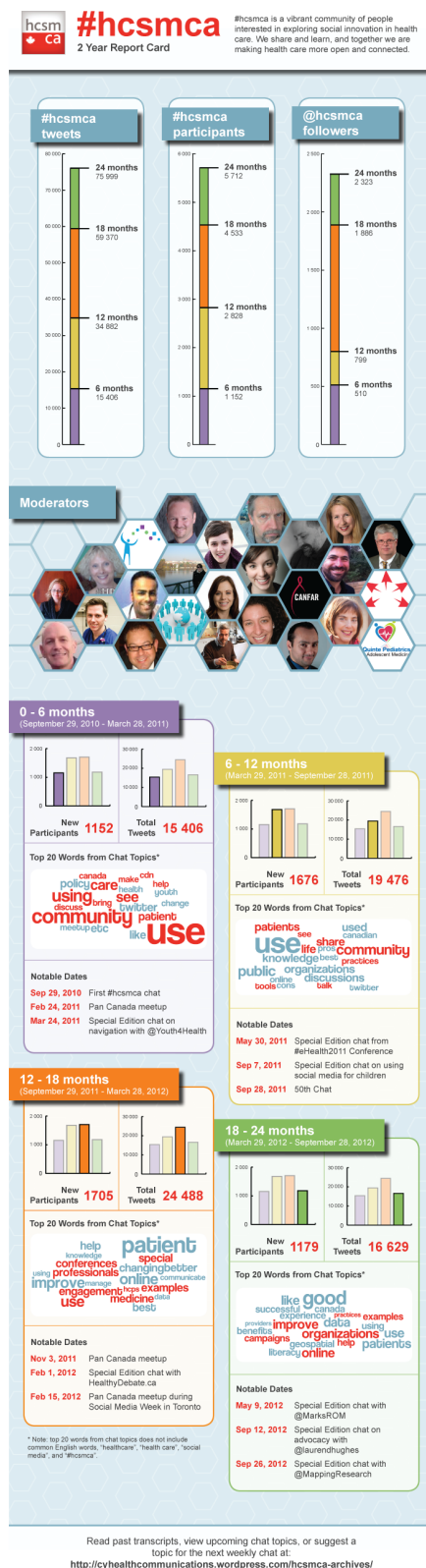
- training core members to assume roles to maintain activity
- stimulating referrals and promoting the community to new members
- intensifying the sense of community
- assessing and optimizing processes
- developing collective value

The community manager should monitor the community and identify opportunities to develop collective value, which in turn

intensifies the sense of community and will help promote the community. Collective value can be developed through co-creation, that is, involving community members to produce something together, such as a community charter or an e-book about the community or on a subject that represents the collective's area of expertise. Members could be asked to write content for third-party publications or to respond to a survey to gather feedback on proposed improvements, giving them ownership of the changes. In the case of hcsma, which is a mature community, members have been invited to become more involved, and some have taken initiative to create new roles and participation in co-creation of documentation and activities, for example:

- moderating weekly online chats
- organizing in-person meetings (meet-ups)
- writing blog posts about their experience in the community or about community activities
- analyzing the community (see Figure 5)
- creating documentation (eg, archiving of chat transcripts)

Many communities remain in the maturity stage for years and, if they maintain sustainable member size, may never reach mitosis or death [8,10].

Figure 5. Infographic created by a hcsma community member about the community.

Designed by @AlainaBCyr | Data collection by @Colleen_Young and @AlainaBCyr using data tracked by The Healthcare HealthTag Project at symplr.com and from the word cloud generator Wordle.net. Data from September 2010 to October 2012.

Mitosis, Adoption, Death

The mitosis phase begins when the community becomes largely self-sustaining and ends when activity and growth start to erode the sense of community [32]. This is a critical juncture in the

community life cycle and community managers need to monitor activity and member attitudes and behaviors to ensure that the community does not become a victim of its own success. When a community grows to a point where the majority of members feel it is too large and too active, they will start to think they

can no longer influence the community and they disengage. If the community manager does not do enough to mitigate this situation, the community will quickly die.

With proper monitoring and data collection during the establishment and maturity phases, however, the community manager will observe special interest groups emerging and will be able to create splinter groups or community subsets [10]. For example, a community for women with breast cancer may create a subcommunity of young women with breast cancer. While the word mitosis implies division into identical cells, a splinter community will develop its own culture and characteristics, not necessarily duplicating those of the parent community.

Each splinter community will return to the inception phase and progress through the life cycle. Because the community management efforts for each splinter community equal the effort of starting a new online community, initiating one splinter group at a time is advised [32].

Millington's adapted version of Iriberri and Leroy's online community cycle appears most applicable. It is practical and quite comprehensive; however, "Adoption" should be added to the life cycle for organizations that can no longer fund or otherwise support their thriving online community. Sometimes burgeoning or thriving online communities are forced to die a premature death when funding priorities shift or dry up despite the success of the community. In such cases, planned adoption can support the survival of the community of people. SharingStrength/FortesEnsemble is one example of successful adoption of a community when the funder's priorities were redirected [11].

Like Iriberri and Leroy, Millington refers to the possibility of the final life stage: death. Understandably, however, his practice guidelines do not focus on this stage because he contends that good community management practices render this stage avoidable [39]. He writes, "You have a short window to reverse a decline before it becomes a death spiral". Community managers can avoid the death spiral by monitoring metrics of growth, activity, and sense of community, in other words, measuring for success.

Measure for Success

Metrics for quantifying the success of online communities vary widely [15,40]. In her paper from 2001, Preece suggests tracking determinants that measure community members' sociability and the usability of the technology to measure success. Sociability refers to how members of a community interact with each other on the community platform or technology; usability is primarily concerned with how users interact with the technology. Analysis of a community at key points in its life cycle using rigorous determinants is highly recommended. For example, Preece outlines the evaluation of the purpose, people, and policy of a community (sociability) and information design, navigation, access, the dialogue and social supports like prompts, use of avatars, etc (usability) [15]. However, collecting such thorough data regularly is likely unrealistic for most community managers.

One can also apply many useful health-related metrics, such as self-efficacy and quality of life, to analyze online communities. For the purpose of this paper, the focus is on metrics specifically pertinent to the community manager's role in developing a thriving online community. This role includes ascertaining the community's needs and identifying relevant community trends and developments to steward the health of the online community—its growth, activity, and sense of community—and to foresee and fix negative trends before they become problems. Thus, it is important to monitor the trends and to gather data that will

- keep track of the growth of the community
- demonstrate activity and engagement (sociability)
- improve the community, discover problems, and validate what works (sociability and usability)
- report progress and demonstrate the value of the community to stakeholders, including community members

Community managers should collect a manageable amount of data regularly, consistently, and accurately over the life of their community. Millington recommends tracking active members to determine growth and engagement [41]. Growth data should include not only the number of registrants (members), but also the number of registrants who contribute (active members) and the number who made a contribution in the past month. Having a high number of registrants relative to the number of participants signals a low conversation rate and may indicate the existence of barriers to participation (usability). Knowing who has not contributed in the past month can inform targeted outreach activities.

Activity and engagement data help determine where the community is in its life cycle (a more engaged community equals a more developed community) and identifies potential problems early, when they can be more easily corrected. Activity and engagement can be assessed by monthly tracking of the number of posts, the average number of contributions per active member, the average number of responses to a post, the average time for a post to receive a response, and the average number of visits per active member [10]. By identifying the activity and engagement level, the community manager can validate the life cycle stage and introduce strategies and tactics appropriate to that stage. Managers can use the data to recognize the type of conversations and activities that generate the highest engagement, which member behaviors are most apt to lead to increased activity, and whether too few members are dominating the conversation.

Table 2 lists examples of growth and activity data that community managers could track monthly to monitor the health of their community. The data can be used to track trends and progress, to help identify which community management strategies and activities are working, and to improve those that are not. These data can also be used to develop tactical activities to promote growth and activity tailored to the community's members, potential members, and technology.

Table 2. Examples of growth and activity data that community managers could collect.

Measurement (monthly)	Key questions
Number of first-time visitors ^a to community	Are people finding the community? What outreach tactics can be used, or technology optimizations made, to increase the number of visits?
Number of new members (registrants) ^b	What is the conversion rate from visitor to member? Is there a usability barrier to registration? Is the platform optimized to motivate visitors to become members? Are outreach tactics attracting the kind of people suited to the community?
Number (or percentage) of active members ^c who made a contribution (post)	What is the conversion rate from registrant to active member? What motivates people to participate (high conversion rate)? Why are people motivated to register, but not to participate (low conversion rate)?
Number of new active members	What influences are successfully motivating new members to participate and then to become active members?
Number of returning active members	Are an increasing number of regular members remaining active? Why or why not?
Total number of active members	Have new members become active members? Why or why not?
Total new posts	Is activity consistently increasing? Where is activity greatest (eg, discussion forums, blogs, groups, polls)?
Average number of contributions per active member	Is this number increasing or decreasing? Should more effort be dedicated to activity of existing members rather than growth? What activities are contributing to increased activity or not?

^aVisitor = someone who has visited the community but has neither registered nor contributed.

^bMember (registrant) = someone who has registered with the community but has not yet made a contribution.

^cActive member = someone who has made a contribution within a determined period of time (eg, past month).

Periodic sample measurements of the community can also be useful. For example, calculating the number of contributions per active member over a given period can demonstrate whether the activity is shared among many or few members, identify the community leaders, and discover who is contributing infrequently. Speed of response, how much response, and what kind of response should be monitored. As mentioned earlier, the quicker that members receive a response, the higher the sociability of the community, which usually leads to more activity [35].

It can be useful to know who is participating in the community. However, gathering all demographic information during the registration process is ill advised—doing so can be a barrier to participation. Instead, members who fill out their profile can be used as a measurement of engagement and sense of community, ie, a measurement of initial successful engagement. Members who have not completed their profile provide an opportunity for community managers or recruited volunteers to contact them and obtain valuable feedback about why they have not completed their profile. Perhaps there is a usability issue or the member needs encouragement or mentoring to feel part of the community.

One could also consider surveying the community to assess its sense of community. As mentioned, Chavis et al developed a sense-of-community index, which Blanchard adapted for online community use in 2007 [30]. But as Millington points out, there is a limit to how often members will agree to be surveyed. Furthermore, members who voluntarily respond to surveys are often the most engaged community members and those who feel the strongest sense of community [10]. Such surveys will likely produce biased results. Nonetheless, community managers should find ways to gather feedback from members and ensure members have clear and simple channels to provide it.

Challenges

Many health organizations are most concerned about disclosure of personal health or other sensitive information and the proliferation of misinformation [3,18,23]. However, experience and observation show that clear policies, proactive community management, and active moderation and community participation render these concerns largely unfounded for online communities associated with reputable organizations. It should be noted that some communities exist for the sole purpose of encouraging unhealthy behaviors, such as pro-anorexia groups [42]. This discussion refers to online communities that support healthy behaviors and how they manage transgressions of their terms of use.

Clearly stated policies make it easy for moderators to modify—and in some cases remove—posts that contravene terms of use, such as commercial postings, advertisements, or impersonations; posts that relate to illegal activity; those that contain disrespectful language, etc [43,44]. For example, Macmillan Cancer Support's online community had to deal with a member posing as a cancer patient. In a blog post [45], the moderation team described the transgression to the community, expressed empathy for the upset it may have caused, explained how members can protect themselves, and gave an opportunity for the community to discuss the situation.

Community managers, moderators, and core members model behavior and can guide members who may have unwittingly shared sensitive information or misinformation. This modeling establishes and maintains the desired tone of a community. Communities with a secure sense of community can rely on responsive self-policing to correct misguided behavior and misinformation. In fact, rather than removing misguided information, allowing and enabling community members to

correct misconceptions and provide balanced debate can be a very productive bonding opportunity that deepens the sense of community and establishes the value of collective knowledge. On hcsma's LinkedIn Group page, a marketing manager made a commercial post about her company's upcoming patient experience conference that did not include patients. This contravened the community's principle of including patients. The ensuing conversation demonstrated hcsma's community cohesiveness, resulted in an informative discussion, and deepened the community's sense of purpose and influence [46]. Because the hcsma community relies on Twitter for the majority of its online interactions, one might think it would be more susceptible to abuse with little recourse to correct misbehavior. However, the tight-knit nature of the community and its unified understanding of the community's purpose guards its principles and guides the behavior of newcomers, quickly correcting or rejecting misuse.

Clinical study recruitment may be desirable in some health communities. If so, guidelines and criteria about how, what, and where to post for recruitment should be readily available for researchers to consult. Recruitment policies may change as a community matures. For example, a request for photo subjects was posted on Virtual Hospice when the community was just starting out [47]. The post was removed and the poster, who had no interest in becoming a community member, was invited to submit the request through more appropriate channels of the organization. In a more mature community, such a request may not have been inappropriate.

As these examples demonstrate, undesirable behavior does happen in online communities, but responsive community management can maintain the integrity, reliability, and value of the collective community knowledge.

The bigger issue challenging the success of online communities is the failure to recognize the time and effort required to build a thriving and reputable online health community. Building a community takes organizational commitment as well as sustainable financial and human resources throughout the community's life cycle [48]. An evidence-informed community management strategy [49] and a dedicated, experienced community manager can ensure an online community's success.

Conclusion

Most of the practices discussed in this paper are not unique to health communities. However, establishing an online health community's purpose and its members' motivations helps community managers modify these practices to tailor the engagement tactics for online health communities. As people increasingly turn to online health communities for information and support, it is vital to realize that community management is more than just moderation (see [Multimedia Appendix 2](#)). Behind each thriving online community is an enabler—a community manager who establishes the tone, proactively initiates and maintains growth through outreach and by encouraging member referrals, ensures regular activity, nurtures core members, hosts events, creates content for and about the community, fosters a sense of community, and constantly gathers data and feedback to guide and improve the community. A successful online community manager will adapt the management strategy as the community evolves through the various stages of its life cycle. “Fostering participation is one of the most difficult, yet crucial, roles for online facilitators” [50].

Acknowledgments

The author would like to thank Jacqueline (Jackie) L Bender, PhD, for her encouragement to submit the abstract for the presentation to Medicine 2.0, 2012, and for her review of this paper. Jackie is a postdoctoral research fellow at ELLICSR Health, Wellness, and Cancer Survivorship Centre and a member of the Phi Group at the Centre for eHealth Innovation, University Health Network, Toronto, Canada. The author would also like to acknowledge Richard Millington, Founder of FeverBee Limited, an online community consultancy and professional community management learning course, for contributing much of the reference material for this paper and for his ongoing mentorship.

Colleen Young was the winner of the JMIR Medicine 2.0 Award 2012 at the 5th Medicine 2.0 World Congress on Social Media, Mobile Apps, and Internet/Web 2.0 in Health, Medicine and Biomedical Research at Harvard University. This award provided a waiver of the article publication fee by JMIR Publications.

The author thanks the JMIR reviewers who provided valuable suggestions.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Online communities mentioned in this paper.

[[PDF File \(Adobe PDF File\), 214KB - jmir_v15i6e119_app1.pdf](#)]

Multimedia Appendix 2

A 12-point summary to building a successful online community.

[PDF File (Adobe PDF File), 215KB - [jmir_v15i6e119_app2.pdf](#)]

References

1. Fox S. Peer-to-peer healthcare.: Pew Internet & American Life Project; 2012 Dec. URL: <http://www.pewinternet.org/Reports/2011/Medicine-20/Part-1.aspx> [accessed 2012-12-15] [WebCite Cache ID 6Cw58c6Uh]
2. Zieband 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. [doi: [10.1111/j.1468-0009.2012.00662.x](https://doi.org/10.1111/j.1468-0009.2012.00662.x)] [Medline: [22709387](#)]
3. O'Grady L, Bender J, Urowitz S, Wiljer D, Jadad A. Promoting and participating in online health forums: A guide to facilitation and evaluation for health professionals. *Journal of Communication In Healthcare* 2010;3(3-4):246-257. [doi: [10.1179/175380710X12870623776478](https://doi.org/10.1179/175380710X12870623776478)]
4. Richardson CR, Buis LR, Janney AW, Goodrich DE, Sen A, Hess ML, et al. An online community improves adherence in an internet-mediated walking program. Part 1: results of a randomized controlled trial. *J Med Internet Res* 2010 Dec;12(4):e71 [FREE Full text] [doi: [10.2196/jmir.1338](https://doi.org/10.2196/jmir.1338)] [Medline: [21169160](#)]
5. 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](https://doi.org/10.2105/AJPH.2009.165449)] [Medline: [20466971](#)]
6. Cobb NK, Graham AL, Byron MJ, Niaura RS, Abrams DB, Workshop Participants. Online social networks and smoking cessation: a scientific research agenda. *J Med Internet Res* 2011 Dec;13(4):e119 [FREE Full text] [doi: [10.2196/jmir.1911](https://doi.org/10.2196/jmir.1911)] [Medline: [22182518](#)]
7. Oprea F, Campo S, Lowe J, Andsager J, Morcuende JA. Online information exchanges for parents of children with a rare health condition: key findings from an online support community. *J Med Internet Res* 2013 Jan;15(1):e16 [FREE Full text] [doi: [10.2196/jmir.2423](https://doi.org/10.2196/jmir.2423)] [Medline: [23470259](#)]
8. Iriberri A, Leroy G. A life-cycle perspective on online community success. *ACM Comput. Surv* 2009 Feb 2009;41(2):1-29. [doi: [10.1145/1459352.1459356](https://doi.org/10.1145/1459352.1459356)]
9. Wenger EC, McDermott R, Snyder W. *Cultivating Communities of Practice*. 1st Edition. Boston, MA: Harvard Business Review Press; 2002.
10. Millington R. FeverBee. 2012. Introducing The Map - A Proven Process For Developing Successful Online Communities URL: <http://www.feverbee.com/2012/01/introducing-the-map-a-proven-process-for-developing-successful-online-communities.html> [accessed 2012-12-17] [WebCite Cache ID 6Cz9xzES8]
11. Young C, Thompson C. SharingStrength. 2013. URL: <http://www.sharingstrength.ca/> [accessed 2013-04-02] [WebCite Cache ID 6FZv1whhP]
12. Young C. About Health Care Social Media Canada: hcsma. 2013. URL: <http://cyhealthcommunications.wordpress.com/hcsma-2/> [accessed 2013-04-02] [WebCite Cache ID 6FZvkD5qv]
13. Sinardo H. CancerConnection.: Canadian Cancer Society; 2013. URL: <http://cancerconnection.ca/> [accessed 2013-04-02] [WebCite Cache ID 6FZvwY8oT]
14. Young C. Canadian Virtual Hospice. 2013. Canadian Virtual Hospice Discussion Forums URL: http://www.virtualhospice.ca/en_US/Main+Site+Navigation/Home/Support/Support/Discussion+Forums.aspx [accessed 2013-04-02] [WebCite Cache ID 6FZwF3Z2S]
15. Preece J. Sociability and usability in online communities: Determining and measuring success. *Behaviour & Information Technology* 2001 Jan 2001;20(5):347-356. [doi: [10.1080/01449290110084683](https://doi.org/10.1080/01449290110084683)]
16. Rheingold H. *The virtual community: homesteading on the electronic frontier*. Cambridge, Mass: MIT Press; 2000.
17. Preece J, Maloney-Krichmar D. Online communities: design, theory and practice. *J Comput Mediat Commun* 2005;10(4):1. [doi: [10.1111/j.1083-6101.2005.tb00264.x](https://doi.org/10.1111/j.1083-6101.2005.tb00264.x)]
18. Preece J, Maloney-Krichmar D. Online Communities. In: Jacko J, Sears A, editors. *Handbook of Human-Computer Interaction*. Mahwah, NJ: Lawrence Erlbaum Associates Inc; 2003:596.
19. iVillage, Inc. 2013. Garden Web Forums URL: <http://forums.gardenweb.com/forums/> [accessed 2013-04-02] [WebCite Cache ID 6FZwsSBkw]
20. BritishExpats. Internet Brands. 2013. URL: <http://britishexpats.com/> [accessed 2013-04-02] [WebCite Cache ID 6FZxCGt6Q]
21. Ickes W. *Empathic accuracy*. New York: Guilford Press; 1997.
22. Bender JL, Jimenez-Marroquin MC, Ferris LE, Katz J, Jadad AR. Online communities for breast cancer survivors: a review and analysis of their characteristics and levels of use. *Support Care Cancer* 2013 May;21(5):1253-1263. [doi: [10.1007/s00520-012-1655-9](https://doi.org/10.1007/s00520-012-1655-9)] [Medline: [23179491](#)]
23. Esquivel A, Meric-Bernstam F, Bernstam EV. Accuracy and self correction of information received from an internet breast cancer list: content analysis. *BMJ* 2006 Apr 22;332(7547):939-942 [FREE Full text] [doi: [10.1136/bmj.38753.524201.7C](https://doi.org/10.1136/bmj.38753.524201.7C)] [Medline: [16513686](#)]

24. Millington R. Pillar Summit. 2012. The Online Community Development Process URL: <http://pillarsummit.com/communitydevelopment.pdf> [accessed 2012-12-17] [WebCite Cache ID 6Cz8UnQA8]
25. Blanchard AL, Markus ML. The experienced "sense" of a virtual community. SIGMIS Database 2004 Feb 2004;35(1):64-71. [doi: [10.1145/968464.968470](https://doi.org/10.1145/968464.968470)]
26. Ahrens S. Crohnology. URL: <http://crohnology.com/> [accessed 2013-04-02] [WebCite Cache ID 6FZxeq89G]
27. BabyCenter Canada. URL: <http://www.babycenter.ca/community> [accessed 2013-04-02] [WebCite Cache ID 6FZxxHBE2]
28. Moore TD, Serva MA. Understanding member motivation for contributing to different types of virtual communities: a proposed framework. In: Proceedings of the. 2007 Presented at: ACM SIGMIS CPR conference; Apr. 19-21, 2007; St. Louis, MO p. 153-158. [doi: [10.1145/1235000.1235035](https://doi.org/10.1145/1235000.1235035)]
29. McMillan DW, Chavis DM. Sense of community: A definition and theory. Journal of Community Psychology 1986;14:6-23. [doi: [10.1002/1520-6629\(198601\)14:1](https://doi.org/10.1002/1520-6629(198601)14:1)]
30. Blanchard AL. Developing a sense of virtual community measure. Cyberpsychol Behav 2007 Dec;10(6):827-830. [doi: [10.1089/cpb.2007.9946](https://doi.org/10.1089/cpb.2007.9946)] [Medline: [18085972](https://pubmed.ncbi.nlm.nih.gov/18085972/)]
31. Association of Cancer Online Resources. URL: <http://acor.org/> [accessed 2013-04-02] [WebCite Cache ID 6FZyB1GFk]
32. Millington R. Pillar Summit. 2012. The Proven Path URL: <http://pillarsummit.com/the-proven-path/theprovenpath.pdf> [accessed 2012-12-17] [WebCite Cache ID 6CzABWLSX]
33. Hernandez M. Tudiabetes Care Team. 2013. URL: <http://www.tudiabetes.org/group/welcomecommittee> [accessed 2013-04-02] [WebCite Cache ID 6FZyT5Stq]
34. Resnick PJ, Janney AW, Buis LR, Richardson CR. Adding an online community to an internet-mediated walking program. Part 2: strategies for encouraging community participation. J Med Internet Res 2010;12(4):e72 [FREE Full text] [doi: [10.2196/jmir.1339](https://doi.org/10.2196/jmir.1339)] [Medline: [21169161](https://pubmed.ncbi.nlm.nih.gov/21169161/)]
35. Bender JL. The web of care: A multi-method study examining the role of online communities as a source of peer-to-peer supportive care for breast cancer survivors - Dissertation. Toronto, Canada: University of Toronto; 2011.
36. Millington R. FeverBee. 2012. Bonding A Community: The Final Step URL: <http://www.feverbee.com/2009/10/bonding-a-community-the-final-step.html> [accessed 2012-12-17] [WebCite Cache ID 6CzALQwg6]
37. 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 Jun;14(3):e66 [FREE Full text] [doi: [10.2196/jmir.1854](https://doi.org/10.2196/jmir.1854)] [Medline: [22732103](https://pubmed.ncbi.nlm.nih.gov/22732103/)]
38. Heywood B, Heywood J. PatientsLikeMe. 2013. URL: <http://www.patientslikeme.com/> [accessed 2013-04-02] [WebCite Cache ID 6FZyp9Sji]
39. Millington R. FeverBee. 2013. Using Data To Identify Potential Issues URL: <http://www.feverbee.com/2012/02/using-data-to-identify-potential-issues.html> [accessed 2013-04-02] [WebCite Cache ID 6FZzYsQ7A]
40. Walther JB, Pingree S, Hawkins RP, Buller DB. Attributes of interactive online health information systems. J Med Internet Res 2005 Jul 1;7(3):e33 [FREE Full text] [doi: [10.2196/jmir.7.3.e33](https://doi.org/10.2196/jmir.7.3.e33)] [Medline: [15998624](https://pubmed.ncbi.nlm.nih.gov/15998624/)]
41. Millington R. FeverBee. 2013. Measuring An Online Community: Master your data to gain an unfair advantage URL: <http://www.feverbee.com/2011/06/measuring.html> [accessed 2013-04-02] [WebCite Cache ID 6Fa03F6c6]
42. Peebles R, Wilson JL, Litt IF, Hardy KK, Lock JD, Mann JR, et al. Disordered eating in a digital age: eating behaviors, health, and quality of life in users of websites with pro-eating disorder content. J Med Internet Res 2012 Oct;14(5):e148 [FREE Full text] [doi: [10.2196/jmir.2023](https://doi.org/10.2196/jmir.2023)] [Medline: [23099628](https://pubmed.ncbi.nlm.nih.gov/23099628/)]
43. Sinardo H, Young C. CancerConnection Community Guidelines.: Canadian Cancer Society; 2013. URL: <http://cancerconnection.ca/discussions/1---getting-started-on-the-site/2010/12/11/community-expectations> [accessed 2013-04-02] [WebCite Cache ID 6Fa0RrHdl]
44. Canadian Cancer Society. CancerConnection/Parlonscancer Online Community Terms of Use. URL: https://secure2.convio.net/ccco/site/SPageNavigator/CANCON_CAN_Terms_of_Use [accessed 2013-04-02] [WebCite Cache ID 6Fa0huFbh]
45. Moderator team. Macmillan Cancer Support. Macmillan Admin Blog - Please Read: Anna Johns URL: <http://community.macmillan.org.uk/blogs/b/admin/archive/2011/11/25/please-read-anna-johns.aspx> [accessed 2013-04-02] [WebCite Cache ID 6Fa3IWPcq]
46. hcsma community. LinkedIn. Thread about Patient Experience on #hcsma LinkedIn Group URL: <http://www.linkedin.com/groups/Connect-other-healthcare-leaders-NATIONAL-3487112.S.224279710> [accessed 2013-04-02] [WebCite Cache ID 6Fa42IBQf]
47. Allan E. Canadian Virtual Hospice. Community Post: Help needed to reach participants for Photo-documentary book "Facing Death" URL: http://www.virtualhospice.ca/en_US/Main+Site+Navigation/Home/Support/Support/Discussion+Forums/Just+Want+to+Talk/2012_05_15_14_10_21_Help+needed+to+reach+participants+for+Photo_documentary+book+quot;Facing+Death+quot;.aspx [accessed 2013-04-02] [WebCite Cache ID 6Fa3bO5OY]
48. van Uden-Kraan CF, Drossaert CH, Taal E, Seydel ER, van de Laar MA. Patient-initiated online support groups: motives for initiation, extent of success and success factors. J Telemed Telecare 2010;16(1):30-34. [doi: [10.1258/jtt.2009.001009](https://doi.org/10.1258/jtt.2009.001009)] [Medline: [20086265](https://pubmed.ncbi.nlm.nih.gov/20086265/)]

49. Millington R. FeverBee. How to Develop Your Online Community Management Strategy URL: <http://www.feverbee.com/2011/06/difficult-community-management.html> [accessed 2012-12-17] [WebCite Cache ID 6CzMdO5HC]
50. Herie M. Online Teaching and Learning. eCOMMUNITY: Int J Mental Health Addict 2004;2(1):22-28 [FREE Full text]

Abbreviations

ACOR: Association of Cancer Online Resources

hcsmdca: Health Care Social Media Canada

Edited by G Eysenbach; submitted 19.12.12; peer-reviewed by L Fernandez-Luque; comments to author 30.01.13; revised version received 02.04.13; accepted 08.05.13; published 11.06.13.

Please cite as:

Young C

Community Management That Works: How to Build and Sustain a Thriving Online Health Community

J Med Internet Res 2013;15(6):e119

URL: <http://www.jmir.org/2013/6/e119/>

doi: [10.2196/jmir.2501](https://doi.org/10.2196/jmir.2501)

PMID: [23759312](https://pubmed.ncbi.nlm.nih.gov/23759312/)

©Colleen Young. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 11.06.2013. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

Virtual Reality for Enhancing the Cognitive Behavioral Treatment of Obesity With Binge Eating Disorder: Randomized Controlled Study With One-Year Follow-up

Gian Luca Cesa^{1,2}, MSc, PsyD; Gian Mauro Manzoni^{3,4}, PsyD, PhD (Psych); Monica Bacchetta¹, PsyD, PhD (Psych); Gianluca Castelnuovo^{3,4}, PsyD, PhD (Psych); Sara Conti¹, MSc, PsyD; Andrea Gaggioli^{1,4}, MSc, PhD (Psych); Fabrizia Mantovani⁵, MSc, PhD (Psych); Enrico Molinari^{3,4}, MSc, PsyD; Georgina Cárdenas-López⁶, PhD; Giuseppe Riva^{1,4}, MSc, PhD (Psych)

¹Applied Technology for Neuro Psychology Lab, Istituto Auxologico Italiano, Verbania, Italy

²Centro Obesità e Nutrizione Clinica (CONC), Ospedale Privato Accreditato Villa Igea, Forlì (FC), Italy

³Psychology Research Laboratory, Ospedale San Giuseppe, Istituto Auxologico Italiano, Verbania, Italy

⁴Department of Psychology, Catholic University of Sacred Heart, Milan, Italy

⁵CESCOM-Centre for Research in Communication Sciences, University of Milan-Bicocca, Milan, Italy

⁶Laboratorio de Enseñanza Virtual y Ciberpsicología, Facultad de Psicología, Universidad Nacional Autónoma de México, México, Mexico

Corresponding Author:

Gian Mauro Manzoni, PsyD, PhD (Psych)

Psychology Research Laboratory

Ospedale San Giuseppe

Istituto Auxologico Italiano

Via Cadorna, 90

Piancavallo (VB)

Verbania, 28824

Italy

Phone: 39 323 514278

Fax: 39 323 514278

Email: gm.manzoni@auxologico.it

Abstract

Background: Recent research identifies unhealthful weight-control behaviors (fasting, vomiting, or laxative abuse) induced by a negative experience of the body, as the common antecedents of both obesity and eating disorders. In particular, according to the allocentric lock hypothesis, individuals with obesity may be locked to an allocentric (observer view) negative memory of the body that is no longer updated by contrasting egocentric representations driven by perception. In other words, these patients may be locked to an allocentric negative representation of their body that their sensory inputs are no longer able to update even after a demanding diet and a significant weight loss.

Objective: To test the brief and long-term clinical efficacy of an enhanced cognitive-behavioral therapy including a virtual reality protocol aimed at unlocking the negative memory of the body (ECT) in morbidly obese patients with binge eating disorders (BED) compared with standard cognitive behavior therapy (CBT) and an inpatient multimodal treatment (IP) on weight loss, weight loss maintenance, BED remission, and body satisfaction improvement, including psychonutritional groups, a low-calorie diet (1200 kcal/day), and physical training.

Methods: 90 obese (BMI>40) female patients with BED upon referral to an obesity rehabilitation center were randomly assigned to conditions (31 to ECT, 30 to CBT, and 29 to IP). Before treatment completion, 24 patients discharged themselves from hospital (4 in ECT, 10 in CBT, and 10 in IP). The remaining 66 inpatients received either 15 sessions of ECT, 15 sessions of CBT, or no additional treatment over a 5-week usual care inpatient regimen (IP). ECT and CBT treatments were administered by 3 licensed psychotherapists, and patients were blinded to conditions. At start, upon completion of the inpatient treatment, and at 1-year follow-up, patients' weight, number of binge eating episodes during the previous month, and body satisfaction were assessed by self-report questionnaires and compared across conditions. 22 patients who received all sessions did not provide follow-up data (9 in ECT, 6 in CBT, and 7 in IP).

Results: Only ECT was effective at improving weight loss at 1-year follow-up. Conversely, control participants regained on average most of the weight they had lost during the inpatient program. Binge eating episodes decreased to zero during the inpatient program but were reported again in all the three groups at 1-year follow-up. However, a substantial regain was observed only in the group who received the inpatient program alone, while both ECT and CBT were successful in maintaining a low rate of monthly binge eating episodes.

Conclusions: Despite study limitations, findings support the hypothesis that the integration of a VR-based treatment, aimed at both unlocking the negative memory of the body and at modifying its behavioral and emotional correlates, may improve the long-term outcome of a treatment for obese BED patients. As expected, the VR-based treatment, in comparison with the standard CBT approach, was able to better prevent weight regain but not to better manage binge eating episodes.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN): 59019572; <http://www.controlled-trials.com/ISRCTN59019572> (Archived by WebCite at <http://www.webcitation.org/6GxHxAR2G>)

(*J Med Internet Res* 2013;15(6):e113) doi:[10.2196/jmir.2441](https://doi.org/10.2196/jmir.2441)

KEYWORDS

virtual reality; obesity; binge eating disorders; allocentric lock hypothesis

Introduction

Binge Eating Disorder (BED) is proposed in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM IV-TR) as a new diagnostic category, requiring further research, within the spectrum of the category Eating Disorders Not Otherwise Specified (EDNOS). BED is characterized by eating a much larger amount of food than most people would consider normal in a discrete amount of time (2 hours). This is associated with a loss of control about what and how much is eaten, but without the compensatory behaviors (vomiting, use of laxatives) typical of bulimia nervosa [1]. On average, individuals suffering from BED have a high prevalence of psychiatric and medical comorbidities, as well as obesity [2]. In fact, BED may occur in up to 30% of extremely obese subjects seeking treatment at weight loss programs [3], and obesity may be observed in approximately 65-70% of people with BED [4]. Cognitive behavioral therapy (CBT) can be considered one of the better clinical approaches available to treat BED [5]. CBT has been shown to reduce binge days and episodes and to improve the psychological features of BED, such as measures of restraint, hunger, and disinhibition surrounding eating. However, CBT alone has not shown strong results in decreased weight and sustained weight reduction in the mid-term (3- and 6-month follow-ups) [2,5].

The rising prevalence of weight-related disorders and the lack of a clear solution are pushing eating disorder and obesity researchers to start collaborating across fields to address them. In particular, their effort is focused on the identification of risk factors that are shared between these weight-related disorders [6]. Apparently, unhealthful weight-control behaviors, such as fasting (going without eating for 24 hours for weight control), vomiting, or laxative abuse, are the common antecedents of both obesity and eating disorders [6-11]. For example, Neumark-Sztainer and colleagues [7] discussed the results of the Project EAT II (Eating Among Teens), a longitudinal study involving 2516 ethnically and socioeconomically diverse adolescents. They report that, 5 years later, the use of unhealthful weight-control behaviors increased the risk for binge eating with loss of control by a factor of six, for being overweight by a factor of three, and for extreme weight-control behaviors, such

as the use of diet pills and self-induced vomiting by a factor between two and five. A similar result was found by Stice and colleagues [11]: in a different longitudinal study fasting was the best predictor for the future onset, 5 years later, of binge eating and bulimia nervosa. These data have an important clinical implication: the evidence that youths practicing unhealthful weight-control behaviors are at higher risk for obesity implies that prevention and treatment interventions should also focus on the causes of these behaviors. A study by Kostanski and Gullone [12] with a sample of 431 Australian pre-adolescent children (7-10 years) offers a possible interpretation: pre-adolescents as young as 7 years of age are unsatisfied with their body appearance and deliberately engage in restrictive eating behaviors. Moreover, a recent study [13] showed that in adolescents frequent self-weighing is associated with lower body satisfaction and higher rates of unhealthy and extreme weight control behaviors.

In particular, according to the allocentric lock hypothesis (ALH) [14,15], individuals with obesity and eating disorders may be locked to an allocentric (observer view) negative image schema [16] of their body that is no longer updated by contrasting egocentric representations driven by perception. In other words, these patients may be locked to a negative image of their body that perception is not able to update even after a demanding diet and a significant weight loss. In sum, they cannot win. Whatever they do to modify their real body, they will always be locked in a negative body perception that they hate. This situation usually has two effects: subjects either start more radical dietary restraint or, at the opposite end, decide to stop any form of food control and start “disinhibited” eating behaviors. The passage from a locked allocentric representation to eating or weight disorders may be explained by social influence: media and culture promote diet as the best way to improve body image satisfaction. However, the impossibility of improving body image, even after a demanding diet, locks the patient into an unsatisfying body experience.

Although the ALH is starting to be backed by research data with anorectic patients [17,18], we decided to include in a CBT approach a virtual reality (VR) protocol aimed both at unlocking the negative memory of the body [15] and at modifying its behavioral and emotional correlates within an inpatient,

medically managed, intensive cognitive-behavioral BED treatment.

The purpose of this study is to evaluate the brief and long-term efficacy of the proposed approach (VR-enhanced CBT for obese inpatients with BED) in a randomized controlled trial. Outcome measures are weight, number of binge eating episodes during the previous month, and body satisfaction. We hypothesize that the VR-enhanced CBT (ECT) is more effective than standard CBT and a control condition in (1) maintaining and further improving weight loss, and (2) maintaining binge eating remission at 1-year follow-up after discharge from a multimodal medically managed inpatient program (IP). Furthermore, we hypothesize that ECT is more effective than standard CBT and a control condition in improving and maintaining body satisfaction.

The study was approved by the Ethical Committee of the Istituto Auxologico.

Methods

Participants and Procedures

This study was a randomized controlled trial (ISRCTN 59019572). In total, 124 consecutive patients seeking treatment at the Eating Disorder Unit of the Istituto Auxologico Italiano, Verbania, Italy, were seen for screening interviews for admission to the study. Criteria for participation included the following: (1) women aged 18-50 years, (2) who met DSM-IV-TR criteria

for BED for at least 6 months prior to the beginning of the study, (3) no other concurrent severe psychiatric disturbance (psychosis, depression with suicidal risk, alcohol or drug abuse), (4) no concurrent involvement in other treatment for BED, including pharmacotherapy, (5) no concurrent medical condition not related to the disorder, and (6) written and informed consent to participate.

Of these, 34 either did not fulfill inclusion criteria or were excluded for other reasons (eg, time constraints). All patients meeting the inclusion criteria were then consecutively and randomly assigned to one of the three experimental conditions described below. The randomization scheme was generated by using a randomization website [19]. After allocation, 24 patients declined participation in the study (Figure 1). In total, 66 female patients (mean age 31.79 ± 7.9 years, mean weight 106.6 ± 17.7 kg, mean height 162 ± 7 cm, mean BMI 40.5 ± 5.2) entered the treatment phase (Figure 1; a detailed description of the inclusion criteria is included in the CONSORT-ehealth form available in Multimedia Appendix 1). The sample characteristics are shown in Table 1. Baseline comparisons among the three groups showed a difference only in marital status. Percentages of married patients were significantly higher in the IP group and, to a lesser extent, in the ECT group than in the CBT group.

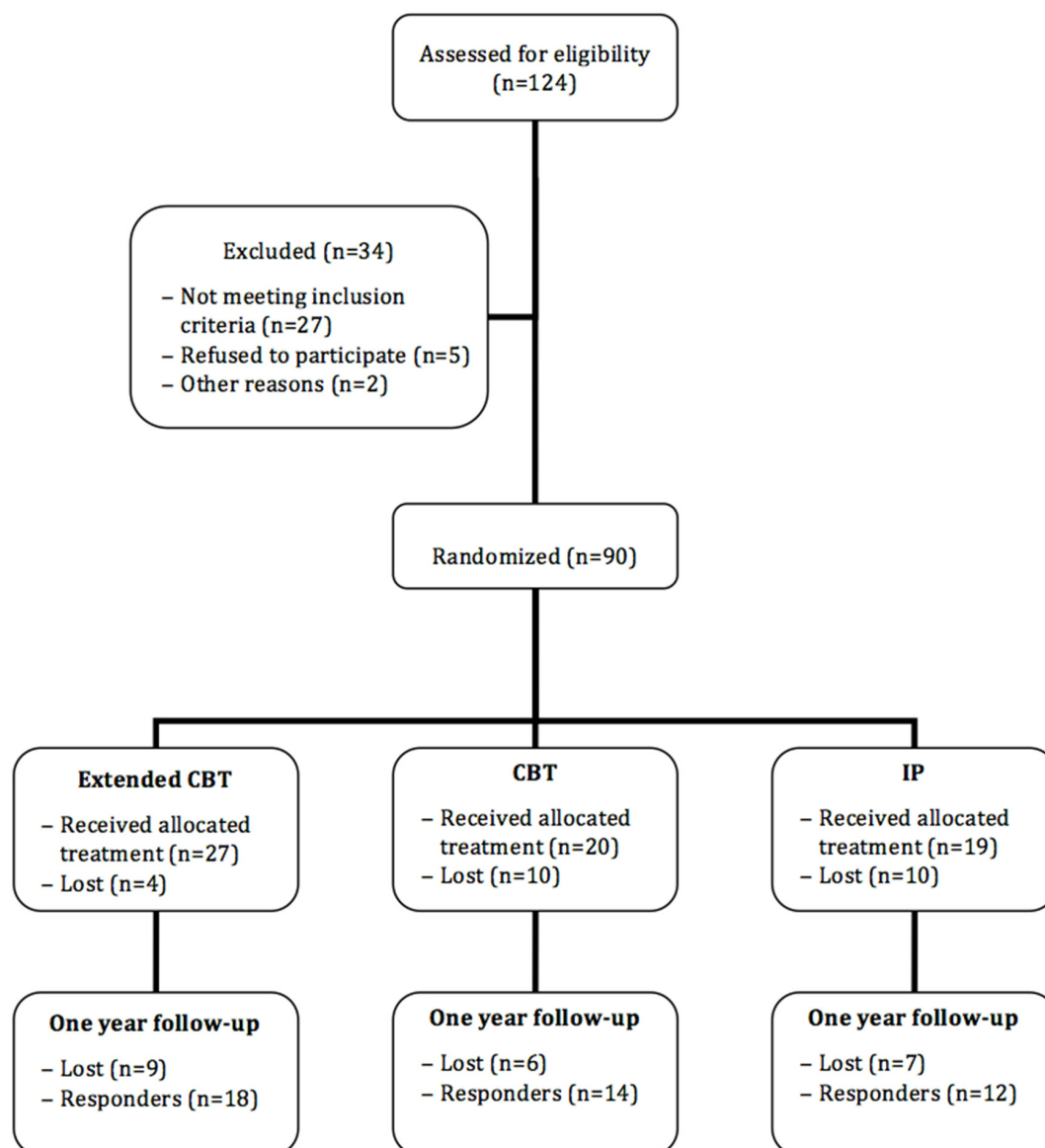
The VR-enhanced CBT and traditional CBT (see below for treatment details) were administered by 2 licensed clinical psychologists and 1 licensed psychotherapist under the supervision of a senior licensed psychotherapist. The 3 therapists were randomized to the two treatment conditions.

Table 1. Baseline characteristics.

	ECT (n=27)	CBT (n=20)	IP (n=19)	P
Age, mean (SD), years	32.9 (8.8)	29.9 (7.95)	32.2 (6.36)	.324 ^a
Weight, mean (SD), kg	103 (18.2)	106.6 (8.9)	111.6 (22.9)	.223 ^a
BMI, mean (SD), kg/m ²	39.2 (5.3)	41.1 (3.3)	41.8 (6.3)	.189 ^a
Education, n (%)				
University	4 (14.8)	1 (5)	2 (10.5)	.481 ^b
High school	14 (51.9)	12 (60)	14 (73.7)	
Lower education	9 (33.3)	7 (35)	3 (15.8)	
Marital status, n (%)				
Married	(44.4)	(25)	(68.4)	.026 ^b

^aKruskall-Wallis test with Monte Carlo *P* estimation.

^bChi-square test with Monte Carlo *P* estimation.

Figure 1. Clinical trial flowchart.

The Integrated Multimodal Medically Managed Inpatient Program

The integrated multimodal medically managed inpatient program (IP) was the common treatment condition for all the participants. It consisted of hospital-based living for a duration of 6 weeks. Inpatients received medical, nutritional, physical, and psychological care. In particular, they maintained a low-calorie diet (tailored to patients' needs), entered weekly nutritional groups held by dietitians, received psychological support both in individual and group settings, and undertook physical training. Participants who were allocated to this condition were considered "controls" and did not enter the following treatments.

Cognitive Behavior Therapy

During the inpatient program, participants allocated to this condition received 15 additional cognitive behavior therapy (CBT) sessions over 5 weeks. Therapists followed a detailed manual that outlined the contents of each session. This manual was based on the CBT approach described by Fairburn and

colleagues [20,21] and by Ricca and colleagues [22]. It was developed during a year of intensive pilot work and adapted to the inpatient setting. In particular, after the first inpatient week, participants entered 5 weekly group sessions and 10 biweekly individual sessions. The first 8 individual sessions were structured according to Stage 1 of the CBT manual for binge eating. They focused on an overview of the goals of the treatment program, the use of self-monitoring records to identify high-risk situations that might trigger binge eating, support in normalizing eating patterns, and the identification of behavioral strategies for coping with high-risk situations for binge eating. The final 2 individual sessions focused on the maintenance of improvement and on relapse prevention.

The group sessions were structured according to Stage 2 of the CBT manual for binge eating. They focused on problem-solving strategies and cognitive interventions targeting concerns about body weight and shape and problematic eating. After hospital discharge, continuity of care and support through telecommunication devices (email, chat, and telephone as

preferred) were offered to each patient. Contacts were not scheduled and were dependent on patients' needs.

VR-Enhanced Cognitive Behavior Therapy

Like the CBT condition, participants allocated to VR-enhanced cognitive behavior therapy (ECT) received 15 additional sessions over 5 weeks. After the inpatients' first week, participants entered 5 weekly group sessions similar to the CBT ones (focused on concerns about body weight and shape and problematic eating) and 10 biweekly VR sessions. ECT treatment was based on a detailed protocol describing the contents of each of the 15 sessions [23,24]. For the virtual reality sessions, NeuroVR open-source software was used [25-27]. NeuroVR includes 14 virtual environments used by the therapist during a 60-minute session with the patient (see Figure 2). The

environments present critical situations related to the maintaining/relapse mechanisms (Home, Supermarket, Pub, Restaurant, Swimming Pool, Beach, Gymnasium) and two body image comparison areas. Through the VR experience, patients practiced both eating/emotional/relational management and general decision-making and problem-solving skills. By directly practicing these skills within the VR environment, patients were helped in developing specific strategies for avoiding and/or coping with triggering situations.

The first session was used to assess any stimuli that could elicit abnormal eating behavior. Specifically, the attention was focused on a patient's concerns about food, eating, shape, and weight. The next 14 sessions were used to assess and modify the following.

Figure 2. A screenshot of the NeuroVR 2 open-source software.



Expectations and Emotions Related to Food and Weight (Functional Analysis)

The therapist helped patients to recognize why they eat and what they need to either avoid or cope with the specific emotional/behavioral triggers. This was achieved by integrating different cognitive-behavioral methods: Countering, Alternative Interpretation, Label Shifting, and Deactivating the Illness Belief.

Strategies Used to Cope With Difficult Interpersonal and Potential Maintenance Situations

The patient practiced old behaviors and tested new ones. This was achieved both by using the Temptation Exposure with Response Prevention [28,29] (skills training) and by working on these three empowering dimensions [30]: perceived control,

perceived competence, and goal internalization (fostering the motivation).

Body Experience of the Subject

To do this, the virtual environment integrates the therapeutic methods used by Butter & Cash [31] and Wooley & Wooley [32] with the body image rescripting protocol based on ALH (see Table 2) [15,33].

In particular, we used the virtual environment in the same way as guided imagery [34] is used in the cognitive and visual/motorial approach.

After hospital discharge, continuity of care and support through telecommunication devices (email, chat, and telephone as preferred) were offered also to patients allocated to this condition. As in the previously described condition, contacts were not scheduled and were dependent on patients' needs.

Table 2. The VR body image rescripting protocol (adapted from Riva, 2011).

Phase 1: Interview	During a clinical interview, the patient is asked to relive the contents of the negative body image and the situation/s in which it was created and/or reinforced (eg, “being teased by my boyfriend at home”) in as much detail as possible. The meaning of the experience for the patient was also elicited.
Phase 2: Development of the VR scene	The clinician reproduces the setting of the identified situation (eg, “the corridor of the classroom where my boyfriend teased me”) using one of the different scenes available in the free NeuroVR software.
Phase 3: Egocentric experience of the VR scene	<p>The patient is asked to re-experience the event in VR from a first person perspective (the patient does not see his/her body in the scene) expressing and discussing his/her feelings. The patient is then asked what was needed to happen to change the feelings in a positive direction. The main cognitive techniques used in this phase, if needed, are:</p> <p>Countering: Once a list of distorted perceptions and cognitions is developed, the process of countering these thoughts and beliefs begins.</p> <p>Label Shifting: The patient first tries to identify the kinds of negative words she uses to interpret situations in her life, such as bad, terrible, obese, inferior, and hateful. The situations in which these labels are used are then listed. The patient and therapist replace each emotional label with two or more descriptive words.</p>
Phase 4: Allocentric experience of the VR scene	<p>The patient is asked to re-experience the event in VR from a third person perspective (the patient sees his/her body in the scene) intervening both to calm and reassure his/her virtual avatar and to counter any negative evaluation. The therapist follows the Socratic approach, for example, “What would need to happen for you to feel better? How does it look through the eyes of a third person? Is there anything you as a third person would like to do? How do the other people respond?” The main cognitive techniques used in this phase, if needed, are:</p> <p>Alternative Interpretation: The patient learns to stop and consider other interpretations of a situation before proceeding to the decision-making stage.</p> <p>Deactivating the Illness Belief: The therapist first helps the client list his/her beliefs concerning weight and eating.</p>

Assessment

Assessments were obtained 1 week after the start of the inpatient program, at the last week, and at 1-year follow-up (by postal mail). Height was measured with a stadiometer, and weight was assessed with the participant in lightweight clothing with shoes removed, on a balance beam scale. A single question extracted from the EDI-Symptom Checklist [35] was administered at each time-point to assess the number of binge eating episodes (with binge eating defined as the consumption of unusually large amounts of food with a subjective sense of loss of control during the last month). Data at follow-up were self-reported. The following self-report questionnaires were also administered 1 week after the start of the inpatient program, at the last week, and at 1-year follow-up (by postal mail):

- *The Italian version [36] of the Body Satisfaction Scale (BSS) [37]:* The scale consists of a list of 16 body parts, half involving the head (above the neck) and the other half involving the body (below the head). The subjects rate their satisfaction with each of these body parts on a 7-point scale: the higher the rating, the more dissatisfied the individual.
- *The Italian version [38] of the Body Image Avoidance Questionnaire (BIAQ) [39]:* The BIAQ is a 19-item self-report questionnaire on avoidance of situations that provoke concern about physical appearance, such as avoidance of tight-fitting clothes, social outings, and physical intimacy. In particular, the questionnaire measures the avoidance behaviors and grooming habits associated with negative body image.
- *Contour Drawing Rating Scale (CDRS) [40]:* This is a set of 9 male and female figures with precisely graduated increments between adjacent size. In this test, subjects rate the figures based on the following instructional protocol: (a) current size and (b) ideal size. The difference between

the ratings is called the “self-ideal discrepancy score” and is considered to represent the individual’s dissatisfaction.

Statistical Analysis

A power calculation was done to evaluate the possibility of detecting statistically significant differences both between the groups (independent measures) and within groups (repeated measures). Given the low/medium statistical power due to the relatively small number of participants, the non-normality of several distributions and the unbalanced groups, we decided to use the exact methods with Monte Carlo approximation: a series of nonparametric statistical algorithms that enable researchers to make reliable inferences when data are sparse, heavily tied or unbalanced, not normally distributed, or fail to meet any of the underlying assumptions necessary for reliable results using the standard asymptotic method [41]. The exact methods with Monte Carlo approximation used for comparisons are the Kruskal-Wallis test with post hoc analysis [42] for independent measures, the Wilcoxon rank-sum test for repeated measures, and Chi-square for categorical variables, with $\alpha=0.5$, two-tailed. Weight data were analyzed with an intention-to-treat (ITT) analysis with nonresponders at follow-up assumed to have regained 0.3 kg per month—an assumption already used in previous studies [43,44]. Also, missing data in number of binge eating episodes during the previous month were replaced with baseline values carried forward (BCF), assuming that nonresponders had worsened. Differently, missing data at follow-up in the BSS, BIAQ, and CDRS questionnaires were not imputed because we had no assumption about the missing process. Odds ratios with confidence intervals were also calculated with respect to the percentage maintaining or improving the weight loss at follow-up for ECT and CBT treatment conditions in comparison with the inpatient program alone. Data were analyzed using SPSS 16.0.

Results

Inpatient Treatment Analysis

Outcome data were available for all the participants at the end of treatment. Weight significantly decreased in all the three conditions (ECT: -6.17 kg, CI -7 to -5.3, $P<.001$; CBT: -7.1 kg, CI -7.9 to -6.2, $P<.001$; IP: -6.6 kg, CI -8.1 to -5.2, $P<.001$) without any significant differences between them (Table 3). This was an expected result because all the participants underwent the same 6-week medically managed inpatient treatment including a low-calorie diet and physical training (30 minutes of walking two times a week as a minimum). For the same reason, the number of binge eating episodes decreased to zero in all the three groups (Figure 3).

Body satisfaction (BSS and CDRS) significantly improved in all the groups too, with no difference across them, while body image concerns (BIAQ-Total) significantly improved only in the ECT condition. This result probably means that experiencing a medically managed inpatient treatment helped in accepting one's own body and in decreasing dissatisfaction towards it, but only the treatment enhanced by VR exposure was effective in improving body satisfaction in relation to typical avoidance situations (Table 3).

Follow-up Analysis

One year follow-up data were available for 66.6% ($n=44$) of the participants who initially entered and completed the treatment phase. Drop-out rates were similar for each group. Patients who did not respond to the follow-up call were not interviewed about their reasons for not attending the final assessments. Differences between those who did and did not respond were tested for all the variables at baseline and at the end of treatment. Statistically significant differences were found only in median numbers of baseline binge eating episodes (12 responders, 16 non-responders). With respect to the two treatment conditions, responders were only those patients who had at least one telepsychological session through telecommunication devices with the clinical psychologist or psychotherapist to whom they had been initially assigned for treatment.

ITT analysis of weight and BMI outcomes after 1-year follow-up show a different picture in comparison with the previous results (Table 3). First, statistically significant weight and BMI increases were observed in the IP group (from 105 kg to 109.3 kg; $P<.001$), while no statistically significant weight and BMI changes were found between the end of the inpatient treatment and the 1-year follow-up in CBT and ECT groups. In addition, statistically significant differences in weight and BMI median scores at follow-up were found across the three groups in favor of ECT (Table 4). In fact, only ECT was effective in further improving weight loss at 1-year follow-up.

Table 3. Means, standard deviations, and medians for BSS, CDRS, and BIAQ-Total at entry to the study and upon completion of the inpatient program.

		ECT (n=27 ^a)			CBT ^a		IP ^a			P^b	
		Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	
BSS											
	At entry	54.85	12.8	55	60.35	8.7	62	57	12.8	57	.281
	Upon completion	45	13.9	43	52	15.5	51	47.84	13	46	.353
CDRS											
	At entry	1.85	0.35	1.8	2.3	1.65	1.8	1.8	0.44	1.8	.886
	Upon completion	1.58	0.36	1.5	2.02	1.69	1.6	1.6	0.35	1.5	.711
BIAQ-Total											
	At entry	34.4	8.5	36	33.85	5.8	33	35.53	7.16	36	.681
	Upon completion	27.2	7.23	28	31.95	6.9	32	33.1	10.26	32	.031

^aNonresponders are assumed to have regained 3.6 kg, 0.3 per month.

^bKruskal-Wallis test with Monte Carlo P estimation across groups at each time point.

Figure 3. Monthly mean number of binge eating episodes at baseline, at the end of the inpatient treatment and at 1-year follow-up (dropouts at follow-up are assumed to have regained the baseline score).

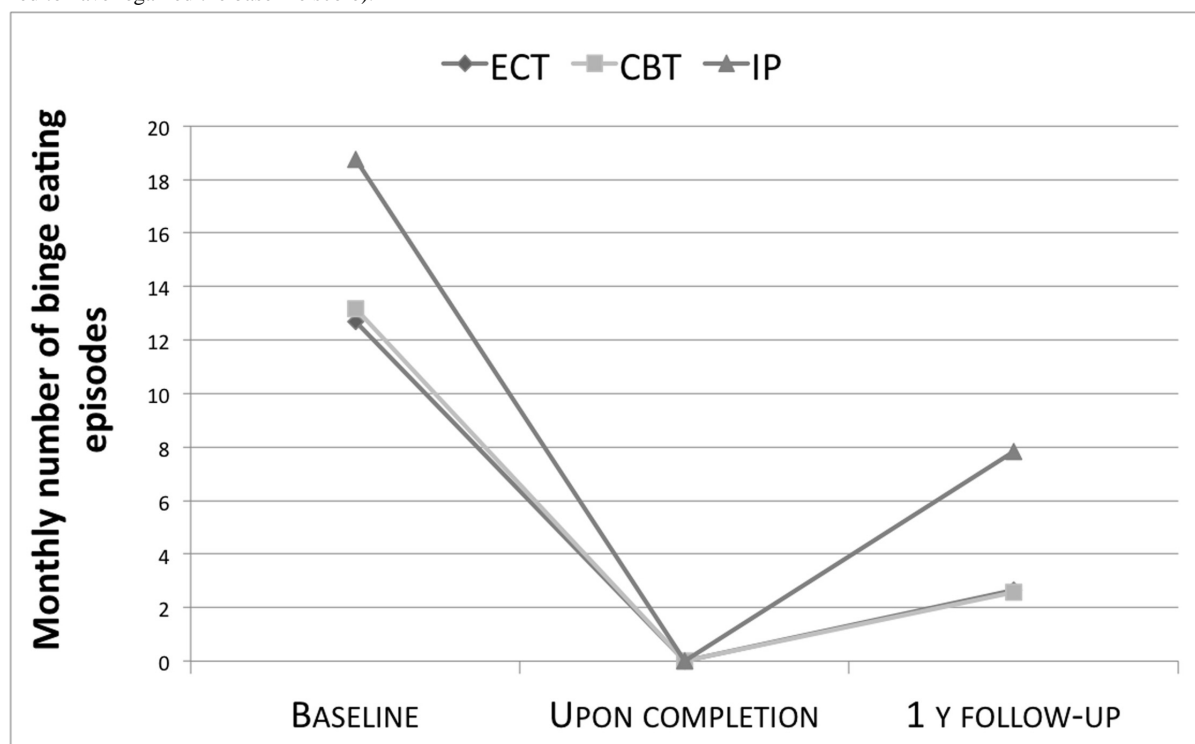


Table 4. Means, standard deviations, and medians for weight and BMI at entry of the study, upon completion of the inpatient program, and at 1-year follow-up by group.

		ECT (n=27 ^a)			CBT ^a		IP ^a			<i>P</i> ^b	
		Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	
Weight (kg)											
	At entry	103	18.2	97.6	106.6	8.9	105.8	111.7	22.9	109	.223
	Upon completion	96.9	16.7	93.6	99.5	7.9	100	105	21.8	102	.251
	1-year follow-up	96	16.3	92	101	9.4	103.7	109.3	22.6	112	.032
BMI											
	At entry	39.2	5.3	38.1	41.1	3.3	40.8	41.8	6.3	42	.189
	Upon completion	36.9	5	36.5	38.3	3	38	39.3	5.9	40.3	.228
	1-year follow-up	36.6	5	36.2	39	3.6	39.1	40.9	6	41.5	.015

^aNonresponders are assumed to have regained 3.6 kg, 0.3 per month.

^bKruskal-Wallis test with Monte Carlo *P* estimation across groups at each time point.

With respect to percentages of weight and BMI reductions at 1-year follow-up from baseline, almost significant differences emerged between the three groups ($P=.052$ for both) in favor of ECT and CBT (Figure 4). Post hoc comparisons showed a significant difference between ECT and IP ($P=.027$). Furthermore, ECT was significantly better after 1-year follow-up in improving or maintaining weight loss after treatment than IP alone. The percentages of participants who succeeded in weight maintenance or in further loss was 44.4% in ECT versus 10.5% in IP (OR 6.8, 95% CI 1.3-35.4, $P=.014$). Also CBT was significantly better after 1-year follow-up in improving or maintaining weight loss than IP alone, with 40% of participants being successful (OR 5.7, 95% CI 1.09-31.5, $P=.035$). With respect to the 5% weight loss criterion (from baseline), ITT

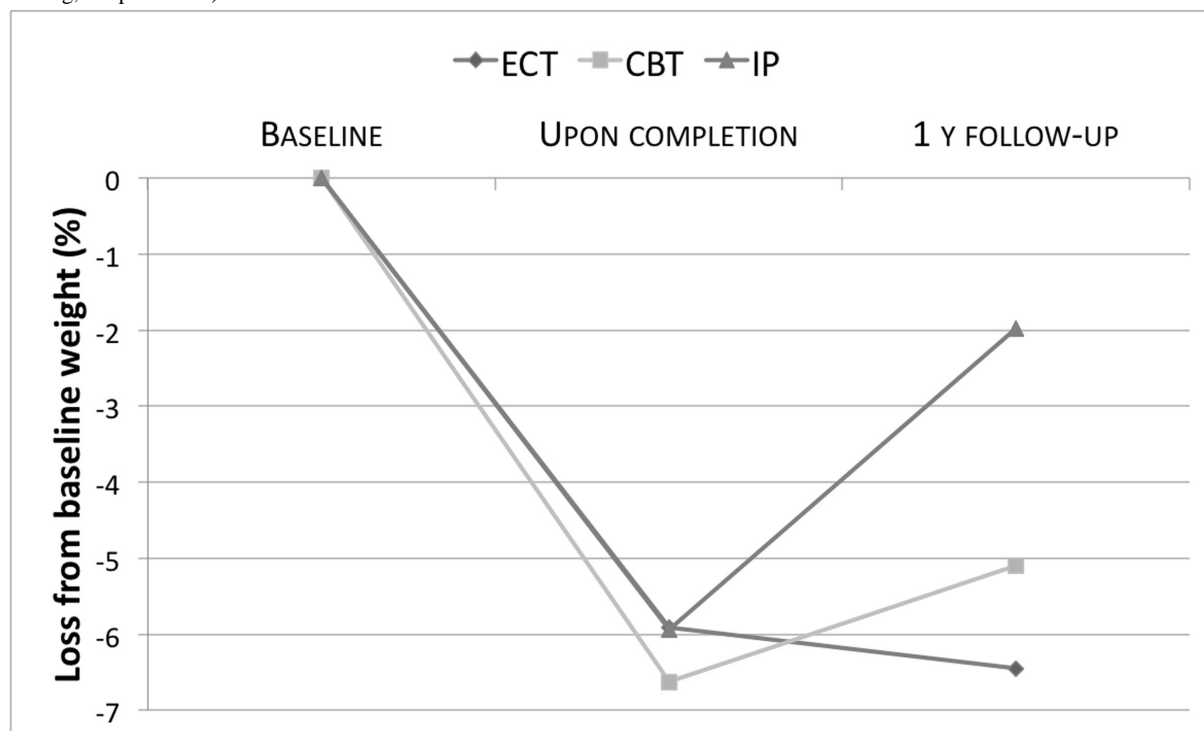
analyses did not detect any statistically significant difference across groups, even if percentages show a trend in favor of ECT (ECT 55.6%, CBT 50%, and IP 31.6%).

Monthly number of binge eating episodes significantly increased from zero in all three groups at 1-year follow-up. ITT analysis with BCF found no statistically significant difference between the groups in follow-up median scores (Figure 3). This was largely due to the relatively great number of nonresponders we handled, assigning them the baseline scores. We decided to impute missing data in order to preserve statistical power, and we used BCF because we assumed that nonresponders had worsened.

Given that drop-out rates were not different among the three groups, we also ran a sensitivity analysis by excluding patients who did not answer the follow-up call. Results were not dissimilar and depicted the same picture we observed with imputed data.

Follow-up analyses were not performed on BSS, BIAQ, and CDRS scores because missing data were not imputed and a responders-only analysis was unfeasible due to the critical lack of information. Statistical power would have been insufficient to detect even large differences across groups and the risk of type II error would have been quite inflated.

Figure 4. Median percent weight reduction at the end of the inpatient treatment and at 1-year follow-up (dropouts at follow-up are assumed to have regained 3.6 kg, 0.3 per month).



Discussion

We presented a randomized controlled clinical trial testing and comparing a new VR-enhanced CBT approach with standard CBT for the treatment of obese individuals with BED. Starting from the ALH [14]—suggesting that individuals with obesity may be locked to an allocentric (observer view) negative memory of the body that is no longer updated by contrasting egocentric representations driven by perception—we decided to include in the standard CBT approach to BED patients a VR-based protocol aimed at both unlocking the negative memory of the body and modifying its behavioral and emotional correlates.

We tested the efficacy of the VR-enhanced CBT on 66 female obese (BMI>40) patients with BED referred to a 6-week medically managed inpatient program (IP). Participants were randomly divided in three groups: IP (control condition), IP+CBT (15 additional sessions of group and individual CBT sessions), and IP+ECT (15 additional sessions of CBT group and VR individual sessions). We found that the inpatient program was effective in reducing weight in a sample of obese patients with BED in a relatively brief period (6 weeks) and independently from the added CBT and ECT treatments. This was a largely expected result given the highly controlled environment in which patients attended the multimodal program. The VR-enhanced CBT (ECT) delivered during the inpatient

program showed to be effective in the long term. In fact, ECT proved superior to the inpatient program alone in preventing weight regain and in further improving the weight loss after 1-year follow-up. Also, standard CBT was more effective than the inpatient program alone in preventing weight regain. However, follow-up scores with nonresponders assumed to have regained 3.6 kg (0.3 per month) show a significant difference in favor of ECT. Indeed, only ECT was effective in further improving weight loss at 1-year follow-up. On the contrary, control participants regained on average most of the weight they had lost during the inpatient program.

Binge eating episodes decreased to zero during the inpatient program (because of the same reasons explained for weight reduction) but were reported again in all three groups at 1-year follow-up. However, in the ITT analysis with BCF, a substantial regain was observed only in the group who received the inpatient program alone, while both ECT and CBT were successful in maintaining a low rate of monthly binge eating episodes (Figure 3).

Limitations

Some study limitations deserve attention. The first and most important one is the medium-high rate of nonresponders (33.4%); thus, our follow-up results have to be considered with caution. In order to preserve the ITT principle and statistical power, we handled missing data by imputation, assuming that patients who did not provide data at 1-year follow-up had

worsened. With respect to weight, we imputed data according to a plausible empirical rule already used in previous studies [43,44], while missing data in a number of binge eating episodes were handled by an imputation method (BCF) that deserves more critical consideration [45]. Second, the CBT treatment used in this study is a reduced (6-week) version of the Oxford outpatient protocol [20,21]. Obviously, major changes have been made to the original 44-week format, so we cannot compare our results with ones obtained in studies that implemented the original protocol. Third, no other psychosocial variable was measured beyond body image satisfaction. As a reviewer suggested, measures of emotional and social functioning would have added further information on the efficacy of the ECT treatment. We agree and hope that future studies on this novel VR-enhanced treatment will include further measures of psychological, emotional, and social variables that can act as outcomes as well as mediators or moderators of treatment efficacy.

Conclusion

The rising prevalence of weight-related disorders and the lack of a clear solution are pushing eating disorder and obesity researchers to collaborate. In particular, their efforts have identified a critical risk factor: unhealthful weight-control behaviors, such as fasting, vomiting, or laxative abuse, are the common antecedents of both obesity and binge eating. But why do these subjects decide to start such radical weight-control behaviors? The answer underlined both by clinical practice and by research studies is simple: because subjects do not like their bodies.

Given the importance of body image satisfaction for the quality of life of these patients, these findings argue for the potential benefits of treatment strategies for improving body satisfaction. Unfortunately, this vision is not shared by the existing treatment protocols. As noted by Rosen 15 years ago, only a third of cognitive-behavioral treatments assessed in his review addressed body image specifically [46]. The situation still has not changed very much. And when it happens, the focus is more on shape concerns and over-evaluation than on the experience and perception of a fat body [47].

Recently the ALH [14,15] offered a clear link between a negative body image and the etiology of obesity and eating disorders, ie, patients may be locked to an allocentric (observer view) negative image schema of their body that is not updated by contrasting egocentric representations driven by perception. In other words, they cannot win: whatever they do to modify their real body, they will always be present in a virtual body that they hate. This situation has usually two effects: subjects either start more radical dietary restraint or, the opposite, they

decide to stop any form of food control and start “disinhibited” eating behaviors.

In this study, we tested a modified CBT approach that included a VR-based treatment to target this issue. VR can be considered an “embodied technology” for its effects on body perceptions [48]. On one side, different authors showed that is possible to use VR both to induce illusory perceptions, eg, a fake limb [49], by altering the normal association between touch and its visual correlate. It is even possible to generate a body transfer illusion [50]: Slater and colleagues substituted the experience of male subjects’ own bodies with a life-sized virtual human female body. On the other side, it is also possible to use VR to improve body image [51,52], even in patients with eating disorders [23,53,54] or obesity [55,56]

Despite the study limitations discussed above, findings support the hypothesis that the integration of a VR-based treatment, aimed at both unlocking the negative memory of the body and at modifying its behavioral and emotional correlates, may improve the long-term outcome of a treatment for obese BED patients. This result is in line with the recent review [57] about the use of VR for the treatment of body image in eating disorders. The authors concluded their paper with the following words: “VR-based therapies seem to be especially suitable for improving body image both in ED patients and in subclinical samples...All these studies showed significantly greater improvement in measures related with body image when the VR component was added.” (p. 9). The ALH discussed in this paper may be a possible explanation for this result. However, due to the integrated structure of the VR intervention and to the assessment tools used in the trial, we are not able to verify a possible direct effect of the proposed treatment on the allocentric lock (only indirect through a better body satisfaction). So, future studies are needed to further explore and verify this specific hypothesis.

As expected, the VR-based treatment, in comparison with the standard CBT approach, was able to better prevent weight regain but not to better manage binge eating episodes. This is because both CBT and ECT shared the same protocol on problem-solving strategies and cognitive interventions targeting problematic eating.

Future research should focus on using meta-analysis techniques in order to test the real strength of effects found in the clinical studies, and on conducting more controlled studies comparing VR-based treatments with traditional ones. To reach this goal, the “Laboratorio de Enseñanza Virtual y Ciberpsicología” at the School of Psychology of the Universidad Nacional Autonoma de Mexico, in cooperation with the Obesity Unit of the Medica Sur Hospital in Mexico City, have recently started a controlled clinical trial [58,59].

Acknowledgments

The present work was supported by the Commission of the European Communities (CEC), through its IST program (Project VEPSY Updated and Project INTREPID) and by the Italian MIUR FIRB program (Project IVT2010).

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V1.6.2 [59].

[PDF File (Adobe PDF File), 987KB - [jmir_v15i6e113_app1.pdf](#)]

References

1. APA. Diagnostic and Statistical Manual of Mental Disorders, 4th ed, Text Revision. Washington, DC: American Psychiatric Association; 2000.
2. Brownley KA, Berkman ND, Sedway JA, Lohr KN, Bulik CM. Binge eating disorder treatment: a systematic review of randomized controlled trials. *Int J Eat Disord* 2007 May;40(4):337-348. [doi: [10.1002/eat.20370](#)] [Medline: [17370289](#)]
3. Hsu LK, Mulliken B, McDonagh B, Krupa Das S, Rand W, Fairburn CG, et al. Binge eating disorder in extreme obesity. *Int J Obes Relat Metab Disord* 2002 Oct;26(10):1398-1403 [FREE Full text] [doi: [10.1038/sj.ijo.0802081](#)] [Medline: [12355338](#)]
4. Grucza RA, Przybeck TR, Cloninger CR. Prevalence and correlates of binge eating disorder in a community sample. *Compr Psychiatry* 2007;48(2):124-131 [FREE Full text] [doi: [10.1016/j.comppsy.2006.08.002](#)] [Medline: [17292702](#)]
5. Yager J. Binge eating disorder: the search for better treatments. *Am J Psychiatry* 2008 Jan;165(1):4-6. [doi: [10.1176/appi.ajp.2007.07101541](#)] [Medline: [18178748](#)]
6. Haines J, Neumark-Sztainer D. Prevention of obesity and eating disorders: a consideration of shared risk factors. *Health Educ Res* 2006 Dec;21(6):770-782 [FREE Full text] [doi: [10.1093/her/cyl094](#)] [Medline: [16963727](#)]
7. Neumark-Sztainer D, Wall M, Guo J, Story M, Haines J, Eisenberg M. Obesity, disordered eating, and eating disorders in a longitudinal study of adolescents: how do dieters fare 5 years later? *J Am Diet Assoc* 2006 Apr;106(4):559-568. [doi: [10.1016/j.jada.2006.01.003](#)] [Medline: [16567152](#)]
8. Neumark-Sztainer D. Preventing obesity and eating disorders in adolescents: what can health care providers do? *J Adolesc Health* 2009 Mar;44(3):206-213. [doi: [10.1016/j.jadohealth.2008.11.005](#)] [Medline: [19237105](#)]
9. Johnston JM. Eating disorders and childhood obesity: who are the real gluttons? *CMAJ* 2004 Dec 7;171(12):1459-1460 [FREE Full text] [doi: [10.1503/cmaj.1041111](#)] [Medline: [15583186](#)]
10. Stice E, Presnell K, Shaw H, Rohde P. Psychological and behavioral risk factors for obesity onset in adolescent girls: a prospective study. *J Consult Clin Psychol* 2005 Apr;73(2):195-202. [doi: [10.1037/0022-006X.73.2.195](#)] [Medline: [15796626](#)]
11. Stice E, Davis K, Miller NP, Marti CN. Fasting increases risk for onset of binge eating and bulimic pathology: a 5-year prospective study. *J Abnorm Psychol* 2008 Nov;117(4):941-946 [FREE Full text] [doi: [10.1037/a0013644](#)] [Medline: [19025239](#)]
12. Kostanski M, Gullone E. Dieting and body image in the child's world: conceptualization and behavior. *J Genet Psychol* 1999 Dec;160(4):488-499. [doi: [10.1080/00221329909595561](#)] [Medline: [10584324](#)]
13. Friend S, Bauer KW, Madden TC, Neumark-Sztainer D. Self-weighing among adolescents: associations with body mass index, body satisfaction, weight control behaviors, and binge eating. *J Acad Nutr Diet* 2012 Jan;112(1):99-103 [FREE Full text] [doi: [10.1016/j.jada.2011.08.036](#)] [Medline: [22717180](#)]
14. Riva G. Neuroscience and eating disorders: the allocentric lock hypothesis. *Med Hypotheses* 2012 Feb;78(2):254-257. [doi: [10.1016/j.mehy.2011.10.039](#)] [Medline: [22100628](#)]
15. Riva G. The key to unlocking the virtual body: virtual reality in the treatment of obesity and eating disorders. *J Diabetes Sci Technol* 2011 Mar;5(2):283-292 [FREE Full text] [Medline: [21527095](#)]
16. Dodge E, Lakoff G. Image schemas: From linguistic analysis to neural grounding. In: Hampe B, editor. *From perception to meaning: image schemas in cognitive linguistics*. Berlin: Walter de Gruyter; 2005:57-91.
17. Riva G, Gaudio S. Allocentric lock in anorexia nervosa: new evidences from neuroimaging studies. *Med Hypotheses* 2012 Jul;79(1):113-117. [doi: [10.1016/j.mehy.2012.03.036](#)] [Medline: [22546757](#)]
18. Gaudio S, Riva G. Body Image Disturbances in Anorexia: The link between functional connectivity alterations and reference frames. *Biological Psychiatry* 2013 May;73(9):e25-e21. [doi: [10.1016/j.biopsych.2012.1008.1028](#)]
19. Randomization. URL: <http://randomization.com/> [accessed 2013-05-14] [WebCite Cache ID 6Gbi6xS5c]
20. Fairburn CG. *Overcoming binge eating*. New York: Guilford Press; 1995.
21. Fairburn CG, Wilson GT. Cognitive-behavioral therapy for binge eating and bulimia nervosa: A comprehensive treatment manual. In: *Binge eating: nature, assessment, and treatment*. New York: Guilford Press; 1993.
22. Ricca V, Mannucci E, Zucchi T, Rotella CM, Faravelli C. Cognitive-behavioural therapy for bulimia nervosa and binge eating disorder. A review. *Psychother Psychosom* 2000;69(6):287-295. [Medline: [11070440](#)]
23. Riva G, Bacchetta M, Baruffi M, Rinaldi S, Vincelli F, Molinari E. Virtual reality-based experiential cognitive treatment of obesity and binge-eating disorders. *Clinical Psychology and Psychotherapy* 2000;7:209-219.

24. Riva G, Bacchetta M, Cesa G, Conti S, Molinari E. Virtual Reality and Telemedicine Based Experiential Cognitive Therapy: Rationale and Clinical Protocol. In: Riva G, Galimberti C, editors. *Towards CyberPsychology: Mind, Cognition and Society in the Internet Age*. Amsterdam: IOS Press; 2001:273-308.
25. Riva G, Gaggioli A, Villani D, Preziosa A, Morganti F, Corsi R, et al. NeuroVR: an open source virtual reality platform for clinical psychology and behavioral neurosciences. *Stud Health Technol Inform* 2007;125:394-399. [Medline: [17377310](#)]
26. Riva G, Gaggioli A, Grassi A, Raspelli S, Cipresso P, Pallavicini F, et al. NeuroVR 2--a free virtual reality platform for the assessment and treatment in behavioral health care. *Stud Health Technol Inform* 2011;163:493-495. [Medline: [21335845](#)]
27. Riva G, Carelli L, Gaggioli A, Gorini A, Vigna C, Corsi R, et al. NeuroVR 1.5 - a free virtual reality platform for the assessment and treatment in clinical psychology and neuroscience. *Stud Health Technol Inform* 2009;142:268-270. [Medline: [19377165](#)]
28. Schlundt DG, Johnson WG. *Eating disorders: assessment and treatment*. Boston: Allyn and Bacon; 1990.
29. Riva G. Virtual Reality vs. Virtual Body: The use of virtual environments in the treatment of body experience disturbances. *CyberPsychology & Behavior* 1998;1:129-137.
30. Menon ST. Psychological Empowerment: Definition, Measurement, and Validation. *Canadian Journal of Behavioural Science* 1999;31:161-164.
31. Butters JW, Cash TF. Cognitive-behavioral treatment of women's body-image dissatisfaction. *J Consult Clin Psychol* 1987 Dec;55(6):889-897. [Medline: [2891736](#)]
32. Garner DM, Garfinkel PE. Intensive out-patient and residential treatment for bulimia. In: *Handbook of treatment for eating disorders*. New York: Guilford Press; 1997.
33. Smucker MR, Dancu C, Foa EB, Niederee JL. Imagery rescripting: A new treatment for survivors of childhood sexual abuse suffering from posttraumatic stress. *Journal of Cognitive Psychotherapy* 1995;9:3-17.
34. Leuner H. Guided affective imagery (GAI). A method of intensive psychotherapy. *Am J Psychother* 1969 Jan;23(1):4-21. [Medline: [4180780](#)]
35. Garner DM. EDI-Symptom Checklist. Odessa, FL: PAR-Psychological Assessment Resources; 1991.
36. Riva G, Molinari E. Factor structure of the Italian version of the Body Satisfaction Scale: a multisample analysis. *Percept Mot Skills* 1998 Jun;86(3 Pt 1):1083-1088. [Medline: [9656310](#)]
37. Slade PD, Dewey ME, Newton T, Brodie D, Kiemle G. Development of the Body Satisfaction Scale (BSS). *Psychology and Health* 1990;4:213-226.
38. Riva G, Molinari E. Replicated factor analysis of the Italian Version of the Body Image Avoidance Questionnaire. *Percept Mot Skills* 1998 Jun;86(3 Pt 1):1071-1074. [Medline: [9656309](#)]
39. Rosen JC, Srebnik D, Saltzberg E, Wendt S. Development of a Body Image Questionnaire. *Psychological Assessment* 1991;1:32-37.
40. Thompson MA, Gray JJ. Development and validation of a new body-image assessment scale. *J Pers Assess* 1995 Apr;64(2):258-269. [doi: [10.1207/s15327752jpa6402_6](#)] [Medline: [7722852](#)]
41. IBM. SPSS User Manual: Exact Tests. Chicago, IL: SPSS Inc; 2000.
42. Siegel S, Castellan NJ, Jr NJC. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill; 1988.
43. Wadden TA, Berkowitz RI, Womble LG, Sarwer DB, Phelan S, Cato RK, et al. Randomized trial of lifestyle modification and pharmacotherapy for obesity. *N Engl J Med* 2005 Nov 17;353(20):2111-2120. [doi: [10.1056/NEJMoa050156](#)] [Medline: [16291981](#)]
44. Wing RR, Tate DF, Gorin AA, Raynor HA, Fava JL. A self-regulation program for maintenance of weight loss. *N Engl J Med* 2006 Oct 12;355(15):1563-1571. [doi: [10.1056/NEJMoa061883](#)] [Medline: [17035649](#)]
45. Salim A, Mackinnon A, Christensen H, Griffiths K. Comparison of data analysis strategies for intent-to-treat analysis in pre-test-post-test designs with substantial dropout rates. *Psychiatry Res* 2008 Sep 30;160(3):335-345. [doi: [10.1016/j.psychres.2007.08.005](#)] [Medline: [18718673](#)]
46. Rosen JC. Body image assessment and treatment in controlled studies of eating disorders. *International Journal of Eating Disorders* 1996;19:341-343.
47. Fairburn CG. *Cognitive behavior therapy and eating disorders*. New York: Guilford Press; 2008.
48. Spagnolli A, Gamberini L. A Place for Presence. *Understanding the Human Involvement in Mediated Interactive Environments*. *PsychNology Journal* 2005;3:6-15 [FREE Full text]
49. Slater M, Perez-Marcos D, Ehrsson HH, Sanchez-Vives MV. Inducing illusory ownership of a virtual body. *Front Neurosci* 2009 Sep;3(2):214-220 [FREE Full text] [doi: [10.3389/neuro.01.029.2009](#)] [Medline: [20011144](#)]
50. Slater M, Spanlang B, Sanchez-Vives MV, Blanke O. First person experience of body transfer in virtual reality. *PLoS One* 2010;5(5):e10564 [FREE Full text] [doi: [10.1371/journal.pone.0010564](#)] [Medline: [20485681](#)]
51. Riva G. Modifications of body-image induced by virtual reality. *Percept Mot Skills* 1998 Feb;86(1):163-170. [Medline: [9530726](#)]
52. Riva G. Virtual environment for body-image modification: Virtual reality system for the treatment of body image disturbances. *Computers in Human Behavior* 1998;14:477-490.

53. Riva G, Bacchetta M, Cesa G, Conti S, Molinari E. Six-month follow-up of in-patient experiential cognitive therapy for binge eating disorders. *Cyberpsychol Behav* 2003 Jun;6(3):251-258. [doi: [10.1089/109493103322011533](https://doi.org/10.1089/109493103322011533)] [Medline: [12855080](https://pubmed.ncbi.nlm.nih.gov/12855080/)]
54. Riva G, Bacchetta M, Baruffi M, Rinaldi S, Molinari E. Virtual reality based experiential cognitive treatment of anorexia nervosa. *J Behav Ther Exp Psychiatry* 1999 Sep;30(3):221-230. [Medline: [10619546](https://pubmed.ncbi.nlm.nih.gov/10619546/)]
55. Riva G, Bacchetta M, Cesa G, Conti S, Castelnovo G, Mantovani F, et al. Is severe obesity a form of addiction? Rationale, clinical approach, and controlled clinical trial. *Cyberpsychol Behav* 2006 Aug;9(4):457-479. [doi: [10.1089/cpb.2006.9.457](https://doi.org/10.1089/cpb.2006.9.457)] [Medline: [16901250](https://pubmed.ncbi.nlm.nih.gov/16901250/)]
56. Riva G, Bacchetta M, Baruffi M, Rinaldi S, Vincelli F, Molinari E. Virtual reality based Experiential Cognitive Treatment of obesity and binge-eating disorders. *Clinical Psychology and Psychotherapy* 2000;7:209-219.
57. Ferrer-García M, Gutiérrez-Maldonado J. The use of virtual reality in the study, assessment, and treatment of body image in eating disorders and nonclinical samples: a review of the literature. *Body Image* 2012 Jan;9(1):1-11. [doi: [10.1016/j.bodyim.2011.10.001](https://doi.org/10.1016/j.bodyim.2011.10.001)] [Medline: [22119329](https://pubmed.ncbi.nlm.nih.gov/22119329/)]
58. Riva G, Cárdenas-López G, Duran X, Torres-Villalobos GM, Gaggioli A. Virtual reality in the treatment of body image disturbances after bariatric surgery: a clinical case. *Stud Health Technol Inform* 2012;181:278-282. [Medline: [22954871](https://pubmed.ncbi.nlm.nih.gov/22954871/)]
59. 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](https://doi.org/10.2196/jmir.1923)] [Medline: [22209829](https://pubmed.ncbi.nlm.nih.gov/22209829/)]

Abbreviations

ALH: allocentric lock hypothesis
BCF: baseline values carried forward
BED: binge eating disorders
BIAQ: Body Image Avoidance Questionnaire
BSS: Body Satisfaction Scale
CBT: cognitive behavioral therapy
CDRS: Contour Drawing Rating Scale
ECT: VR-enhanced cognitive behavior therapy
IP: integrated multimodal medically managed inpatient program
ITT: intention-to-treat
VR: virtual reality

Edited by G Eysenbach; submitted 15.11.12; peer-reviewed by V Cardi, J Guiterrez, P Bordnick, A Tal; comments to author 22.01.13; revised version received 13.03.13; accepted 12.04.13; published 12.06.13.

Please cite as:

Cesa GL, Manzoni GM, Bacchetta M, Castelnovo G, Conti S, Gaggioli A, Mantovani F, Molinari E, Cárdenas-López G, Riva G
 Virtual Reality for Enhancing the Cognitive Behavioral Treatment of Obesity With Binge Eating Disorder: Randomized Controlled Study With One-Year Follow-up
J Med Internet Res 2013;15(6):e113
 URL: <http://www.jmir.org/2013/6/e113/>
 doi: [10.2196/jmir.2441](https://doi.org/10.2196/jmir.2441)
 PMID: [23759286](https://pubmed.ncbi.nlm.nih.gov/23759286/)

©Gian Luca Cesa, Gian Mauro Manzoni, Monica Bacchetta, Gianluca Castelnovo, Sara Conti, Andrea Gaggioli, Fabrizia Mantovani, Enrico Molinari, Georgina Cárdenas-López, Giuseppe Riva. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 12.06.2013. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Letter to the Editor

The Systematic Removal of Participants Post-Randomization Can Lead to Alternate Explanations of the Results

John A Cunningham^{1,2}, PhD

¹Centre for Mental Health Research, The Australian National University, Canberra, Australia

²Centre for Addiction and Mental Health, Toronto, ON, Canada

Corresponding Author:

John A Cunningham, PhD

Centre for Mental Health Research

The Australian National University

Building 63

Canberra, 0200

Australia

Phone: 61 2 6125 1450

Fax: 61 2 6125 0733

Email: john.cunningham@anu.edu.au

(*J Med Internet Res* 2013;15(6):e128) doi:[10.2196/jmir.2631](https://doi.org/10.2196/jmir.2631)

KEYWORDS

research methods; randomized controlled trial; confound; alternate explanation

The strength of randomized controlled trials is that they allow causal statements to be made about the efficacy of an intervention. This is because the randomization of participants to experimental condition distributes participants with different characteristics to each experimental condition by chance, including variations in the outcome variables of interest. Such randomization allows the use of statistical tests incorporating probability statements regarding the chance that the difference observed between conditions is due to chance (eg, 1 in 20 chance; $P < .05$).

Any systematic removal of participants post-randomization interferes with the assumption of the randomness of participant allocation to experimental condition. This can introduce a potential confound or alternative explanation of the results. The recently published study by Rooke et al [1] removed participants from the analysis who reported other treatment use during the study ($n=5$). The removal of participants was done post-randomization and in a systematic fashion (ie, anyone who reported using other treatment was not included in the analysis). All participants who reported receiving other treatment were in the control condition and were allocated to receive the Web address of an education only website. Even if some participants in the intervention condition had also accessed other treatment, this would still be a systematic, post-randomization removal of participants. However, it is easier to develop alternate

explanations of the findings of this trial because all participants using other treatment were from the control condition.

From one perspective, it is possible that the excluded participants were those who were experiencing the most serious problems with their Cannabis use. If this is the case, it could make it less likely that significant differences would be observed between experimental condition by reducing the variance between participants. Alternatively, perhaps these were the five participants in the control condition who were the most motivated to reduce their Cannabis use. This could mean that participants in the control condition were less motivated, on average, than those in the intervention condition, to do something about their Cannabis use. This could serve as an alternate explanation of the findings in this study.

Given that only five participants were removed out of 225 (or 230 if control and intervention group participant totals are added?) randomized at baseline, it is quite possible that this alternate explanation is untrue. However, the authors do expose themselves to alternate explanations of the findings by the removal of these participants. If the systematic removal of participants post-randomization is deemed necessary, one possible solution would be to run (and report on) the analysis with and without these participants included. Such sensitivity analyses would go a long way towards addressing any possible confounds that may have been introduced.

Conflicts of Interest

None declared.

Reference

1. Rooke S, Copeland J, Norberg M, Hine D, McCambridge J. Effectiveness of a self-guided Web-based cannabis treatment program: randomized controlled trial. J Med Internet Res 2013;15(2):e26 [FREE Full text] [doi: [10.2196/jmir.2256](https://doi.org/10.2196/jmir.2256)] [Medline: [23470329](https://pubmed.ncbi.nlm.nih.gov/23470329/)]

Edited by G Eysenbach; submitted 25.03.13; peer-reviewed by S Rooke; accepted 11.06.13; published 24.06.13.

Please cite as:

Cunningham JA

The Systematic Removal of Participants Post-Randomization Can Lead to Alternate Explanations of the Results

J Med Internet Res 2013;15(6):e128

URL: <http://www.jmir.org/2013/6/e128/>

doi: [10.2196/jmir.2631](https://doi.org/10.2196/jmir.2631)

PMID: [23796473](https://pubmed.ncbi.nlm.nih.gov/23796473/)

©John A Cunningham. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 24.06.2013. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

<https://www.jmirpublications.com/>