Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 11 (2009), Issue 3 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Original Paper

Systematic Review on Internet Support Groups (ISGs) and Depression (1): Do ISGs Reduce Depressive Symptoms?

Kathleen M Griffiths[1], PhD; Alison L Calear[1], PhD; Michelle Banfield[1], BA

Centre for Mental Health Research, The Australian National University, Canberra, Australia

Corresponding Author:
Kathleen M Griffiths, PhD
Centre for Mental Health Research
The Australian National University
Professor & Deputy Director
Canberra ACT 0200
Australia
Phone: +61 2 6125 9723
Fax: +61 2 6125 0733
Email: kathy.griffiths@anu.edu.au

Abstract

Background: Internet support groups (ISGs) enable individuals with specific health problems to readily communicate online. Peer support has been postulated to improve mental health, including depression, through the provision of social support. Given the growing role of ISGs for both users with depression and those with a physical disorder, there is a need to evaluate the evidence concerning the efficacy of ISGs in reducing depressive symptoms.

Objective: The objective was to systematically review the available evidence concerning the effect of ISGs on depressive symptoms.

Method: Three databases (PubMed, PsycINFO, Cochrane) were searched using over 150 search terms extracted from relevant papers, abstracts, and a thesaurus. Papers were included if they (1) employed an online peer-to-peer support group, (2) incorporated a depression outcome, and (3) reported quantitative data. Studies included both stand-alone ISGs and those used in the context of a complex multi-component intervention. All trials were coded for quality.

Results: Thirty-one papers (involving 28 trials) satisfied the inclusion criteria from an initial pool of 12,692 abstracts. Sixteen trials used either a single-component intervention, a design in which non-ISG components were controlled, or a cross-sectional analysis, of which 10 (62.5%) reported a positive effect of the ISG on depressive symptoms. However, only two (20%) of these studies employed a control group. Only two studies investigated the efficacy of a depression ISG and neither employed a control group. Studies with lower design quality tended to be associated with more positive outcomes (P = .07). Overall, studies of breast cancer ISGs were more likely to report a reduction in depressive symptoms than studies of other ISG types (Fisher P = .02), but it is possible that this finding was due to confounding design factors rather than the nature of the ISG.

Conclusions: There is a paucity of high-quality evidence concerning the efficacy or effectiveness of ISGs for depression. There is an urgent need to conduct high-quality randomized controlled trials of the efficacy of depression ISGs to inform the practice of consumers, practitioners, policy makers, and other relevant users and providers of online support groups.

(J Med Internet Res 2009;11(3):e40) doi:10.2196/jmir.1270

KEYWORDS

Depression; consumer participation; Internet; self-help groups

Introduction

Internet support groups (ISGs) provide individuals with specific health problems an opportunity to share experiences and to seek, receive, and provide information, advice, and emotional support online. It has been estimated that millions of people visit online peer-to-peer discussion groups daily [1], and there is evidence that over 28% of Internet users have visited an online support group at least once [2].

Internet users seeking health information frequently access information about depression [3], and online depression groups have been reported to be among the most common ISGs on the

Internet [4]. It is also known that there is a high level of depression among individuals with a physical illness [5]. Thus, many users seeking to join health ISGs may have elevated depressive symptoms or may be at risk of developing depression.

Peer support has been postulated to improve mental health, including depression, through the provision of social support, which alters cognitions, attitudes, self-attributions, and coping, which, in turn, leads to a reduction in depressive symptoms [6]. Given the growing role of ISGs for both consumers with depression and other health conditions, there is a need to evaluate the evidence concerning the effect of these groups on depressive symptoms. One research group has conducted a high-quality, systematic review of studies on the effect of health ISGs on a range of outcomes [1]. The review did not, however, focus on depression outcomes in detail and was confined to articles published prior to October 2003.

The current paper aims to provide a systematic and comprehensive review of the available evidence concerning the effect of ISGs on depressive symptoms regardless of the ISG health condition. A more detailed review of depression ISGs specifically is provided in a companion paper, which reports the scope and findings from all qualitative and quantitative empirical studies of depression ISGs (see [7]).

Methods

Databases

Three databases (PubMed, PsycINFO, Cochrane) were searched using keywords and phrases for the period prior to August 2007. The search was undertaken at two time points, the first in May 2005 and the second in July 2007.

Search Methodology

The search terms and strategies were based on those reported by Eysenbach et al [1], which involve the following concepts: (computer/Internet communication *and* support) *or* e-community venue. In addition, a further 48 relevant search terms were extracted from research papers on ISGs, abstracts extracted by running database searches using the resulting search terms, and an online thesaurus searching for similes of key terms [8].

Study Identification

A multi-step process was employed to select relevant studies for the current review and the review of depression ISGs reported in the companion paper to this study [7] (see Figure 1). In the first stage, each of the 12,692 abstracts returned by the database searches was screened by one of the three authors (AC, MB, KG). The aim of this stage was to screen out clearly irrelevant abstracts and, in particular, to eliminate any reference that clearly did not satisfy the following inclusion criteria:

1. Study discussed or investigated peer-to-peer interaction.
2. Study discussed or investigated at least one of the following: online/electronic support groups, online/electronic social or peer support, online/computer-based communication or interaction, collaborative virtual environments or interventions.
3. The support "group" discussed or investigated was health/psychology related (eg, biological illness, mental illness, health risk factors, bereavement, group counseling), or the article measured a health/psychology related outcome in relation to the support group.

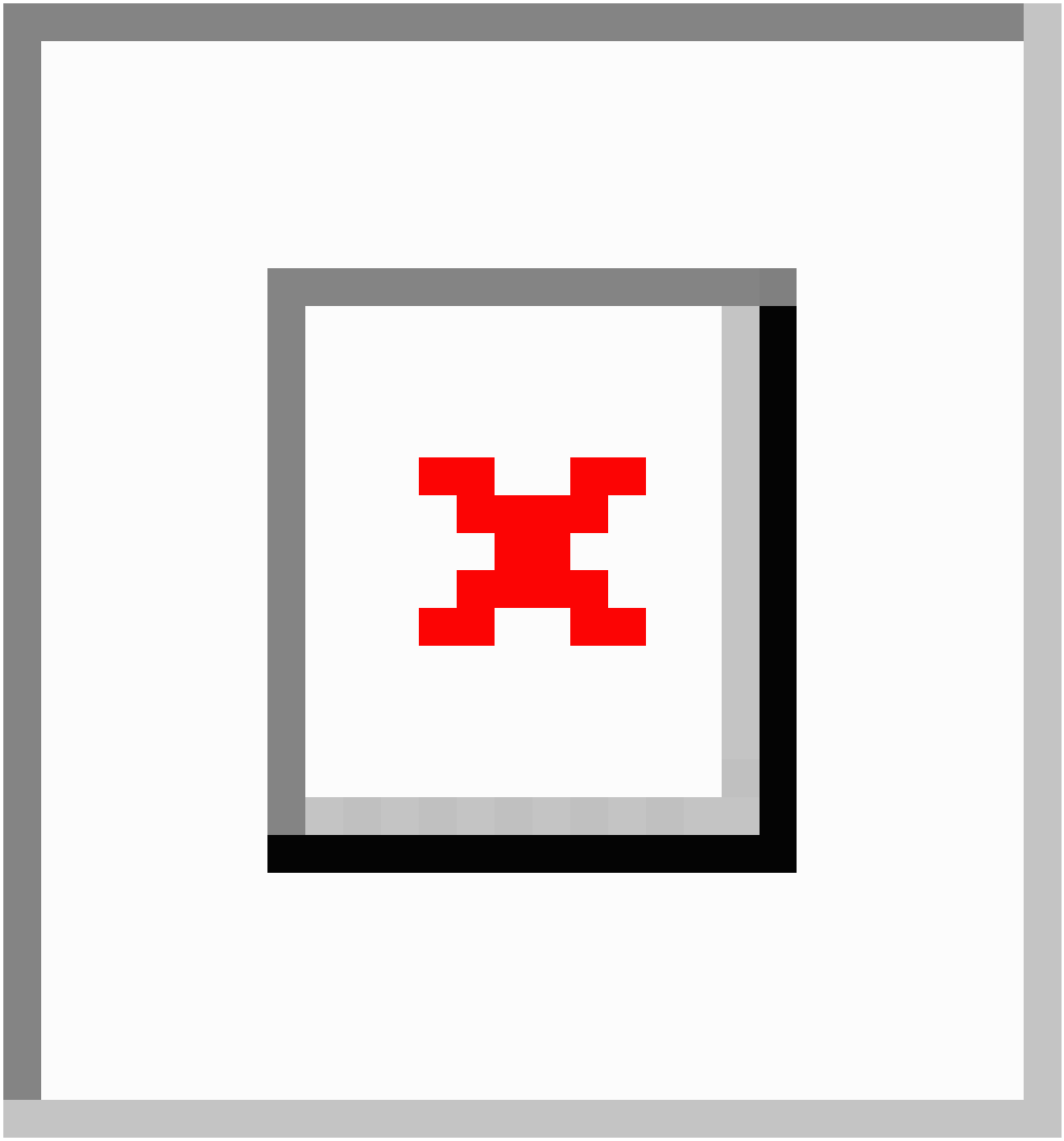
After removing duplicate papers (Stage 2), the remaining abstracts (n = 859) were coded as relevant, not relevant, or possibly relevant according to the following inclusion criteria:

1. Employed an online peer-to-peer support group
2. Incorporated either a depression outcome or involved a unipolar depression ISG
3. Reported either quantitative or qualitative empirical data (Stage 3)

Studies were included whether they incorporated a stand-alone ISG or involved a complex multi-component intervention. Reviews of ISGs satisfying the first two criteria were identified and analyzed separately. Abstracts were coded by one author (AC or KG) and checked by a second author (KG or AC). Any disagreement was resolved by discussion. After excluding the irrelevant abstracts, 158 papers were obtained, read (if in English), and coded against the inclusion criteria by one author (KG). The coding was checked by a second author (AC). Those papers that did not report a depression outcome or did not concern an ISG exclusively devoted to depression were excluded (Stage 4), as were any duplicate papers generated as a result of conducting a two-phase searches process (n = 2). In addition, two papers were judged to be non-English versions of an English-language publication and were excluded [9,10]. Nine other non-English papers of possible but not definite relevance were excluded for pragmatic reasons (cost of translation) [11-19]. It is unclear how many of these would have been retained in the review had they been formally translated. However, one did not satisfy the inclusion criteria based on a translation by the first author [18], and only three of the remaining non-English papers were rated as probable or definite relevance based on the English abstract and a perusal of the content of the tables in the untranslated paper [11,14] or a partial translation supplied by a colleague [19].

The above process yielded a total of 38 relevant papers and five systematic reviews. Two additional relevant papers were identified from the five reviews, and a further two papers cited in at least one of the 38 relevant papers were included among the pool of relevant papers (Stage 5). This resulted in a total of 42 relevant papers of which 31 papers comprising 28 separate trials incorporated a depression outcome (Stage 6) and 11 (studies of depression ISGs) did not. The current paper focuses on the 28 trials reporting a depression outcome.

Figure 1. Study identification flow diagram: PubMed (PM), PsycINFO (PI), Cochrane (C)



Coding of the Included Papers

The 31 papers reporting a depression outcome were independently coded by two raters (KG, AC), and discrepancies were subsequently resolved by discussion between the two raters.

Quantitative studies that included depression outcomes were coded for ISG, participant and study characteristics, and depression outcomes. The ISG characteristics extracted included the psychological or physical condition experienced by members of the group, the format of the ISG (newsgroup, bulletin board, chatroom), whether moderated (yes, no, don't know), and, if so, by whom (consumer, health professional, both, don't know), ISG type (public, research, other restricted access), and ISG origin (United States, Europe, other). Participant characteristics recorded included age (median older than 25 years or 25 years and younger), gender, education, ethnicity, and rurality. Study design characteristics and quality were also coded, including sample size, attrition, design type (randomized controlled trial [RCT], controlled trial, historical control, pre-post, cross-sectional, case series), appropriateness of randomization process and reporting, whether the study employed an intent-to-treat (ITT) analysis (yes, no), and how missing data were treated (last observation carried forward, multiple imputation, other). Each study was also rated as to whether it involved a multi-component design of which the ISG was just one component, or whether the study evaluated a stand-alone ISG or at minimum used a control group that controlled for the

non-ISG components of the intervention. Intervention characteristics recorded included duration of intervention and length of longest follow-up. The depression outcome measures used in each study were recorded, and each sample was rated according to whether it yielded a statistically significant positive outcome. Finally, raters coded the type of publication (thesis, journal, book), country of primary author (United States, Europe, other), and whether consumers were actively involved in the design or conduct of the research.

Analyses

A formal quantitative meta-analysis was not conducted due to the low quality of the studies meeting the inclusion criteria and the heterogeneous nature of the conditions studied. However, the possible role of different characteristics and quality were explored by comparing the characteristics of samples reported to have yielded positive, statistically significant results with those that did not, using a series of Fisher exact tests for categorical attributes and Mann-Whitney tests for other data. For the purposes of this analysis, data were analyzed at the comparison rather than the study level. In addition, for descriptive purposes, where possible, Cohen's d standardized effect sizes were calculated and reported. For uncontrolled studies, the pre-post standardized effect size was calculated from the mean pre-test and post-test scores and standard deviations. For controlled studies, the study effect size was the difference between the pre-post effect size for the control group and the pre-post effect size for the intervention group. In a study involving the comparison between depression scores for high-use compared to low-use Internet users, effect size was based on the standardized difference for the two groups. Effect sizes were not calculated in several instances. Where only the *t* test value for dependent (or equivalent) samples was available [20], no effect size was estimated as such *t* values are based on the standard error of the difference rather than a pooled standard deviation and therefore overestimate the effect size. For the same reason, an effect size was not calculated from the F value of simple effects analysis of residualized change in depression [21]. In addition, effect sizes were not calculated for studies in which only baseline adjusted means [22] and baseline adjusted difference in change [23] were reported and for one study containing apparent inconsistencies in reported sample standard deviations [24].

Results

Of the 28 studies with depression outcomes, five reported results separately for two different populations (patient versus carer [21,25], mothers versus fathers [26], adolescents versus young adults [27], heterogeneous versus homogenous group composition [28]), and one involved two arms differing in intervention duration [24]. Thus, there were a total of 34 samples. In reporting the findings below, the term "samples" will be used to refer to these 34 different populations or arms, and the term "studies" will be reserved to describe the 28 trials.

Study Characteristics

Of the 28 studies with depression outcomes, 16 involved the evaluation of stand-alone ISGs or used a design that controlled for the use of intervention components other than the peer-to-peer component or involved cross-sectional studies of online groups (single component). The remaining studies incorporated a multi-component intervention that comprised the discussion group plus at least one additional component such as health education, skills training, or decision aids. Table 1 and Table 2 present the characteristics of each of the single-component and multi-component studies with a depression outcome. Table 3 summarizes the intervention and design characteristics across studies (ISG format and type, level of evidence) or, where appropriate, across samples (conditions, participant characteristics). Complete data were not available for all variables.

Table 1. Study characteristics and findings for single-component or cross-sectional studies[a]

Study	Participants	Design/Control	Intervention/Nature of ISG	Outcome Measures/Follow-Up	ITT	Completer No. and % Dropout (d/o)	Results/Effect Size[b]	Significant?
Breast Cancer								
Winzelberg 2003 [29] USA	N = 72 women with BC, diagnosed in past 32 mths I = 36; C =36 Recruitment: Advertisements in media and oncology offices	RCT/WLC Randomization method not specified	12-wk Web-based structured newsgroup ISG (Bosom Buddies) One topic/week introduced by psychologist moderator (3 consecutive groups: n = 10, n = 11, n = 15)	CES-D Baseline 12 wks	Yes LOCF	N = 58 (19.4% d/o) I = 28 (22.2% d/o) C = 30 (16.6% d/o) Baseline measures did not predict dropout	Greater reduction in depressive symptoms in ISG group than control ES = 0.60 (completers)	Yes
Lieberman 2003 [30] USA	N = 32 women with BC Recruitment: Online advertisement on BC websites and via media, physicians, hospitals, and community centers	Pre-post	16 week × 1.5 hr chatroom sessions with experienced leader therapist plus 24 hr/day bulletin board access	CES-D Baseline 16-20 wks	No	I = 26 (18.8% d/o) Predictors of nonadherence: poorer coping with anxiety, more fatalistic, pain interfered less with life, less perceived change in relationships/personal strength	Significant reduction in depressive symptoms after use of ISG ES = 1.05	Yes
Lieberman 2005 [31] USA	N = 114 women with BC who joined 1 of 5 frequently used public bulletin boards < 8 wks previously Recruitment: Advertisement on the online bulletin board	Pre-post	6- to 8-mth membership on public BC moderated bulletin board ISG providing emotional support	CES-D "Baseline" 6 mths post baseline	No	6 mths I = 91 (20% d/o) NS difference between completers and non-completers demographics, clinical characteristics, depression severity, posttraumatic growth/psychosocial well-being	Significant reduction in depressive symptoms after use of ISG ES = 4.52	Yes
Lieberman 2006 [20] USA	N = 74 women with BC who joined 1 of 4 frequently used bulletin boards < 8 wks previously Recruitment: Advertisement on the online bulletin board	Pre-post	6- to 8-mth membership on public BC bulletin board providing emotional support No information about moderator status	CES-D "Baseline" 6 mths post baseline	No	6 mths I = 61 (17.6% d/o) Baseline depression severity did not predict dropout	Significant reduction in depressive symptoms after use of ISG	Yes
Rodgers 2005 [32] USA	N = 100 randomly selected women with BC who posted to a BC bulletin board during particular 1-wk period	Pre-post	Variable duration membership (mean 247 days; range 44-1001 days) of public BC bulletin board	Thematic analysis of mood		I = 100 (only followed up while members)	Significant association between frequency of posting and improved mood 43.3% participants improved mood (no data on poorer mood)	Possibly
Mental Disorder								

Study	Participants	Design/Control	Intervention/Nature of ISG	Outcome Measures/ Follow-Up	ITT	Completer No. and % Dropout (d/o)	Results/Effect Size[b]	Significant?
Andersson 2005 [33] Sweden	N = 60 participants with depression (CIDI diagnosis major depression and MADRS-S score 15-30 [mild to moderate depression]) Recruitment: Press release/media	Pre-post arm[c]	10-wk moderated bulletin board ISG	BDI MADRS-S (completer analysis only) Baseline 10 wks 36 wks	Yes LOCF	Post-treatment I = 35 (41.7% d/o) NS between completers and non-completers in baseline depression, quality of life, treatment history, demographic characteristics	NS reduction in depressive symptoms with use of ISG MADRS-S: ES = 0.34 (10 wks) ES = 0.87 (36 wks) (ES values not ITT)	No (ITT and completers)
Houston 2002 [34] USA	N = 103 users of public depression ISGs N = 89, 86.4% depressed on CES-D Recruitment: Requests for volunteers on listservs/bulletin boards	Pre-post	Participation in public listservs/bulletin boards at least 12 mths	CES-D "Depression" = CES-D ≥ 23 at least 1-2 mths after start bulletin board "Baseline" 6 mths 12 mths	No	6 mths I = 72 (30.1% d/o) 12 mths I = 66 (35.9% d/o) Of those depressed at baseline, 79 completed at least 1 follow-up (20.2% d/o) Attrition not predicted by baseline severity of depression, frequency of ISG use, or social support	Resolution of depression greater in more frequent ISG users after adjustment for baseline depression severity/ demographic variables ($P < .03$)	Yes
Golkaramnay 2007 [35] Germany	N = 228 adults discharged from psychiatric hospital with non-psychotic mental disorder I = 114 (with TK insurance - 61 mood disorder) C = 114 (without TK insurance - 59 mood disorder)	CT/TAU	12- to 15-wk exposure to psychotherapist-guided chatroom ISG comprising 8-10 people for 90 mins/wk	LIFE semi- structured interview 1 wk 12 mths	No	I = 97 (14.9% d/o)[d] C = 104 (8.8% d/o)[d]	NS difference in the percentage of ISG and control participants with a diagnosis of disorder at 12 mths follow-up	No
Diabetes								
McKay 2002 [36] USA Glasgow 2003 (12 month f/up) [37] USA	N = 160 primary care DB patients aged 40 to 75 yrs with no Internet access at home or work I_1= 40[e]; C = 40 Recruitment: Letters sent by primary care physicians to their patients with DB	RCT /info control[f] Randomization method not specified	I_1: 10-mth professionally moderated bulletin board/chatroom and information[g]	CES-D 3mths 10 mths	No	3 mths: N= 133 (16.9% d/o) I_1= 30 (25% d/o) C= 33 (17.5% d/o) 10 mths: 18% d/o overall; further details not provided Characteristics of completers and dropouts did not differ	No effect of ISG on reduction in depressive symptoms at either follow-up period ES = 0.15 (3 mths)	No
Renal								

Study	Participants	Design/Control	Intervention/Nature of ISG	Outcome Measures/ Follow-Up	ITT	Completer No. and % Dropout (d/o)	Results/Effect Size[b]	Significant?
Quick 1999 [38] USA	N = 3 people undergoing dialysis for renal disease Recruitment: Dialysis clinics, dialysis websites	MT (single case)	5-wk participation in a pre-existing public email discussion list ISG for renal patients No information about moderation status	BDI 3 time points	Yes	N = 3 (0% d/o)	No improvement in depressive symptoms over time	No
No Disorder								
Gross 2006 [27] USA	N = 77 adolescents aged 11 to 15 yrs N = 81 first-year college students Recruitment: Adolescents - Summer camps/after-school programs College students - Fliers/ announcements in college dorms/halls/classrooms; in person recruitment at halls Rewards for participation/ completing consent form	RCT[h] Randomization method not specified	12 mins of instant online messaging to an unknown peer after experimental induction of low mood in control and intervention group	Dysphoria measure devised for study Baseline and immediate post intervention	No	Adolescents[i]: N = 50 (35.1% d/o, including 1 participant dropped by researchers) College students: N = 60 (25.9% d/o, including 14.8% dropped by researchers)	Adolescents[j]: Mood improvement greater for peer-to-peer intervention group than control College students: No difference in mood change for peer-to-peer group compared to control	A:Yes C:No
Shaw 2002 [39] USA	N = 46[k] introductory psychology university students Recruitment: Advertisement on a psychology course Web page	MT	4-8 wks of online chat sessions with the same anonymous partner Participant provided with topics for the chat	CES-D Pretest, mid-test, post-intervention	No	I = 40 (13%[k] d/o)	Significant reduction in depressive symptoms following use of ISG ES = 0.47	Yes
Morgan 2003 [40] USA	N = 287 (or 256) first-year residential university students Recruitment: Postal notification followed by email	XS	Chatroom unspecified /instant messaging	Modified CES-D (11-item, Iowa version)	N/A	N/A	Significant correlation between chatroom hrs and depressive symptoms r = −.13, $P < .05$ Increased chatroom hrs predicted decreased depression after controlling for demographic variables/social support $P < .01$	Yes
Sun 2005 [41] USA	N = 2373 7th grade students (age 11 to 16 yrs) Recruitment: Invitation via school	XS	Chatroom unspecified	Not specified	N/A	N/A	Daily chatroom users more depressed than those with Internet access who did not use chatrooms OR 1.2, $P < .05$	Yes (-ve effect)

Study	Participants	Design/Control	Intervention/Nature of ISG	Outcome Measures/Follow-Up	ITT	Completer No. and % Dropout (d/o)	Results/Effect Size[b]	Significant?
Campbell 2006 [42] Australia	N = 188 self-selected global sample of online users of whom 137 were frequent chat users and 51 were not Recruitment: Passive recruitment via website advertisement (eg, on APA website)	XS	High chatroom (unspecified) use Control low chatroom (unspecified) use	ZDS DASS	N/A	N/A	NS difference in depressive symptoms for high chatroom compared to low chatroom use ES = −.06 (ZDS) ES = 0.02 (DASS)	No
Kang 2007 [43] USA	N = 158 chatroom users from US university community (57% female) Recruitment: Not reported	XS	Chatroom unspecified	CES-D Kraut depression items	N/A	N/A	Higher chatroom use predicted lower depression β = −0.29, $P <$ 0.001	Yes

[a] APA = American Psychological Association; BC = breast cancer; BDI = Beck Depression Inventory; C = control sample size; CES-D = Center for Epidemiologic Studies Depression Scale; CIDI = Composite International Diagnostic Interview; CT = controlled trial; DASS = Depression Anxiety Stress Scales; DB = diabetes; ES = effect size; I = intervention sample size; ITT = intent to treat; LIFE = Longitudinal Interval Follow-up Evaluation; LOCF = last observation carried forward; MADRS-S = Montgomery-Asberg Depression Rating Scale; MT = multiple time points; N/A = not applicable; OR = odds ratio; NS = no significant difference; RCT = randomized controlled trial; TAU = treatment as usual; TK = Techniker Krankenkassde; WLC = wait list control; XS = cross-sectional; ZDS = Zung Depression Scale.

[b] Pre-post standardized effect size (for pre-post design) or difference between intervention and control pre-post effect sizes (for controlled designs).

[c] This study was an RCT involving an intervention group comprising CBT self-help and an ISG and a control group involving an ISG alone. This design does not permit an evaluation of the effect of ISG alone. Therefore, only the data for the control group (pre-post) are presented here.

[d] Did not complete both baseline and follow-up assessments; other dropout information not available.

[e] Also, I_2 = 40, I_3 = 40.

[f] Online articles on diabetes (information only).

[g] Also two other conditions: I_2: access to professional coach and blood glucose tracking; I_3: a combination of I_1 & I_2.

[h] Participants randomized to one of three groups: (1) control, (2) intervention, (3) intervention group partners.

[i] These figures are for participants across all groups including dyad partners who had not undergone negative mood induction. Sample size and dropout figures were not available for the groups separately.

[j] Outcome measures recorded and analyzed for mood induction intervention and control samples only.

[k] Unclear if n = 46 before or after consent.

Table 2. Study characteristics for multi-component interventions[a]

Study	Participants	Design/Type of Control	Intervention/Nature of ISG	Outcome Measures/ Follow-Up	ITT	Completer No. and % Dropout (d/o)	Results/Effect Size[b]	Significant?
Cancer								
Owen 2003 [22] USA	N = 59 women with BC I = 29; C = 30 Recruitment: Contact with patients in medical oncology clinics, advertisements in hematology/oncology outpatient clinic, health websites, community nurse referral, media $10 for completing each survey	RCT/WLC Randomization: Random number generator	12-wk SURVIVE online program comprising health professional, moderated bulletin board group, cancer information, resources, self-management advice, art/poetry forum, structured coping skills exercises (including stress management, assertiveness, and structured problem solving training) Up to 20 participants per group	HADS Baseline 12 wks	No	I = 25 (13.8% d/o) C = 27 (10% d/o)	NS difference in baseline adjusted mean at 12 wks for intervention and control groups	No
Van Den Brink 2007 [23] Netherlands	N = 184 people post-surgery for head or neck cancer I = 39; C = 145 Recruitment: Tertiary university hospital–treated patients recruited by doctor independent of treating physicians	CT/TAU	6-wk electronic health information support system comprising peer-to-peer forum and email communication; information and monitoring via electronic questionnaire	"Feelings of depression" Baseline 6 wks 3 mths	No	N = 163 (11.4% d/o) I = 35 (10.3% d/o) C = 128 (11.7% d/o)	NS baseline adjusted difference in change at 6 or 3 mths for intervention compared to control groups	No
Neurological								
Brennan 1995 [44] USA	N = 102 caregivers of people with Alzheimer's disease I = 51; C = 51 Recruitment: Research registry, support groups	RCT/TAU Randomization: Not specified	12-mth access to bulletin board moderated by nurse who posted messages to "foster systematic group cohesion" and information and decision support (expert Q&A)	CES-D Baseline 12 mths (intervening variable)	No	I = 47 (7.8%[c] d/o) C = 49 (3.9% d/o)	Depression was treated as a intervening variable rather than an outcome ES = 0.24	N/R
Liebermann 2005[d] [28,45] USA	N = 66 or 65 patients with PD assigned to: Heterogeneous (Het) groups - variable age and time since diagnosis Homogenous (Hom) groups - homogenous age and time since diagnosis[d] Recruitment: Fliers to support groups, PD clinic, practitioners, online posts, newsletter	Pre-post	20 wks × 1.5 hrs weekly health-professional moderated chatroom and bulletin board available at all times and Q&A weekly health education session with an expert	CES-D Baseline 20 weeks	No	Dropout rates could not be calculated separately for Hom and Het Combined: I = 32 (39% d/o)[d] NS differences in baseline measures between dropouts and completers	Significant reduction in depressive symptoms following intervention involving Hom but not Het ISG[d]	Hom: Yes Het: No

Study	Participants	Design/Type of Control	Intervention/Nature of ISG	Outcome Measures/ Follow-Up	ITT	Completer No. and % Dropout (d/o)	Results/Effect Size[b]	Significant?
Chronic Illness								
Battles [46] USA	N = 32 children (age 8-19 yrs) with serious chronic illness (HIV, cancer, granuloma, neurofibromatosis) participating as residential out patients in pediatric clinical trials at the NIH Recruitment: Playroom staff at NIH residential center identified potentially eligible participants Researchers approached eligible participants/parents	(1) Restricted randomly alternating (A, B) treatment design Control = normal playroom activity (2) Pre-post	4 × 30 min sessions on the STARBRIGHT World (SBW) program comprising network connection to other children in a hospital (video) Connect/Find a Friend and information about medical conditions and entertainment and distraction Sessions administered across multiple NIH residential visits over unspecified time period	Depression Analogue Scale CBCL-anxious depressed (parent) Usefulness in reducing sadness-depression (parent) Pre-post session Pre-post intervention	d/k	d/k	NS improvement in depression ratings or symptoms 24% parents reported positive effects of the program on mood Estimated ES (CBCL) = −0.06	No
Hill 2006 [47] USA	N = 120 female, rural residents (35 to 65 yrs) with chronic illness (diabetes/rheumatoid condition/ heart condition/multiple sclerosis/cancer) I = 61; C = 59 Recruitment: Mass media, agency and service organization newsletter, and word of mouth	RCT[d] Randomization: Method not specified	22-wk professionally moderated online support group and online health information modules The support group was described as an “asynchronous chatroom”	CES-D Baseline 22 wks	No	I = 43 (29.5% d/o) C = 57 (3.4% d/o)	NS differences in reduction in depressive symptoms in intervention compared to the control group ES = 0.15	No
Carers								
Bragadottir 2004 [26] USA thesis, Icelandic sample	N = 21 parents of children who had completed cancer treatment within past 5 yrs Mothers: I = 11 Fathers: I = 10 Recruitment: From files of Icelandic hospital responsible for treating children with cancer	Pre-post	4-mth access to health professional–moderated mailing list Professionals facilitated and joined in group discussions, answered questions, directed parents to resources, corrected misconceptions/misinformation, monitored appropriateness of discussions	SCL-90 depression subscale Baseline 3 mths 4 mths	No	3 mths and 4 mths N= 16 (23.8% d/o) Mothers: I = 8 (27.3% d/o) Fathers: I = 8 (20% d/o)	Mothers: NS reduction in depressive symptoms ES = −0.10 (3 mths) ES = 0.20 (4 mths) Fathers: NS reduction depressive symptoms ES = −0.22 (3 mths) ES = 0.40 (4 mths)	Mothers: No Fathers: No

Study	Participants	Design/Type of Control	Intervention/Nature of ISG	Outcome Measures/ Follow-Up	ITT	Completer No. and % Dropout (d/o)	Results/Effect Size[b]	Significant?
Carers and Heart Recipients								
Dew 2004 [21] USA	N = 124 heart recipients and family caregivers Recipients: I = 24; C = 40 Caregivers: I = 20; C = 40 Recruitment: Letter from transplant team asking if had Internet access Those with access "approached" to participate	Controlled/ "Historic" TAU comparison group enrolled in other longitudinal studies and matched for demographic distribution and assembled before or after intervention	4-mth HeartNet programs comprising discussion groups (online moderated bulletin boards, separate caregiver and recipient boards) and interactive online stress and medical regimen management skills training grounded in CBT principles and Ask an Expert (online questions to transplant team expert plus Q&A Library plus archived responses to Ask and Expert plus Health living tips plus Resources plus References Library)	SCL-90 Depression subscale Baseline, 4 mths (I), and 4-6 mths (C)	No	Recipients: I= 20 (16.7% d/o) C = 34 (15% d/o) Caregivers: I= 17 (15% d/o) C = 34 (15% d/o)	Recipients: Receiving intervention showed a greater reduction in depressive symptoms than the control group Caregivers: NS difference in reduction in depressive symptoms in intervention compared to the control group	Recipients:Yes Caregivers: No
Diabetes								
McKay 2001 [48] USA	N = 78 sedentary people with type 2 diabetes aged 40 years or older I = 38; C = 40 Recruitment: Email postings to online diabetes groups and websites	RCT /online information, blood glucose tracking Control Randomization: Automatic system allocated	8-wk D-Net Active Lives program comprising tailored online physical activity program with tracking of daily physical activity, information about a physical activity plus online personal coach counseling plus health professional moderated online peer support (Active Lives Support Group)	CES-D Baseline 8 wks	No	N = 68 (13% d/o) I = 35 (7.9% d/o) C = 33 (17.5% d/o) Predictors of drop out: None	NS difference in reduction in depressive symptoms in intervention compared to control group ES = 0.35	No (*P* = .10)
HIV								
Gustafson 1994 [49] USA Gustafson 1999 [24] USA	N = 219; I = 118; C = 97 with HIV 3-mth intervention: I = not specified; C = not specified 6-mth intervention: I = not specified; C = not specified Recruitment: Posters, newspaper advertisement, HIV clinics/organizations Paid to complete surveys	RCT/TAU Randomization: Independent third party using random number table	6 mths (Cohort 1) and 3 mths (Cohorts 2 and 3) CHESS program comprising online facilitated bulletin board discussion group plus Q&A plus Instant Library (information articles) plus Ask an Expert (communication with medical experts) plus Getting help/support plus Referral Directory plus Personal stories plus assessment (of lifestyle risks) plus Decision Aid plus Action Plan for implementing decisions	MOSdepression subscale 3-mth Int: Baseline, 2 mths, 5 mths 6-mth Int: Baseline, 2 mths, 5 mths, 9 mths	No	Dropout rates could not be calculated separately for 3-mth and 6-mth intervention groups All cohorts at 2 mths: I = 97 (17.7% d/o) C = 90 (9.3% d/o) All cohorts who "completed trial": I = 94 (21% d/o) C = 89 (9.2% d/o)	NS differences in reduction in depressive symptoms in intervention compared to control group for any follow-up/cohort combination	3-mth Int (5-mth f/up): No 6-mth Int (9-mth f/up): No

Study	Participants	Design/Type of Control	Intervention/Nature of ISG	Outcome Measures/ Follow-Up	ITT	Completer No. and % Dropout (d/o)	Results/Effect Size[b]	Significant?
Mental Disorder								
Taylor 2006 [50] USA	N = 480 college women (18 to 30 yrs) at high risk of developing an eating disorder I = 244; C = 236 Recruitment: Flyers at colleges, campus mailings, mass media	RCT/WLC Randomization: Stratified by school; computer-generated sequences produced by study coordinator	8-wk professionally modified bulletin board and cognitive behavioral intervention	CES-D Baseline 8 weeks 60 weeks	No	I = 191 (21.7% d/o)[e] C = 198 (16.1% d/o)[e] NS demographic or baseline differences between completers and non-completers	NS difference in reduction in depressive symptoms in intervention compared to control group ES = 0.04 (8 wks) ES = 0.11 (60 wks)	No ($P < .07$)
IVF								
Tuil 2006 [25] Netherlands	N = 244 participants undergoing IVF or ICSI treatment in authors' hospital Males: I = 61; C = 61 Females: I = 61; C = 61 Recruitment: From author IVF clinic	RCT "Randomization": Alternating allocation to intervention or control	Access to professionally moderated bulletin board and chatroom (for communication with peers and professionals) plus information and access to own records during period of IVF/ICSI treatment cycle	Beck Depression Index for Primary Care Baseline Post-intervention	No	Males: I= 51 (16.4% d/o) C = 38 (37.7% d/o) Females: I= 51 (16.4% d/o) C = 40 (34.4% d/o)	Males: ES = −0.25 Females: ES = 0.18	Males: No Females: No

[a] BC = breast cancer; C = control sample size; CBCL = Child Behavior Checklist; CBT = cognitive behavioral therapy; CES-D = Center for Epidemiologic Studies Depression Scale; CT = controlled trial; d/k – don't know; ES = effect size; HADS = Hospital Anxiety & Depression Scale; I = intervention sample size; ITT = intent to treat; MOS = Medical Outcomes Study; NIH = National Institutes of Health; N/R = not reported; PD = Parkinson's disease; RCT = randomized controlled trial; SCL-90 = Symptom Checklist 90; TAU = treatment as usual; WLC = wait list control.

[b] Pre-post standardized effect size (for pre-post design) or difference between intervention and control pre-post effect sizes (for controlled designs).

[c] Includes three (5.9%) dropouts "not able to have computer installed."

[d] Due to apparent inconsistencies within and between the two papers on this study, effect sizes have not been computed, individual sample sizes are not reported, and individual dropout rates not computed.

[e] Computed for completers of CES-D only; data for overall completers not available.

Table 3. Study and sample characteristics[a]

Study (Sample[c]) Variable	Total n = 28 (n = 34)[c]	Single Component n = 16 (n = 17)[c]	Multi-Component n = 12 (n = 17)[c]
Source of study			
Journal article	24 (87.5)	14 (87.5)	10 (83.3)
Thesis	4 (14.3)	2 (12.5)	2 (16.7)
Country of senior author			
United States	23 (81.2)	13 (81.3)	10 (83.3)
Europe	4 (14.3)	2 (12.6)	2 (16.7)
Australia	1 (3.6)	1 (6.3)	-
Level of evidence			
Randomized controlled trial	10 (35.7)	3 (18.8)	7 (58.3)
Controlled trial	2 (7.1)	1 (6.3)	1 (8.3)
Historic control	1 (3.6)	-	1 (8.3)
Pre-post	9 (32.1)	7 (43.8)	2 (16.7)
Pre-post + single case randomization	1 (3.6)	-	1 (8.3)
Cross-sectional	4 (14.3)	4 (25.0)	-
Case series	1 (3.6)	1 (6.3)	-
ISG format			
Bulletin Board	9 (32.1)	4 (25.0)	5 (41.7)
Chatroom	5 (17.9)	5 (31.3)	-
Mailing list/newsgroup	2 (7.1)	1 (6.3)	1 (8.3)
Instant Messaging	2 (7.1)	1 (6.3)	1 (8.3)
Combination	6 (25.0)	3 (18.9)	3 (25.0)
Mailing list or bulletin board	2 (7.2)	2 (12.5)	-
Unclear	2 (7.2)	-	2 (16.6)
ISG origin			
Public, accessible	9 (32.1)	9 (56.3)	0 (0)
Closed, research ISG	17 (60.7)	7 (43.8)	10 (83.3)
Restricted access hospital	2 (7.1)	-	2 (16.7)
Moderation status			
Moderated	14 (50)	6 (37.5)	8 (66.7)
Some moderated	1 (3.6)	1 (56.3)	-
Not specified	13 (46.4)	9 (6.3)	4 (33.3)
Type of moderation	(n = 15)	(n = 7)	(n = 8)
Health professional	11 (73.3)	5 (71.4)	6 (75)
Don’t know	4 (26.7)	3 (28.6)	2 (25)
Median duration intervention (n = 29)[b]	16 wks (n = 23)	15.5 wks (n = 10)	17 wks (n = 13)
Median longest follow-up (n = 29)[b] (from intervention commencement)	22 wks (n = 22)	26 wks (n = 12)	18.5 wks (n = 10)
Condition (n = 34)[c]			

Study (Sample[c]) Variable	Total n = 28 (n = 34)[c]	Single Component n = 16 (n = 17)[c]	Multi-Component n = 12 (n = 17)[c]
Cancer	7 (20.6)	5 (29.4)	2 (11.8)
No disorder	7 (20.6)	7 (41.2)	-
Diabetes	2 (5.9)	1 (5.9)	1 (5.9)
Carers	4 (11.8)	-	4 (23.5)
Chronic illness	2 (5.9)	-	2 (11.8)
Neurological	2 (5.9)	-	2 (11.8)
Depression	2 (5.9)	2 (11.8)	-
Other mental disorder	2 (5.9)	1 (5.9)	1 (5.9)
Cardiovascular	1 (2.9)	1 (5.9)	1 (5.9)
Renal	1 (2.9)	1 (5.9)	-
HIV/AIDS	2 (5.9)	-	2 (11.8)
IVF	2 (5.9)	-	2 (11.8)
Participant mean/median age (n = 34)[c]			
11 to 17 yrs	3 (8.8)	2 (11.8)	1 (5.9)
18 to 25 yrs	4 (11.8)	3 (17.6)	1 (5.9)
26 to 40 yrs	5 (14.7)	2 (11.8)	3 (17.6)
41 to 65 yrs	11 (32.4)	4 (23.5)	7 (41.2)
Not certain	11 (32.4)	6 (35.3)	5 (29.4)
Gender (n = 34)[c]			
> 70% women	16 (47.1)	9 (56.3)	7 (46.7)
> 70% men	4 (11.8)	-	4 (25.0)
Neither gender > 70%	11 (32.4)	7 (43.8)	4 (50)
Don't know	3 (8.8)	1 (6.3)	2 (11.8)
Rural (n = 34)[c]			
> 50% rural	1 (2.9)	0 (0)	1 (5.9)

[a] Values are no. (%) unless otherwise specified.

[b] Multiple samples receiving different intervention durations treated separately (one study: [24])

[c] Multiple samples treated separately (six studies: [21,24-28])

Origin

The majority of studies were reported in published journal articles, and, in most cases, the senior author was located in the United States.

Interventions

The studies primarily employed bulletin boards, chatrooms, or mailing lists, either alone or in combination (see Table 3). Approximately two-thirds were closed ISGs, typically developed for research purposes. Half of the studies specified that the ISGs were moderated, and of these the majority of moderators were health professionals. The duration of the interventions ranged from 12 minutes to 12 months (median 16.5 weeks), and length of time to follow-up ranged from immediately post-intervention to 12 months post-intervention.

Participants

More samples were focused on ISGs for breast cancer than any other condition. In addition, a significant percentage of the samples related to depression and ISG use in those without a physical or psychological condition. As noted above, only two samples were exposed to depression ISGs. The median age of participants in the samples typically fell between 26 and 65 years. Some of the samples comprised college-aged or younger adolescents. None was concerned specifically with older people, although the median age of one sample was 64 years [28]. Significantly, only a minority of samples focused on men, whereas almost one half contained a predominance of, or all, women. Only one study focused on rural participants [47]; two others mentioned the inclusion of some rural residents [26,30].

Outcome Measures

Half of the studies (n = 14) used the Center for Epidemiologic Studies Depression Scale (CES-D) as an outcome measure, with the next most common measures (with two trials each) being the Symptom Checklist 90 (SCL-90) and the Beck Depression Inventory (BDI). Each of the remaining measures was administered in one trial only.

Study Quality

One third of the studies involved an RCT, and almost half of the 28 studies employed a control group. The majority of the remaining studies used a pre-post design. Of the 23 studies that used at least a pre-post design, only three (13%) used an ITT design, with a further study neither specifying if an intent-to-treat design was employed nor indicating the level of dropout if any [46]. Two of the four ITT studies [29,33] used the last observation carried forward method for treating missingness. The third inferred mood from initial and final posts on a bulletin board, thus ensuring that there was no dropout [32]. No study used multiple imputation for estimating missingness. Of the nine studies said to have employed an RCT design, only three [22,24,48] both adequately specified the randomization procedure and employed an appropriate method of randomization [51].

Intervention and control sample sizes ranged from 10 to 244 (median 46) and 30 to 236 (median 51), respectively, for samples derived from studies of at least pre-post test quality. Cross-sectional study sample sizes ranged from 158 to 2373 (median 230). Dropout among samples in studies of at least pre-post test quality ranged from 7.9% to 41.7% and 0% to 37% for intervention and control conditions, respectively. Of the 22 studies of at least pre-post design with some dropout, 46% (n = 10) compared the characteristics of completers and non-completers. All but one of these (n = 9, 90%) reported no difference in baseline characteristics for these groups.

ISG Efficacy for Depression

The outcomes for single and multiple studies are discussed separately.

Single-Component Studies

Of the 17 intervention samples (16 studies) involving a peer-to-peer component alone or a cross-sectional design, 10 (59%) yielded a positive effect of the ISG on depressive symptoms. However, only two of these involved a controlled trial.

The largest number of single-component samples involved women with breast cancer (n = 5) [20,29-32]. Of these, four yielded significant effects of moderate to large size [20,29-31], and the fifth was associated with a small, significant association between board use and improved mood [32]. However, only one of these trials employed a controlled design [29].

Three samples (three studies) involved ISGs comprising members with a mental disorder, two of them depression [33-35]. One of these produced a positive result. In particular, Houston et al [34] found that more frequent depression ISG users were significantly more likely to recover from depression after adjustment for baseline depression severity and demographic variables. However, the study did not include a control group. The second depression ISG comparison involved the control arm of an RCT of an online cognitive behavior therapy intervention for depression in which a research bulletin board was used as a control condition [33]. There was no significant effect of the bulletin board.

There were two other single-component samples (2 studies) involving medical conditions, one of them involving a trial of an ISG for diabetes [36,37], the other the use of an ISG for renal patients undergoing dialysis [38]. The ISG did not produce an effect on depressive symptoms in either of these studies, but the latter involved only three cases.

Finally, seven samples (six studies) involved people with no psychological or physical disorder [27,39-43]. Three samples (two studies) involved experimental studies of the effect on mood of online communication between peer dyads [27,39]. Two of these reported a positive effect of the dyad on mood. The remaining four samples (four studies) involved cross-sectional studies of survey data designed to investigate the association between frequency of chatroom use and mood in community samples. Two of these studies involved university communities and found that higher chatroom use predicted lower depression [40,43]. A third, cross-sectional study of general users on the Internet did not find an association between frequency of use and mood but employed a dichotomized measure of frequency and may therefore have lacked statistical power [42]. The final study, which involved adolescents aged 11 to 16 years, found a reverse effect, with higher Internet use being associated with a higher level of depressive symptoms [41]. In summary, there is weak evidence that chatroom use among people without a disorder may be associated with lower levels of depression, but the quality of evidence is poor and the findings inconsistent.

Multi-Component Studies

Of the 17 samples (12 studies) that involved intervention components in addition to the ISG, only two (12%) reported a positive effect [21,28]. The first, involving a homogenous group of patients with Parkinson's disease, employed a pre-post design only and incorporated a health professional education component as well as the ISG [28]. The second, involving heart recipients, employed a historical control differing in depression severity and comprised many potentially active components in addition to the ISG, including stress skills training [21].

Association Between Positive Results and Study Characteristics

Multi-component studies were significantly less likely to yield significant, positive outcomes than stand-alone interventions and cross-sectional studies (Fisher exact test, $P = .01$). Breast cancer ISGs were more successful than other ISGs (Fisher exact test, $P = .02$), but most of the breast cancer studies originated from a single research group. Outcome was not affected by the use of synchronous (chatroom) compared to asynchronous (bulletin board, listserv/newsgroups) ISGs (Fisher exact test, $P = .99$), whether or not the study reported using a moderator (Fisher exact test, $P = .72$) or whether the board was public, research, and/or restricted access (Fisher exact test, $P = .11$).

There was no effect on outcome for the duration of the intervention (Mann-Whitney U = 57, *P* = .23) or the length of follow-up (Mann-Whitney U = 75.5, *P* = .83). Nor was there a significant association between age (25 years and younger vs older) and success, but there were few studies of young people (Fisher exact test, *P* = .64). Considering only the samples that were predominantly comprised of males (n = 4) or females (n = 16), there was no association between outcome and sex (*P* = .59), but the sample size of males was very small.

With respect to study quality, there was a trend toward an association between lower design quality and positive outcomes, with 19% (n = 3) of samples involving controlled comparisons (RCT, controlled trial, historic control) and 53% (n = 9) of uncontrolled effects yielding significant positive findings. However, this association fell short of statistical significance (Fisher exact test, *P* = .07). A similar non-significant trend (Fisher exact test, *P* = .13) was noted for samples involving RCTs compared to other designs. In the latter case, only 17% (n = 2) of the RCTs yielded a positive effect and none of these employed an ITT design. By contrast, 48% (n = 10) of the lower-quality trials yielded significant positive outcomes. There was no association between total sample size of study intervention groups and outcome (Mann-Whitney U = 62, *P* = .26).

Discussion

The most salient finding of this review was the paucity of high-quality studies of the impact of depression or other ISGs on depression outcomes. Only a minority of the identified studies employed a control group, and two-thirds of RCTs either failed to use an adequate method of randomization or failed to specify the method of randomization. In addition, only 13% of studies of at least pre-post quality used an ITT analysis, and no study used multiple imputation for treating missingness. This low level of quality is a cause for concern, particularly given the trend toward an association between significant positive findings and low design quality.

Despite the apparent popularity of the Internet as a source of support for people with depression, there were only two studies of the effectiveness or efficacy of depression ISGs in improving mood. One comprised the control arm in a study of the effectiveness of a psychological therapy, and the other involved an uncontrolled multi-time-point study of an existing public depression ISG. Although the findings from the latter study were promising, neither study was of sufficient quality to evaluate whether depression ISGs improve or do not improve depression outcomes. Clearly, there is a need to undertake an RCT of the effect of a depression ISG on depression status.

Although there were more studies of the effect on depression for ISGs for conditions other than depression, many of these studies were of low quality and almost 50% employed multi-component interventions of which the ISG was only one component. Indeed, only two studies employed both a controlled design and a single-component intervention [27,29]. The first involved a structured 12-week breast cancer newsgroup intervention facilitated by a psychologist. There was a greater reduction in depressive symptoms among the ISG than the control group using ITT analyses. The second involved a sample of well adolescents and a sample of well college students who, after exposure to a negative mood induction manipulation, were provided with the opportunity to interact online with an unknown peer. There was an improvement in mood for the adolescents assigned to online peer interaction relative to control adolescents, but no such effect for college students. Thus, the results of the two highest quality studies are encouraging and suggest that further studies of ISGs of all types are warranted.

The finding that breast cancer ISGs were significantly more likely to be associated with positive results than ISGs of other types requires further investigation given that women with breast cancer are known to be at increased risk of depression [52]. If found to be effective in reducing depressive symptoms, such ISGs could provide an important mental health self-care and prevention tool for women with breast cancer. However, the status of the current results is unclear given that the majority of findings were derived from one research group and the studies were typically of low quality.

The finding that chatroom use tends to be associated with lower levels of depression among participants without depression or other medical conditions raises the possibility that chatroom usage may protect against depression in universal samples of members of the community. However, much of the evidence is based on cross-sectional surveys. Thus, the direction of causation cannot be determined, and chatroom usage may be associated with other behaviors and these rather than the chatroom use may mediate the depression levels.

Theoretically, online support groups could be particularly relevant and appropriate for users who are isolated or not able to access conventional or face-to-face services, either due to lack of mobility or geographic location. It is therefore of some concern that none of the studies investigated ISGs among older people and that only one study specifically focused on the effectiveness of an ISG for rural participants.

Limitations

A limitation of this study is that it does not include trials published after July 2007. To investigate this, a further search was conducted by the first author incorporating the time period from August 2007 to May 2009 and using the same search terms employed in the reported searches but limiting results to those incorporating the terms “depression” or “depressive” or “mood.”

After excluding a published study reporting data from a dissertation already incorporated into the review [53], 14 new relevant papers were identified. Of these, six involved experimental studies [54-59] and the remainder were non-experimental [60-67]. No new descriptive studies of depression ISGs were identified. Of the experimental studies, all but two [54,55] incorporated potentially active components in addition to an ISG. Only one of the six employed an ITT design [58], and although three were RCTs [56-58], none specified the method of randomization. The remaining three experimental studies were controlled trials [54,55,59], but one employed a non-contemporaneous control [54]. Of the two single-component studies, one involved an ISG for Spanish-speaking immigrant women with breast cancer [54]

and the other an ISG for Asian American women with a lesbian or bisexual orientation [55]. Neither resulted in a positive effect on depressive symptoms relative to a control.

Of the four multi-component trials [56-59], three reported a greater reduction in depressive symptoms in the intervention group [57-59]. The first of these studies involved an ISG and educational films for people with chronic pain or burnout ([57], RCT), but the effect was not sustained at follow-up. The second employed a discussion group in addition to a therapist-facilitated online group and an offline cognitive behavioral therapy program, but the latter is a known effective treatment for depression ([58], RCT). The third comprised a computer and Internet educational program for older people that incorporated, but was not limited to, participation in forums and virtual communities ([59], controlled trial). The remaining multi-component trial found no effect of a complex intervention incorporating an ISG component for rural-residing women with a chronic illness ([56], RCT). This study found that an intensive intervention involving peer-to-peer online support, expert-facilitated online group discussion, and online expert advice resulted in no greater reduction in depression than an information intervention alone or no intervention [56]. The 11 non-experimental studies identified investigated the relationship between chatroom (unspecified) use and depression, and most used a cross-sectional design. The findings were mixed. In summary, studies published since mid-2007 shed little additional light on the effectiveness of ISGs in reducing depressive symptoms and provide no further evidence concerning the efficacy of depression ISGs.

Conclusions

There is a need for high-quality research to investigate the effect of ISGs on depression outcomes. We acknowledge that there are significant challenges associated with designing and undertaking efficacy studies of ISGs. We acknowledge too that the appropriateness and feasibility of conducting such research on online self-help groups have been questioned [68]. However, we believe that creative researchers, together with consumers, can find a way to shed further light on an issue of unquestionable practical significance for millions of consumers worldwide.

Acknowledgments

KG is supported by NHMRC Fellowship 425413, AC is supported by NHMRC Capacity Grant No 418020, and MB is supported by an Australian Rotary Health Ian Scott PhD Scholarship. KG conceived the project, designed the study, rated the abstracts, undertook the statistical analyses, and drafted the paper. AC designed the search terms; identified, screened, and rated the abstracts; and edited the paper. MB screened abstracts and edited the paper. Ada Tam provided editorial assistance.

Conflicts of Interest

None declared.

References

1. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. BMJ 2004 May 15;328(7449):1166 [FREE Full text] [Medline: 15142921] [doi: 10.1136/bmj.328.7449.1166]
2. Horrigan JB, Rainie L, Fox S. Online communities: networks that nurture long-distance relationships and local ties. Washington, DC: Pew Internet & American Life Project; 2001.
3. Fox S. Online health search 2006. Washington, DC: Pew Internet & American Life Project; 2006.
4. Davison KP, Pennebaker JW, Dickerson SS. Who talks? The social psychology of illness support groups. Am Psychol 2000 Feb;55(2):205-217. [Medline: 10717968] [doi: 10.1037/0003-066X.55.2.205]
5. Benton T, Staab J, Evans DL. Medical co-morbidity in depressive disorders. Ann Clin Psychiatry 2007;19(4):289-303. [Medline: 18058286] [doi: 10.1080/10401230701653542]
6. Lloyd C. Understanding social support within the context of theory and research on the relationship of life stress and mental health. In: Brugha TS, editor. Social Support and Psychiatric Disorder. Cambridge, UK: Cambridge University Press; 1995:41-60.
7. Griffiths KM, Calear A, Banfield M, Tam A. Systematic Review on Internet Support Groups (ISGs) and Depression (2): What is Known About Depression ISGs? J Med Internet Res 2009;11(3):e41. [doi: 10.2196/jmir.1303]
8. Thesaurus.com. URL: http://thesaurus.reference.com/ [WebCite Cache ID 5jRrzpYJR]
9. Davidson KP, Pennebaker JW, Dickerson SS, Noferi G. Who speaks? The social psychology of self-help groups for disease. Bollettino di Psicologia Applicata 2001;234:3-19.
10. Fekete S, Osvath E. The Internet as a new source of information on suicides, depression and anxiety: preliminary study. Psychiatr Hung 1997;12(4):209-216. [doi: 10.1080/13811110214533]
11. Kordy H, Golkaramnay V, Wolf M, Haug S, Bauer S. Internet chat groups in psychotherapy and psychosomatics. Acceptance and effectiveness of an internet-bridge between hospital and everyday-life. Psychotherapeut 2006;51(2):144-153. [doi: 10.1007/s00278-005-0458-6]
12. Kordy H, Haug S, Wolf M, Dogs P. An internet connecting bridge between specialized institution and the daily routine. MMW Fortschr Med 2006 Aug 3;148(31-32):27-29. [Medline: 16937873]

13. Pfeiffer-Gerschel T, Niedermeier N, Hegerl U. Modern discussion forum "depression, suicidality". MMW Fortschr Med 2006 Aug 3;148(31-32):22-5, 27. [Medline: 16937872]
14. Neidermeier N, Pfeiffer-Gerschel T, Hegerl U. Learning from our patients - a field report after five years running an online discussion forum in the framework of the German Research Network on Depression and Suicidality. Nervenheilkunde 2006;25:361-367.
15. Herrmann-Traulsen C, Götz T. Study of virtual self help groups after prenatal loss: not being alone in the grief process. Pflege Z 2006 Jul;59(7):418-421. [Medline: 16883739]
16. Katayama K. The present situation regarding research into a certain electronic support group that aims to heal grief. Nippon Koshu Eisei Zasshi 2006 Jun;53(6):424-431. [Medline: 16881530]
17. Doring N. Rewards and punishments on the Internet: behavioral control in chat forums. Gruppendynamik 2001;32(2):109-143. [doi: 10.1007/s11612-001-0014-y]
18. Paquet MJ, Leprohon J, Cantin L. The virtual community in heart health practice. Perspect Infirm 2004;1(3):21-29. [Medline: 14964075]
19. Agrell H, Wålinder J. Internet health service can provide support to depressed individuals. Lakartidningen 2002;99:4152-4154, 4157.
20. Lieberman MA, Goldstein BA. Not all negative emotions are equal: the role of emotional expression in online support groups for women with breast cancer. Psychooncology 2006 Feb;15(2):160-168. [Medline: 15880627] [doi: 10.1002/pon.932]
21. Dew MA, Goycoolea JM, Harris RC, Lee A, Zomak R, Dunbar-Jacob J, et al. An internet-based intervention to improve psychosocial outcomes in heart transplant recipients and family caregivers: development and evaluation. J Heart Lung Transplant 2004 Jun;23(6):745-758. [Medline: 15366436] [doi: 10.1016/j.healun.2003.07.002]
22. Owen JE. Internet support and information for women with breast cancer [dissertation]. Birmingham, AL: University of Alabama; 2003.
23. van den Brink JL, Moorman PW, de Boer MF, Hop WCJ, Pruyn JFA, Verwoerd CDA, et al. Impact on quality of life of a telemedicine system supporting head and neck cancer patients: a controlled trial during the postoperative period at home. J Am Med Inform Assoc 2007;14(2):198-205 [FREE Full text] [Medline: 17213498] [doi: 10.1197/jamia.M2199]
24. Gustafson DH, Hawkins R, Boberg E, Pingree S, Serlin RE, Graziano F, et al. Impact of a patient-centered, computer-based health information/support system. Am J Prev Med 1999;16(1):1-9. [doi: 10.1016/S0749-3797(98)00108-1]
25. Tuil WS, Verhaak CM, Braat DDM, de Vries Robbé PF, Kremer JAM. Empowering patients undergoing in vitro fertilization by providing Internet access to medical data. Fertil Steril 2007 Aug;88(2):361-368. [Medline: 17416366] [doi: 10.1016/j.fertnstert.2006.11.197]
26. Bragadottir H. Developing a computer-mediated support group intervention for parents whose children have been diagnosed with cancer [dissertation]. Iowa City, IA: University of Iowa; 2004.
27. Gross EF. Logging on, bouncing back: An experimental investigation of online communication following social exclusion [dissertation]. Los Angeles, CA: University of California; 2006.
28. Lieberman MA, Winzelberg A, Golant M, Di Minno M. The impact of group composition on internet support groups: homogenous versus heterogeneous Parkinson's groups. Group Dyn 2005;9(4):239-250. [doi: 10.1037/1089-2699.9.4.239]
29. Winzelberg AJ, Classen C, Alpers GW, Roberts H, Koopman C, Adams RE, et al. Evaluation of an internet support group for women with primary breast cancer. Cancer 2003 Mar 1;97(5):1164-1173 [FREE Full text] [Medline: 12599221] [doi: 10.1002/cncr.11174]
30. Lieberman MA, Golant M, Giese-Davis J, Winzlenberg A, Benjamin H, Humphreys K, et al. Electronic support groups for breast carcinoma: a clinical trial of effectiveness. Cancer 2003 Feb 15;97(4):920-925 [FREE Full text] [Medline: 12569591] [doi: 10.1002/cncr.11145]
31. Lieberman MA, Goldstein BA. Self-help on-line: an outcome evaluation of breast cancer bulletin boards. J Health Psychol 2005 Nov;10(6):855-862. [Medline: 16176962] [doi: 10.1177/1359105305057319]
32. Rodgers S, Chen Q. Internet community group participation: psychosocial benefits for women with breast cancer. Journal of Computer-Mediated Communication [serial online] 2005;10(4) [FREE Full text] [WebCite Cache ID 5jGp9vu3P]
33. Andersson G, Bergström J, Holländare F, Carlbring P, Kaldo V, Ekselius L. Internet-based self-help for depression: randomised controlled trial. Br J Psychiatry 2005 Nov;187(5):456-461 [FREE Full text] [Medline: 16260822] [doi: 10.1192/bjp.187.5.456]
34. Houston TK, Cooper LA, Ford DE. Internet support groups for depression: a 1-year prospective cohort study. Am J Psychiatry 2002 Dec;159(12):2062-2068 [FREE Full text] [Medline: 12450957] [doi: 10.1176/appi.ajp.159.12.2062]
35. Golkaramnay V, Bauer S, Haug S, Wolf M, Kordy H. The exploration of the effectiveness of group therapy through an Internet chat as aftercare: a controlled naturalistic study. Psychother Psychosom 2007;76(4):219-225. [Medline: 17570960]
36. McKay HG, Glasgow RE, Feil EG, Boles SM, Barrera M J. Internet-based diabetes self-management and support: initial outcomes from the diabetes network project. Rehabil Psychol 2002;47(1):31-48. [doi: 10.1037/0090-5550.47.1.31]
37. Glasgow RE, Boles SM, McKay HG, Feil EG, Barrera M. The D-Net diabetes self-management program: long-term implementation, outcomes, and generalization results. Prev Med 2003 Apr;36(4):410-419. [Medline: 12649049] [doi: 10.1016/S0091-7435(02)00056-7]

38. Quick B. The role of support groups on the internet for those suffering from chronic kidney disease [dissertation]. Stockton, CA: University of the Pacific; 1999.
39. Shaw LH, Gant LM. In defense of the internet: the relationship between Internet communication and depression, loneliness, self-esteem, and perceived social support. Cyberpsychol Behav 2002 Apr;5(2):157-171 [FREE Full text] [Medline: 12025883] [doi: 10.1089/109493102753770552]
40. Morgan C, Cotten SR. The relationship between internet activities and depressive symptoms in a sample of college freshmen. Cyberpsychol Behav 2003 Apr;6(2):133-142. [Medline: 12804025] [doi: 10.1089/109493103321640329]
41. Sun P, Unger JB, Palmer PH, Gallaher P, Chou CP, Baezconde-Garbanati L, et al. Internet accessibility and usage among urban adolescents in Southern California: implications for web-based health research. Cyberpsychol Behav 2005 Oct;8(5):441-453. [Medline: 16232037] [doi: 10.1089/cpb.2005.8.441]
42. Campbell AJ, Cumming SR, Hughes I. Internet use by the socially fearful: addiction or therapy? Cyberpsychol Behav 2006 Feb;9(1):69-81. [Medline: 16497120] [doi: 10.1089/cpb.2006.9.69]
43. Kang S. Disembodiment in online social interaction: impact of online chat on social support and psychosocial well-being. Cyberpsychol Behav 2007 Jun;10(3):475-477. [Medline: 17594274] [doi: 10.1089/cpb.2006.9929]
44. Brennan PF, Moore SM, Smyth KA. The effects of a special computer network on caregivers of persons with Alzheimer's disease. Nurs Res 1995;44(3):166-172. [Medline: 7761293] [doi: 10.1097/00006199-199505000-00007]
45. Lieberman MA, Winzelberg A, Golant M, Wakahiro M, DiMinno M, Aminoff M, et al. Online support groups for Parkinson's patients: a pilot study of effectiveness. Soc Work Health Care 2005;42(2):23-38. [Medline: 16390834] [doi: 10.1300/J010v42n02_02]
46. Battles HB, Wiener LS. STARBRIGHT world: effects of an electronic network on the social environment of children with life-threatening illnesses. Child Health Care 2002;31(1):47-68. [doi: 10.1207/S15326888CHC3101_4]
47. Hill W, Weinert C, Cudney S. Influence of a computer intervention on the psychological status of chronically ill rural women: preliminary results. Nurs Res 2006;55(1):34-42 [FREE Full text] [Medline: 16439927] [doi: 10.1097/00006199-200601000-00005]
48. McKay HG, King D, Eakin EG, Seeley JR, Glasgow RE. The diabetes network internet-based physical activity intervention: a randomized pilot study. Diabetes Care 2001 Aug;24(8):1328-1334 [FREE Full text] [Medline: 11473065] [doi: 10.2337/diacare.24.8.1328]
49. Gustafson DH, Hawkins RP, Boberg EW, Bricker E, Pingree S, Chan CL. The use and impact of a computer-based support system for people living with AIDS and HIV infection. Proc Annu Symp Comput Appl Med Care 1994:604-608 [FREE Full text] [Medline: 7949999]
50. Taylor CB, Bryson S, Luce KH, Cunning D, Doyle AC, Abascal LB, et al. Prevention of eating disorders in at-risk college-age women. Arch Gen Psychiatry 2006 Aug;63(8):881-888 [FREE Full text] [Medline: 16894064] [doi: 10.1001/archpsyc.63.8.881]
51. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al; CONSORT GROUP (Consolidated Standards of Reporting Trials). The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001 Apr 17;134(8):663-694 [FREE Full text] [Medline: 11304107]
52. Fann JR, Thomas-Rich AM, Katon WJ, Cowley D, Pepping M, McGregor BA, et al. Major depression after breast cancer: a review of epidemiology and treatment. Gen Hosp Psychiatry 2008;30(2):112-126. [Medline: 18291293] [doi: 10.1016/j.genhosppsych.2007.10.008]
53. Bragadóttir H. Computer-mediated support group intervention for parents. J Nurs Scholarsh 2008;40(1):32-38. [Medline: 18302589] [doi: 10.1111/j.1547-5069.2007.00203.x]
54. Changrani J, Lieberman M, Golant M, Rios P, Damman J, Gany F. Online cancer support groups: experiences with underserved immigrant Latinas. Prim Psychiatry 2008;15(10):55-62.
55. Van Brunt RAZ. An online support group intervention for Asian American lesbian and bisexual women [dissertation]. Iowa City, IA: Iowa State University; 2008.
56. Weinert C, Cudney S, Hill WG. Rural women, technology, and self-management of chronic illness. Can J Nurs Res 2008 Sep;40(3):114-134 [FREE Full text] [Medline: 18947095]
57. Brattberg G. Internet-based rehabilitation for individuals with chronic pain and burnout II: a long-term follow-up. Int J Rehabil Res 2007 Sep;30(3):231-234. [Medline: 17762769] [doi: 10.1097/MRR.0b013e32829fa545]
58. Heinicke BE, Paxton SJ, McLean SA, Wertheim EH. Internet-delivered targeted group intervention for body dissatisfaction and disordered eating in adolescent girls: a randomized controlled trial. J Abnorm Child Psychol 2007 Jun;35(3):379-391. [Medline: 17243014] [doi: 10.1007/s10802-006-9097-9]
59. Shapira N, Barak A, Gal I. Promoting older adults' well-being through Internet training and use. Aging Ment Health 2007 Sep;11(5):477-484. [Medline: 17882585] [doi: 10.1080/13607860601086546]
60. Selfhout MHW, Branje SJT, Delsing M, ter Bogt TFM, Meeus WHJ. Different types of Internet use, depression, and social anxiety: the role of perceived friendship quality. J Adolesc 2009 Aug;32(4):819-833. [Medline: 19027940] [doi: 10.1016/j.adolescence.2008.10.011]

61. van den Eijnden RJJM, Meerkerk GJ, Vermulst AA, Spijkerman R, Engels RCME. Online communication, compulsive Internet use, and psychosocial well-being among adolescents: a longitudinal study. Dev Psychol 2008 May;44(3):655-665. [Medline: 18473634] [doi: 10.1037/0012-1649.44.3.655]
62. Beaudoin CE, Tao CC. Benefiting from social capital in online support groups: an empirical study of cancer patients. Cyberpsychol Behav 2007 Aug;10(4):587-590. [Medline: 17711369] [doi: 10.1089/cpb.2007.9986]
63. Beaudoin C, Tao CC. Modeling the impact of online cancer resources on supporters of cancer patients. New Media & Society 2008;10(2):321-344. [doi: 10.1177/1461444807086477]
64. Bessière K, Kiesler S, Kraut R, Boneva B. Effects of internet use and social resources on changes in depression. Information, Communication & Society 2008;11(1):47-70. [doi: 10.1080/13691180701858851]
65. Fortson BL, Scotti JR, Chen YC, Malone J, Del Ben KS. Internet use, abuse, and dependence among students at a southeastern regional university. J Am Coll Health 2007;56(2):137-144. [Medline: 17967759] [doi: 10.3200/JACH.56.2.137-146]
66. Mittal VA, Tessner KD, Walker EF. Elevated social Internet use and schizotypal personality disorder in adolescents. Schizophr Res 2007 Aug;94(1-3):50-57 [FREE Full text] [Medline: 17532188] [doi: 10.1016/j.schres.2007.04.009]
67. Slicton D. Lured into the Web: information superhighway or dead end? [dissertation]. Colorado Springs, CO: The University of the Rockies; 2006.
68. Barak A, Grohol JM, Pector E. Methodology, validity, and applicability: a critique on Eysenbach, et al. BMJ. 2004. URL: http://www.bmj.com/cgi/eletters/328/7449/1166#59643 [WebCite Cache ID 5jGpd7S5r]

Abbreviations

ISG: Internet support group
ITT: intent to treat
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 28.05.09; peer-reviewed by J Powell; comments to author 22.06.09; revised version received 15.07.09; accepted 31.07.09; published 30.09.09.

Please cite as:
Griffiths KM, Calear AL, Banfield M
Systematic Review on Internet Support Groups (ISGs) and Depression (1): Do ISGs Reduce Depressive Symptoms?
J Med Internet Res 2009;11(3):e40
URL: http://www.jmir.org/2009/3/e40/
doi:10.2196/jmir.1270
PMID:19793719

Original Paper

Systematic Review on Internet Support Groups (ISGs) and Depression (2): What Is Known About Depression ISGs?

Kathleen M Griffiths[1], PhD; Alison L Calear[1], PhD; Michelle Banfield[1], BA; Ada Tam[1], BSc

Centre for Mental Health Research, The Australian National University, Canberra, Australia

Corresponding Author:
Kathleen M Griffiths, PhD
Centre for Mental Health Research
The Australian National University
Professor & Deputy Director
Canberra ACT 0200
Australia
Phone: +61 2 6125 9723
Fax: +61 2 6125 0733
Email: kathy.griffiths@anu.edu.au

Abstract

Background: Internet support groups (ISGs) are a popular means by which consumers with depression communicate online. A number of studies have evaluated the nature and impact of depression-specific ISGs. However, to date there have been no published systematic reviews of this evidence.

Objective: The aim was to systematically identify and summarize the available evidence concerning the scope and findings of studies of depression ISGs.

Methods: Three databases (PubMed, PsycINFO, Cochrane) were searched using over 150 search terms extracted from relevant papers, abstracts, and a thesaurus. Papers were included if they employed an online peer-to-peer depression-specific support group and reported either quantitative or qualitative empirical data. The objective of each study was coded according to a 20-category classification system, which included the effect on depression and other outcomes, including help seeking; user characteristics, activity, satisfaction, perceived benefits, perceived disadvantages; the reason for using the ISG; the nature of ISG posts; characteristics of depression ISGs compared to other ISG types, face-to-face groups, and face-to-face counseling; ISG structure and longitudinal changes; and predictors of ISG adherence.

Results: Thirteen papers satisfied the inclusion criteria from an initial pool of 12,692 abstracts. Of these, three collected data using survey questionnaires, nine analyzed samples of posts, and one both collected survey data and analyzed a sample of posts. The quality of most studies was not high, and little data were collected on most key aspects of depression ISGs. The most common objective of the studies was to analyze the nature of the posts (eight studies) and to describe site usage (six studies) and user characteristics (five studies). The most prevalent types of social support were emotional, informational, and social companionship.

Conclusions: Given the popularity of depression ISGs and the paucity of available evidence about them, there is a need for high-quality, systematic studies of these groups, their impact, and the characteristics of their members and users. Such information is required to inform decision making by consumers, provider and educational organizations, guideline developers, policy makers, and funding bodies considering using, recommending, providing, or funding such groups.

(J Med Internet Res 2009;11(3):e41) doi:10.2196/jmir.1303

KEYWORDS

Depression; consumer participation; Internet; self-help groups

Introduction

Depression is a recurring, debilitating condition that is the primary cause of disability burden in developed countries [1]. Although frequently chronic in nature [2], depression is typically managed as an acute condition. In addition, depression and help seeking for the condition are stigmatized [3,4]. A substantial minority of people with depression do not seek formal help [5], and those who do may find that formal services do not meet all their needs [6], particularly in relation to practical advice and ongoing emotional support. Moreover, it has been calculated

that optimal treatment using the best currently available evidence-based interventions would avert only 34% of the burden associated with depression even if it were provided to all people with the condition [7].

It is perhaps not surprising then that consumers seek less-formal methods to assist them in coping with depression. Peer-to-peer depression Internet support groups (ISGs) provide one potential means of obtaining such support. In fact, there is evidence that such groups are among the most common support groups on the Internet [8]. Many ISGs enjoy a large membership. For example, at the time of writing, one depression group reported a registered membership of over 30,000 members with as many as 3993 visiting on one day [9].

Given the prevalence of depression ISGs and their potential to play a role in the management of depressive disorders, it is important to understand who, why, and in what way consumers use depression ISGs, their benefits and risks, and how such groups are best structured for optimal consumer outcomes. In a companion paper, we have reported the results of a review of the effect of health ISGs on depressive symptoms [10]. However, the review was not focused on depression ISGs specifically and incorporated only those studies reporting depression outcome data. As noted above, other attributes of depression ISGs are of interest. To our knowledge, there are no published reviews of the scope or findings of empirical research on depression-specific ISGs.

The current study seeks to address this gap by reporting the results of a systematic review of the available quantitative and qualitative evidence concerning depression ISGs. In particular, the review aims to document what is known about the demographic and clinical characteristics of depression ISG users, the nature and quantity of depression ISG usage, consumer attitudes about depression ISGs, whether depression ISGs influence help seeking and user attitudes about conventional health care, and how these online groups compare with other types of ISGs and with face-to-face mutual support groups for depression.

Methods

The methods employed in the current review have been reported in a companion study of the efficacy of health ISGs in reducing depression symptoms, [10] and the reader is referred to that paper for further details. Briefly, the review methodology entailed a search of three databases for the period prior to August 2007 using an extended version of a search strategy reported by Eysenbach et al [11]. The procedure for identifying studies involved a multi-step process: (1) eliminating clearly irrelevant abstracts, (2) identifying definitely or possibly relevant abstracts (two reviewers), (3) collecting and eliminating papers not satisfying inclusion criteria (two reviewers), and (4) identifying any additional relevant studies cited in systematic reviews (two reviewers) or included studies. Inclusion criteria for the current study were that the study reported qualitative or quantitative empirical data on a peer-to-peer depression support group. Studies were excluded if the target ISG was not specific to depression [12-14]. Of the 12,692 abstracts initially returned by the database searches, 13 studies satisfied the inclusion criteria, including two efficacy studies [15,16] that were also included in the efficacy review [10]. A flow diagram of the above process is available in the companion paper to this review [10]. As noted by Griffiths et al [10], it is possible that additional papers may have been published since the original searches. To investigate this, a search was conducted by the first author incorporating the period of 2007 to May 2009 and using the same search terms employed in the reported searches but limiting results to those incorporating the terms “depression” or “depressive” or “mood.” No new empirical studies of depression ISGs were identified.

Coding of the Included Studies

Each of the identified 13 papers was independently coded by two raters (KG, AT) and discrepancies subsequently resolved by discussion. Variables coded included type of ISG (format, moderation characteristics, country of origin), country of origin of the primary author, study design (experimental, observational, descriptive), and the type of data analyzed (survey, ISG posts). In addition, each study was rated for the presence (yes, no) of each of a series of potential aims: to measure the (1) effectiveness of ISGs on depression or (2) another outcome (eg, quality of life); (3) effect on formal service use; (4) user satisfaction with depression ISGs; (5) reason for using depression ISGs; (6) user perceived benefits; (7) user perceived disadvantages; (8) source of referral to the ISG; (9) demographic characteristics; (10) clinical characteristics; (11) service use of depression ISG users; (12) activity/usage of depression ISGs; (13) the nature of depression ISG posts; (14) comparative characteristics of depression and other medical ISGs, (15) face-to-face support groups, and (16) face-to-face counseling; (17) changes in depression ISGs over time; (18) predictors of adherence to depression ISGs; (19) aspects liked best and/or least about ISG; and (20) ISG structure. In practice, it proved difficult to differentiate between categories (5) and (6), and these were therefore combined into a single category, “reason for using depression ISGs and/or user perceived benefits.”

Results

Table 1 summarizes the characteristics of each of the 13 identified studies. The majority were descriptive (n = 11), one was an observational trial, and one was an experimental trial. The quality of most studies was not high, and little data had been collected on most key aspects of depression ISGs. The most common objective of the studies was to analyze the nature of the posts (eight studies) and to describe site usage (six studies) and user characteristics (five studies). Two of the studies analyzed different aspects of the same data [17,18]. Findings from the 13 studies of depression ISGs are detailed below.

Table 1. Design characteristics of depression ISG studies: ISG user survey studies, ISG post studies, and combined user and post studies[a,b]

Study	Setting/Type of ISG	Design	Recruitment (surveys)/ Sampling method (posts)	Aims (for which findings reported)	Measures
ISG user survey studies					
Houston 2002 [15] USA	Five moderated public depression bulletin boards and listservs	Observational (survey) Descriptive (survey)	Participants recruited for survey over 2-mth period through posts on the five ISGs	• Effectiveness depression • Effectiveness other outcome • Characteristics • Perceived benefits/ disadvantages • Activity/usage • Compare with face-to-face counseling	• CES-D (depression) • MOS-SS (social support) • Age, gender, marital status, education, employment, nationality, years since first diagnosis, type/quantity/quality face-to-face care • Value of ISG (5-point Likert), effect on face-to-face care • ISG use past 2 wks (hrs), self-report • Preference for type of interaction (face-to-face vs ISG)
Powell 2003 [19] UK	Six Netdoktor depression bulletin boards (Austria, Denmark, Germany, Norway, Sweden, UK) Moderation status N/S	Descriptive (survey)	Visitors to Netdoktor ISGs over 4-wk period in May/June 2002 responding to survey offered in pop-up window	• Characteristics • Perceived benefits/disadvantages	• Age, sex, history of depression/consultations, MDI, nationality of ISG, reason for visit (self, friend, family member, etc) • Self-perceived effects of use
Andersson 2005 [16] Sweden	Moderated, research depression bulletin board	Experimental Control arm of a randomized controlled trial (survey)	Recruitment of participants for an Internet research project on depression through print media Self-selected participants included if probability of 0.55 of diagnosis of MDD (CIDI-SF) and clinically significant symptoms of depression (MADRS-S)	• Efficacy depression • Efficacy other outcome • Predictor of adherence	• BDI, MADRS-S (depression) • QOLi (quality of life), BAI (anxiety)
ISG post studies					
Davison 2000 [8] USA	Highest volume English-language Internet depression newsgroup and AOL depression support bulletin board	Descriptive (posts)	All posts to ISG analyzed over 2-wk period	• Activity/usage • Compare ISGs	• Number of posts
Salem 1997 [20] USA	Public depression newsgroup	Descriptive (posts, 2 independent coders)	Analyzed all posts to ISG in each of 2 wks, 1 mth apart in 1995 (involving 533 posters)	• Characteristics • Activity/usage • Nature of posts • Compare with face-to-face group	• Gender, work status, whether depression professionals (inferred from posts); whether depressed or a carer (inferred from posts) • Frequency of posts, no. of posters, mean and range posts/person, target of post (group vs individual) • 13 coding categories (5 general categories) derived from Roberts [21] and face-to-face mutual support literature • Measures used for characteristics and nature of posts

Study	Setting/Type of ISG	Design	Recruitment (surveys)/ Sampling method (posts)	Aims (for which findings reported)	Measures
Fekete 2002 [22] Hungary	Depression newsgroup	Descriptive (posts)	Analyzed all posts to ISG over 3-mth period (involving n = 45 posters)	• Characteristics • Activity/usage • Nature • Compare ISGs	• Gender (inferred from posts), nationality (inferred) • Frequency of posts, no. posters, no posting once only • Syntax/grammar analysis/speech patterns/verbal features (modified Weingraub [23,24] content analytic method)
Muncer 2000 [17] UK	One depression newsgroup, selected because it was "particularly active"	Descriptive (posts)	Random sample of all postings to ISG made over 1 mth (involved n = 118 participants)	• Usage • Nature	• Mean thread length, no posting > 6 messages • Cohen and Wills [25] typology of social support (coded at level of thread)
Muncer 2000 [18] UK	Depression newsgroup from Muncer et al [17] Diabetes newsgroup	Descriptive (posts)	Analyzed posts made over 1 mth to depression ISG (sampled as above) and posts made over "longer period" to diabetes ISG Usage analyses based on all posters (n = 118 depression, n = 132 diabetes); network structure on frequent posters (n = 26 depression, n = 10 diabetes)	• Activity/usage • Network structure • Compare ISGs	• Mean thread length, posters/thread, posts/person, no. posting once only, no. posting > 6 messages • Multidimensional scaling routine form UCINET [26]
Alexander 2002 [27] USA	Three public depression ISGs selected randomly from a list of non-professionally run depression groups on public e-groups website Moderation status N/S	Descriptive (posts, partial double coding to establish and verify reliability)	Analyzed 1 mth of posts to the three ISGs at 3, 6, 9, and 12 mths after group commencement Posts from 1998 to 2001	• Activity/usage • Nature (for leaders/most active participants)	• Frequency of posts/mth, no. posters, no. leaders • Modified Cutrona's Support Behaviour Codes [28]
Witt 1999 [29] USA	Public depression bulletin board, public smoking cessation bulletin board Method of selection, moderation status N/S	Descriptive	Analyzed over 1000 posts on ISGs collected over a "2- to 3-mth period"	• Nature of posts • Comparison of ISG types	• Contract, Thinking Feeling (CTF) coding system [30]
Macius 2005 [31] UK	Public "message board"	Descriptive (posts, 3 coders, reliabilities computed)	Analyzed one or two threads on ISG for a week in June 2002	• Nature of posts • Comparison of ISG types	• Coded for type of support, type of medical/drug content discussed
Lamerichs 2003 [32] Netherlands	Public depression "discussion forum" (senior citizens)	Descriptive (posts)	Selected extracts from ISG, chosen to illustrate limitations of existing cognitive models of online interactions	• Nature of posts	• Identity management
Survey and posts					

Study	Setting/Type of ISG	Design	Recruitment (surveys)/ Sampling method (posts)	Aims (for which findings reported)	Measures
Alexander 2003 [33] USA	Newsgroup Method of selection and moderation status N/S	Descriptive (posts, 3 independent coders)	Posts: Analyzed all posts made to the ISG over 3 consecutive wks (n = 74 members posting) Survey: Volunteers recruited through post on the ISG (n = 19)	• Characteristics • Usage • Nature	• Depression status (disclosed in posts) • Frequency of posts, no. of posters • Cutrona's Support Behaviour Codes [28] • 6 satisfaction items (survey)

[a] AOL = America Online; BAI = Beck Anxiety Inventory; BDI = Beck Depression Inventory; CES-D = Center for Epidemiologic Studies Depression Scale; CIDI-SF = Composite International Diagnostic Interview Short-Form; MADRS-S = Montgomery-Asberg Depression Rating Scale; MDI = Major Depression Inventory; MDD = Major Depressive Disorder; MOS-SS = Medical Outcomes Study Social Support Scale; N/S = not specified; QOLi = quality of life.

[b] Only measures for which data or findings were reported are included in table.

What Are the Characteristics of Depression ISG Users?

There is limited research on the clinical and demographic characteristics of the users of public depression ISGs. Moreover, data collected thus far have been derived only from surveys or advertisements posted on bulletin boards or inferred from bulletin board posts. No study has reported data on characteristics of all users who register with an ISG.

Clinical Status

Four studies reported information about users' current or past history of depression [15,19,20,33]. Assessments employed included formal measures and self-reported status as well as mental health status inferred from the content of ISG posts. The evidence suggests that the majority of depression ISG users are consumers with depression or a history of depression and that a substantial percentage (50-80%) are currently depressed on formal assessment. In particular, Houston et al [15] found that 99% of respondents to a survey of depression ISG members reported having received a diagnosis of depression from a health professional, and 86% scored above the cutoff for depression on the CES-D screening test for depression at the time of completing the survey 1 to 2 months after joining the ISG. Powell et al [19] reported that 52% of ISG users had current major depression as measured by the Major Depression Inventory and that 7% of respondents were friends or family members of a person with depression. They argued that the discrepancy between their findings and those reported by Houston et al [15] may be due to differences in recruitment methods employed in the two studies. The remaining two studies reported the rates of self-identification or inferred depression status from an analysis of posts [20,33]. The first found that all members posting on a depression ISG self-identified as suffering from depression, [33] and the second, inferring the clinical status of active users of a newsgroup from their posts, reported that 92% identified themselves as depressed and 2% as carers of people with depression [20].

Clinical Treatments/Service Use

Two studies have reported data on depression ISG users' receipt of professional treatments or services [15,19]. Houston et al [15] reported a high level of such help, with 92% of ISG respondents to their survey currently receiving antidepressants and 65% receiving counseling. ISG users surveyed by Powell et al [19] were less likely to have received professional help. Reporting only on the 50% of depression ISG participants with current major depression, the authors found that 26% were currently receiving psychological treatment, 44% medication for the condition, and 51% either psychological treatment, medication, or both. Although 64% had consulted a health professional in the previous year, one-third (34%) had never received either treatment. Again, it is unclear if the discrepancy in the results of these studies relates to differences in the study recruitment methods or to differential treatment rates in the jurisdictions served by the ISGs [19].

Gender

Four studies analyzed the gender distribution of users [15,19,20,22]. The results of these studies were mixed. Two reported a preponderance of females (79% and 70%) [15,19], and two reported more male users (61% and 66%) [20,22]. The first two studies focused on bulletin boards, and gender was based on self-report among survey respondents; the second two studies involved newsgroups, and gender was inferred from posts.

Age

Two studies reported on the age distribution of respondents. These studies suggest that bulletin board users are commonly between their mid- to late 20s and mid-40s. Houston et al [15] reported that 21%, 49%, and 30% of bulletin board participants were aged 18-29 years, 30-45 years, and over 45 years, respectively. Powell et al [19] reported that 27%, 33%, 23%, and 17% of participants were aged < 26 years, 26-35 years, 36-45 years, and over 45 years, respectively.

Other Demographics

Only one study evaluated other demographic characteristics of depression ISG users. Houston et al [15] reported that 45% of users had achieved a college education, 42% were unemployed, and 44% were married.

Source of Referral

Although one survey included a question about source of referral to the depression ISG, the authors did not report the findings for this item [33].

What Is Known About Depression ISG Usage?

Six studies (seven papers) provided some information about usage of public depression ISGs [8,15,17,18,20,22,27].

Posting Rate

Three studies reported the rate of posting on an ISG, with markedly differing results. For comparative purposes, we converted these rates to posts per user per week. Reported rates of posting varied from 0.3 [22] to 25.4 [27] per user per week. More specifically, in studies of depression newsgroup posts, Salem et al [20] reported 1.8 posts per user per week (533 users over 2 weeks) and Fekete [22], 0.29 posts per user per week (45 users over a 3-month period, 29 [64%] of these posting only once). Alexander [27] found that the number of users and rate of posting on a depression ISG varied across time and group. Twelve months after commencement of the group, usage on three different depression ISGs was 5.6 (20 participants), 25.4 (21 participants), and 1.2 (69 participants) posts per user per week. Usage of the ISGs also changed over time (see below). Davison [8] reported that the highest-volume depression ISGs on America Online and the wider Internet yielded total weekly posts of 124 and 389, respectively. However, the author failed to report the number of users generating these posts.

Thread Length

Little is known about average thread length on depression ISGs. One group reported an average thread length of 8.04 posts across 61 threads randomly selected from an unspecified number of threads posted during a 1-month period on a "particularly active newsgroup," with 60% of messages posted by 26 (22%) of the participants [17].

Time Spent on ISGs

Only one study investigated the time members devote to depression ISGs. Over half (53.4%) of the ISG users in the study reported spending at least 5 hours on the depression ISG over a 2-week period in the early stages of their membership in the group [15].

What Is the Nature of Posts on Depression ISGs?

Eight studies (nine papers) provided some information about the nature of depression ISG posts [17,18,20,22,27,29,31-33]. Four of these reported quantitative information about the prevalence of different types of social support in posts on a total of seven newsgroups [17,20,27,33]. Some of the latter studies also coded for other characteristics, including the prevalence of disclosure, help seeking, different types of knowledge [20], and task maintenance [17,33].

Each employed a pre-formulated system for coding the typology of posts. Coding systems included the Cutrona Support Behavior Code [28] (studies [27,33]), an adaptation of a typology developed by Roberts and collaborators for face-to-face support groups [21] (studies [20,21]), and a typology of social support described by Cohen and Wills [25] (studies [17,27]). Each of these systems differs subtly, rendering synthesis of study findings difficult. For example, cognitive guidance messages were coded as informational in some studies [17,27,33] but separately in another [20]. Esteem support and emotional support were treated separately in two studies [27,33] but coded only in the category "esteem support" in another [17]. Studies also differed with respect to the unit of analysis (thread vs post) and whether the unit could be allocated a single code or multiple codes. One study confined the analysis to messages that the authors deemed "helping posts" [20], and another limited the analysis to a very small number of members designated as group "leaders" [27]. The characteristics of the depression ISGs also differed across and within studies.

Social Support

The relative prevalence of different types of social support varied across studies. However, overall emotional support was common (mean 34%, range 22-48%, n = 5 groups). Informational support also accounted for a substantial proportion of posts (mean 26.2, range 8-46%, n = 6 groups). In the single study that investigated it [20], cognitive guidance accounted for only a minority of posts (7.2%). Posts sharing experiential knowledge (14%) were more common than posts containing "second-hand" professional knowledge (3%) [20]. The only study that reported prevalence of social companionship (eg, social chit-chat) found that it accounted for almost one quarter of posts [17]. Considering only the two studies that coded pure self-esteem, one reported a substantial minority of self-esteem posts (24%, [33]), whereas the other, which was confined to posts of ISG leaders, reported few self-esteem posts (5%, [27]). Unlike other forms of social support, tangible support was absent in ISG posts [17,33].

Other Characteristics of Posts

Salem et al [20] reported that self-disclosure was common (50.6%) but that only a minority of posts involved requests for help (13%). Posts relating to the group structure and group identification were also common (20%). However, in his study of three depression ISG groups, Alexander [27] identified relatively low levels of task maintenance (eg, monitoring group norms, keeping discussion on track; mean 5%, range 2-8%, n = 3 groups) and relational maintenance (reinforcement of the cohesiveness of the group; mean 9.3%, range 8-11%, n = 3).

In a syntactical and grammatical analysis of ISG posts, another study reported that personal "I/me" references, feelings, expressions involving judgments of goodness/badness, adverbial intensifiers, dichotomous expressions, retractions and explanations were more frequently employed on a depression ISG than a control group journalism discussion group [22].

What Predicts the Nature of Posts?

Salem et al [20] found that high-frequency users were significantly more likely to post socially supportive and humorous messages, to agree with others, and to respond to individuals as opposed to the group. However, they disclosed less, sought less help, and offered less experiential and second-hand knowledge than less-frequent ISG users. Salem et al [20] also investigated difference in posts according to the inferred gender of the user. A larger percentage of men's posts were experiential, whereas women's posts were more likely to involve group structure and process.

Are Users Satisfied With Depression ISGs?

Only one study has formally measured level of satisfaction with a depression ISG [33]. Based on a survey posted on a depression ISG, the authors reported that members were extremely satisfied with the group (score 6.47 out of 7). The response rate represented 26% of 74 members posting during the period the survey was available.

Why Do Members Use ISGs?: Perceived Benefits and Disadvantages

Three studies systematically investigated the nature of benefits and/or the reason for participating in a depression ISG [15,19,33]. They provide some evidence that emotional support and information are perceived by members as an advantage of depression ISGs and that these groups are perceived to be effective in reducing depressive symptoms. There is little evidence concerning the disadvantages of ISGs.

In particular, Houston et al [15] reported that emotional support was the most commonly cited reason for participating in a depression ISG. Of the 103 members who participated in their survey, 98 (95%) reported that the ISG had helped their symptoms. Powell et al [19] found that of those members who visited a depression ISG more than once and completed an online survey, the majority (71%) reported having learned "more about medication," half indicated that they were "able to discuss subjects that they felt unable to discuss elsewhere," and 44% indicated that they "felt less isolated" due to their participation in the ISG [19]. There were also positive effects on formal help seeking (see below). Finally, Alexander et al [33] reported that user-cited benefits of a depression ISG were freely given support, caring, and affirmation from other members, the provision of an outlet for expression, and a place to turn when alone. Although the authors requested that users also indicate the aspects of the depression ISG they liked least, users typically replied that the groups did not require any improvement.

Do Depression ISGs Improve Outcomes?

Two studies have investigated the effectiveness of depression ISGs for improving outcomes [15,16], in particular for depression [15,16], anxiety [16], quality of life [15], and social support [15]. As noted by Griffiths et al [10], the studies employed different designs, recruitment procedures, and ISG types and yielded different findings for depression, with one reporting a decrease in depressive symptoms among frequent public ISG users relative to low-frequency users and the other showing no significant reduction in depressive symptoms following participation in a research ISG. The only study to examine the effect of an ISG on anxiety symptoms yielded no change following participation in a research ISG [16]. The latter study also failed to find a change in quality of life following participation in the research ISG. There was no effect of participation on social support [15].

There were no observational or experimental studies of the effect of depression ISG participation on other outcomes such as knowledge of depression, attitudes to depression, self-efficacy, self-esteem, behaviors, or user empowerment.

Does Participation in Depression ISGs Affect Formal Help Seeking?

Little is known about the effect of ISGs on the use of other health services. However, Powell et al [19] found that over one-third (37%) of participants in their survey reported that participation had encouraged them to seek professional help, although a small minority considered that their participation had delayed such a consultation (9%) or reduced their trust in their doctor (11%) [19].

What Are the Similarities and Differences Between Face-to-Face Support Groups and ISGs?

No study has directly compared face-to-face and online support groups in a single trial. However, one group [20] did compare their findings from a depression ISG with those from a previously reported study of a face-to-face group [21].

Type of Interaction

Basing their coding system on that used in the interactions in the face-to-face support group, the researchers concluded that there were similarities and differences between the nature of ISG and face-to-face interactions. In each, "positive, supportive" statements were more frequent than negative comments. However, self-disclosure, which was rarely seen in the face-to-face group, was the most common type of communication in the ISG group. The authors [20] asserted that this difference was unlikely to be due entirely to differences in anonymity in the two formats since many users employed their own identities on the ISG. Rather, they proposed that the absence of physical cues minimizes perceived differences between members and allows participants to focus on the communication and their shared concerns. A second difference between ISG and face-to-face groups was that whereas for ISGs advice and information and emotional support were more common than cognitive guidance (of the type used in conventional therapy), the reverse was true for face-to-face interactions.

Demographic and Other Characteristics

The authors of the above study [20] also concluded that the gender composition of ISGs and face-to-face groups differed. They noted that whereas more women than men use face-to-face mutual support groups, 61% of ISG members in their study were male. However, as already noted, they inferred rather than directly collected gender information in their ISG study. It is possible that men were more likely to disclose their gender online.

Do ISGs for Depression Differ From ISGs for Other Conditions?

Seven studies have provided some comparative information about ISGs for depression and other conditions [8,18,22,27,29,31,33]. These have provided data on the differences in post frequency (n = 1), post content (n = 4), ISG structure (n = 1), and satisfaction with the ISG (n = 1).

Activity

Davison [8] investigated the extent of participation (number of posts over a 2-week period) as a function of different types of

ISG. Depression ISGs were the third and fourth most active ISGs on America Online and the general Internet, respectively.

Content

Some differences have also emerged with respect to the typology of the posts for depression and other online groups.

Alexander et al [33] compared the frequency of types of social support on a depression, Alcoholics Anonymous (AA), attention deficit disorder (ADD), and cancer ISG. Whereas emotional support was the most common type of support provided on the depression ISG, informational support was the most common support on each of the other three ISGs. The extent to which these differences were due to the nature of the condition supported by the ISG as opposed to the smaller membership and greater homogeneity of the depression group is unclear. In another study, Alexander et al [27] compared the type of social support seen on ISGs for eight different conditions, one of which was depression. Three different ISGs were analyzed for each condition. However, it was difficult to draw conclusions about the comparative frequency of different types of support across conditions due to the variability in the results between groups within the same condition. In addition, since only the messages of "leaders" were coded and there were few of these in a number of instances, the generalizability of the results to all members is unclear. In another study of the content of discussions, Macias et al [31] compared the nature of posts on ISGs for depression, panic/anxiety, breast cancer, prostate cancer, infertility, HIV/AIDS, arthritis, type II diabetes, high cholesterol, irritable bowel syndrome, and obesity/overweight groups. They reported that the depression ISG participants were the most likely to seek and provide advice but were among the least likely of the groups to discuss medical treatments and procedures. The depression group did not differ from the other ISGs with respect to seeking encouragement, expressing concern, or providing personal information not related to the illness.

Fekete [22] analyzed the syntactical and grammatical characteristics of a depression, suicide, and panic ISG and a journalism discussion group. Compared to the other three groups, the depression participants were significantly more likely to make value judgments (goodness-badness, right-wrong). They were also significantly more likely than the journalism group to use dichotomous (polarized) expressions such as "always" or "never," to provide a reason for their thoughts, beliefs, or actions ("because," "therefore"), to use words retracting another statement (eg, "but," "on the other hand," "except"), and to refer to themselves ("I," "me"). The depression group was more likely to use adverbial intensifiers ("I *really* like it") and to express feelings (eg, "I love," "I hate") than the journalism or panic groups. The suicide group used significantly more negative ("no," "never") and dichotomized expressions than did the depression group. In another study that analyzed the terms in a post, Witt [29] compared the content of a depression and smoking ISG using a computerized rating system of four bipolar dimensions: emotion, cognition, contract, and performance (see [30]). She concluded that compared to the smoking group, the depression group's language incorporated more negative affect (emotion negative) and that the depressive group was more active in asking for help (negative cognition) and offering help and information (positive cognition).

Network Analysis

In a comparison of the structure of a depression and a diabetes ISG using network analysis, Muncer et al [18] reported that the depression ISG networks were "denser and more vibrant," more likely to involve "cliques," and more likely to be characterized by social support than the diabetes group, in which participants were more likely to exchange information.

Satisfaction

Alexander et al ([33], see above) found that participants in the depression ISG reported a higher level of overall satisfaction than did the members of the AA or ADD ISGs. The depression group also reported higher satisfaction with esteem support than did the AA, ADD, or cancer group members.

How Do Depression ISGs Change Over Time?

Only one study has investigated change associated with ISGs over time [27]. The trend in the pattern of posts over time varied across depression groups, with one steadily decreasing, one increasing, and the third decreasing initially and then leveling out. Similarly, there were no consistent patterns of change across three depression groups in type of social support over a 1-year period.

What Predicts Adherence to ISGs?

Andersson et al [16] reported a withdrawal rate of 18% over a period of 10 weeks among participants in a stand-alone depression ISG created as part of a research protocol, a figure that was lower than the dropout recorded for an online cognitive behavioral therapy program over the same period. There have been no other reports of adherence rates for depression ISGs and no studies of the predictors of ISG adherence.

Preferences for ISG Support Compared to Face-to-Face Counseling?

Of those depression ISGs users responding to a survey, 38% indicated that they preferred ISG support to face-to-face counseling, approximately half (51%) preferred counseling, and the remainder expressed no preference [15].

Discussion

Findings

The systematic search yielded only 13 relevant studies for the period under study. These studies addressed a range of issues, including efficacy, user characteristics, activity levels, the nature of online interactions, satisfaction, and reasons for visiting depression ISGs. They also compared ISG activity and interactions over time and across conditions. However, the available data on each of these facets of depression ISGs were limited either due to the small number of studies undertaken, to methodological limitations, or both.

For example, little is known about the demographic characteristics of users, and although there have been four studies of gender distribution, results have been mixed and the methodology employed inadequate. Data that have been

collected on clinical characteristics of users suggest that a majority of ISG users have a history of depression and that a substantial percentage of those with current depression were receiving concurrent conventional depression treatments. The latter finding suggests that ISGs often serve as an adjunct to, rather than a replacement for, formal help seeking. However, one study did find that one-third of members with a current depressive disorder had never received formal treatment [19]. Together with evidence that many members report that a depression ISG facilitated their help seeking [19], this finding raises the possibility that depression ISGs may be a fertile setting in which to encourage formal help seeking. There are no available data concerning the means by which users are referred to or arrive at ISGs.

Although seven papers provided information about usage of depression ISGs, overall posting ratings varied across studies and there were insufficient studies to draw conclusions about thread length and time spent on depression ISGs. The strongest focus for depression ISG research has been on the nature of posts (eight studies, of which four investigated types of social support). The primary types of social support provided on depression ISGs were emotional, informational, and social companionship. In the one study that coded for it, self-disclosure was high. The effect of self-disclosure, including its potential advantages and disadvantages, has not been explored in the depression ISG domain, although there is evidence that high-frequency users are less likely to disclose [20] and that disclosure is higher on ISGs than in face-to-face support groups. There would also appear to be differences across conditions in the typology of posts and structure of ISGs. However, it is difficult to determine the extent to which variations across conditions in ISG interactions and structure are attributable to differences in the conditions or other factors (eg, size of the group, duration of group). One study reported variations across ISG depression groups in activity and type of interactions over time, with no consistent pattern evident across the groups.

The lack of efficacy studies of depression ISGs has already been discussed in the companion paper to this study [10]. There was a similar gap in evidence concerning the impact of depression ISGs on other well-being and mental health outcomes. Moreover, there were no observational or experimental studies of the effect of these support groups on user knowledge, user behaviors, or user attitudes.

Nevertheless, there is some evidence that users believe depression ISGs are useful. The single study to formally investigate user satisfaction with depression ISGs reported satisfaction to be high both in absolute terms and relative to ISGs for other health conditions. However, member response rate in this study was low, so it is unclear if the results are representative of all members. There is some evidence that depression ISG members perceive emotional support [15,33], information [19], and effectiveness in improving mood [15] as advantages of their depression ISG, but again these results may not be representative of all users. Significantly, there is little evidence concerning the disadvantages of ISGs despite the practical importance of this issue.

It is notable that none of the studies addressed the question of what factors promote greater acceptability of and satisfaction with depression ISGs, retention of members, activity levels, or efficacy of depression ISGs. No studies systematically manipulated variables such as group size, presence or absence of a moderator, level of moderator participation, content of the board rules, or level of ISG accessibility to evaluate the effect of these on process or outcome variables. Moreover, there were no comprehensive user reports addressing these issues or naturalistic comparative studies of groups differing in these attributes. It is notable that despite the potential relevance of depression ISGs to specific target groups such as adolescents, who are high users of social networking technologies, and older people and rural residents, whose access to other forms of peer-to-peer mutual support may be limited, there are no published studies of the use of ISGs by these populations. Without such studies, organizations or individuals planning or providing depression ISG services must do so without the benefit of an evidence-based framework.

Conclusions

The conclusions from this study are clear. There is a need for high-quality research, both quantitative and qualitative, to investigate all aspects of ISGs as they relate to depression. Currently, the evidence is not of sufficient quality or strength to inform decision making by consumers, ISG providers, health professionals, policy makers, or funding bodies considering the use of ISGs for depression. Given the popularity of peer-to-peer depression ISGs, appropriately targeted studies of these mutual support groups has the potential to contribute substantially to the identification, design, and implementation of suitable self-care models for depression.

Acknowledgments

KG is supported by NHMRC Fellowship 425413, AC is supported by NHMRC Capacity Grant No 418020, and MB is supported by an Australian Rotary Health Ian Scott PhD Scholarship. KG conceived the project, designed the study, rated the abstracts, undertook the statistical analyses, and drafted the paper. AC designed the search terms; identified, screened, and rated the abstracts; and edited the paper. MB screened abstracts and edited the paper. AT coded abstracts and edited the paper.

Conflicts of Interest

None declared.

References

1. Andrews G, Titov N. Depression is very disabling. Lancet 2007 Sep 8;370(9590):808-809. [Medline: 17826154] [doi: 10.1016/S0140-6736(07)61393-2]
2. Andrews G. Should depression be managed as a chronic disease? BMJ 2001 Feb 17;322(7283):419-421 [FREE Full text] [Medline: 11179166] [doi: 10.1136/bmj.322.7283.419]
3. Barney LJ, Griffiths KM, Jorm AF, Christensen H. Stigma about depression and its impact on help-seeking intentions. Aust N Z J Psychiatry 2006 Jan;40(1):51-54. [Medline: 16403038] [doi: 10.1111/j.1440-1614.2006.01741.x]
4. Griffiths KM, Nakane Y, Christensen H, Yoshioka K, Jorm AF, Nakane H. Stigma in response to mental disorders: a comparison of Australia and Japan. BMC Psychiatry 2006;6(1):21 [FREE Full text] [Medline: 16716231] [doi: 10.1186/1471-244X-6-21]
5. Andrews G, Issakidis C, Carter G. Shortfall in mental health service utilisation. Br J Psychiatry 2001 Nov;179(5):417-425 [FREE Full text] [Medline: 11689399] [doi: 10.1192/bjp.179.5.417]
6. McNair BG, Highet NJ, Hickie IB, Davenport TA. Exploring the perspectives of people whose lives have been affected by depression. Med J Aust 2002 May 20;176 Suppl:S69-S76 [FREE Full text] [Medline: 12065001]
7. Andrews G, Issakidis C, Sanderson K, Corry J, Lapsley H. Utilising survey data to inform public policy: comparison of the cost-effectiveness of treatment of ten mental disorders. Br J Psychiatry 2004 Jun;184(6):526-533 [FREE Full text] [Medline: 15172947] [doi: 10.1192/bjp.184.6.526]
8. Davison KP, Pennebaker JW, Dickerson SS. Who talks? The social psychology of illness support groups. Am Psychol 2000 Feb;55(2):205-217. [Medline: 10717968] [doi: 10.1037/0003-066X.55.2.205]
9. Depression Forums.org. URL: http://www.depressionforums.org/ [accessed 2009-07-15] [WebCite Cache ID 5iE3wIs93]
10. Griffiths KM, Calear A, Benfield M. Systematic Review on Internet support groups (ISGs) and depression (1): Do ISGs reduce depressive symptoms? J Med Internet Res 2009;11(3):e40. [doi: 10.2196/jmir.1270]
11. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. BMJ 2004 May 15;328(7449):1166 [FREE Full text] [Medline: 15142921] [doi: 10.1136/bmj.328.7449.1166]
12. Kummervold PE, Gammon D, Bergvik S, Johnsen JAK, Hasvold T, Rosenvinge JH. Social support in a wired world: use of online mental health forums in Norway. Nord J Psychiatry 2002;56(1):59-65. [Medline: 11869468] [doi: 10.1080/08039480252803945]
13. Chen JY, Lin CC, Bai YM. Satisfaction with eGroups among persons with psychiatric disorders. Psychiatr Serv 2004 Aug;55(8):945 [FREE Full text] [Medline: 15292552] [doi: 10.1176/appi.ps.55.8.945]
14. Hsiung RC. The best of both worlds: an online self-help group hosted by a mental health professional. Cyberpsychol Behav 2000;3(6):935-950. [doi: 10.1089/109493100452200]
15. Houston TK, Cooper LA, Ford DE. Internet support groups for depression: a 1-year prospective cohort study. Am J Psychiatry 2002 Dec;159(12):2062-2068 [FREE Full text] [Medline: 12450957] [doi: 10.1176/appi.ajp.159.12.2062]
16. Andersson G, Bergström J, Holländare F, Carlbring P, Kaldo V, Ekselius L. Internet-based self-help for depression: randomised controlled trial. Br J Psychiatry 2005 Nov;187(5):456-461 [FREE Full text] [Medline: 16260822] [doi: 10.1192/bjp.187.5.456]
17. Muncer S, Burrows R, Pleace N, Loader B, Nettleton S. Births, deaths, sex and marriage ... but very few presents? A case study of social support in cyberspace. Crit Public Health 2000;10(1):1-18. [doi: 10.1080/713658221]
18. Muncer S, Loader B, Burrows R, Pleace N, Nettleton S. Form and structure of newsgroups giving social support: A network approach. Cyberpsychol Behav 2000;3(6):1017-1029. [doi: 10.1089/109493100452282]
19. Powell J, McCarthy N, Eysenbach G. Cross-sectional survey of users of Internet depression communities. BMC Psychiatry 2003 Dec 10;3(1):19 [FREE Full text] [Medline: 14664725] [doi: 10.1186/1471-244X-3-19]
20. Salem DA, Bogat GA, Reid C. Mutual help goes on-line. J Community Psychol 1997;25(2):189-207. [doi: 10.1002/(SICI)1520-6629(199703)25:2]
21. Roberts LJ, Luke DA, Rappaport J, Seidman E, Toro PA, Reischl TM. Charting uncharted terrain: a behavioral observation system for mutual help groups. Am J Community Psychol 1991 Oct;19(5):715-737. [Medline: 1763786] [doi: 10.1007/BF00938041]
22. Fekete S. The Internet - a new source of data on suicide, depression and anxiety: a preliminary study. Arch Suicide Res 2002;6(4):351-361. [doi: 10.1080/13811110214533.]
23. Weintraub W, Aronson H. Verbal behavior analysis and psychological defense mechanisms. VI. Speech pattern associated with compulsive behavior. Arch Gen Psychiatry 1974 Mar;30(3):297-300. [Medline: 4813132]
24. Weintraub W. Verbal behavior in everyday life. New York: Springer; 1989.
25. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychol Bull 1985 Sep;98(2):310-357. [Medline: 3901065] [doi: 10.1037/0033-2909.98.2.310]
26. Borgatti S, Everett MG, Freeman L. UCINET Version 1. Natick, MA: Analytic Technologies; 1999.
27. Alexander SC. Leadership and social support on the internet: a longitudinal analysis of mental and physical illness groups [dissertation]. Urbana-Champaign, IL: University of Illinois; 2002.
28. Cutrona C, Suhr J, MacFarlane R. Interpersonal transactions and the psychological sense of support. In: Duck S, Silver R, editors. Personal Relationship and Social Support. London, UK: Sage; 1990:30-45.

29. Witt RT. Comparison of self-help computer-mediated communication groups: a CTF analysis [dissertation]. Fresno, CA: Alliant International University; 2000.
30. Canfield ML. Computerized rating of the psychotherapeutic process. Psychotherapy 1991;28(2):304-316.
31. Macias W, Lewis LS, Smith TL. Health-related message boards/chat rooms on the Web: discussion content and implications for pharmaceutical sponsorships. J Health Commun 2005;10(3):209-223. [Medline: 16036729] [doi: 10.1080/10810730590934235]
32. Lamerichs J, Te Molder HFM. Computer-mediated communication: from a cognitive to a discursive model. New Media & Society 2003;5(4):451-473. [doi: 10.1177/146144480354001]
33. Alexander SC, Peterson JL, Hollingshead AB. Help is at your keyboard: support groups on the Internet. In: Frey LR, editor. Group communication in context: Studies of bona fide groups. 2nd edition. London, UK: Lawrence Erlbaum Associates; 2003:309-334.

Abbreviations

ISG: Internet support group

Edited by G Eysenbach; submitted 15.07.09; peer-reviewed by J Powell; accepted 31.07.09; published 30.09.09.

Please cite as:
Griffiths KM, Calear AL, Banfield M, Tam A
Systematic Review on Internet Support Groups (ISGs) and Depression (2): What Is Known About Depression ISGs?
J Med Internet Res 2009;11(3):e41
URL: http://www.jmir.org/2009/3/e41/
doi:10.2196/jmir.1303
PMID:19793718

Original Paper

Effectiveness of Active-Online, an Individually Tailored Physical Activity Intervention, in a Real-Life Setting: Randomized Controlled Trial

Miriam Wanner[1,2], MSc; Eva Martin-Diener[1], MPH; Charlotte Braun-Fahrländer[2], MD; Georg Bauer[3,4], MD, DrPH; Brian W Martin[1,3], MD, MPH

[1]Swiss Federal Institute of Sport Magglingen, Magglingen, Switzerland

[2]Institute of Social and Preventive Medicine, University of Basel, Basel, Switzerland

[3]Institute of Social and Preventive Medicine, University of Zurich, Zurich, Switzerland

[4]Center for Occupational and Organizational Sciences ETH Zurich, Zurich, Switzerland

Corresponding Author:
Miriam Wanner, MSc
Swiss Federal Institute of Sport Magglingen
Brésil, Alpenstrasse 21+23
CH-2532 Magglingen
Switzerland
Phone: +41 32 327 6138
Fax: +41 32 327 6404
Email: Miriam.Wanner@baspo.admin.ch

Abstract

Background: Effective interventions are needed to reduce the chronic disease epidemic. The Internet has the potential to provide large populations with individual advice at relatively low cost.

Objective: The focus of the study was the Web-based tailored physical activity intervention Active-online. The main research questions were (1) How effective is Active-online, compared to a nontailored website, in increasing self-reported and objectively measured physical activity levels in the general population when delivered in a real-life setting? (2) Do respondents recruited for the randomized study differ from spontaneous users of Active-online, and how does effectiveness differ between these groups? (3) What is the impact of frequency and duration of use of Active-online on changes in physical activity behavior?

Methods: Volunteers recruited via different media channels completed a Web-based baseline survey and were randomized to Active-online (intervention group) or a nontailored website (control group). In addition, spontaneous users were recruited directly from the Active-online website. In a subgroup of participants, physical activity was measured objectively using accelerometers. Follow-up assessments took place 6 weeks (FU1), 6 months (FU2), and 13 months (FU3) after baseline.

Results: A total of 1531 respondents completed the baseline questionnaire (intervention group n = 681, control group n = 688, spontaneous users n = 162); 133 individuals had valid accelerometer data at baseline. Mean age of the total sample was 43.7 years, and 1146 (74.9%) were women. Mixed linear models (adjusted for sex, age, BMI category, and stage of change) showed a significant increase in self-reported mean minutes spent in moderate- and vigorous-intensity activity from baseline to FU1 (coefficient = 0.14, P = .001) and to FU3 (coefficient = 0.19, P < .001) in all participants with no significant differences between groups. A significant increase in the proportion of individuals meeting the HEPA recommendations (self-reported) was observed in all participants between baseline and FU3 (OR = 1.47, P = .03), with a higher increase in spontaneous users compared to the randomized groups (interaction between FU3 and spontaneous users, OR = 2.95, P = .02). There were no increases in physical activity over time in any group for objectively measured physical activity. A significant relation was found between time spent on the tailored intervention and changes in self-reported physical activity between baseline and FU3 (coefficient = 1.13, P = .03, intervention group and spontaneous users combined). However, this association was no longer significant when adjusting for stage of change.

Conclusions: In a real-life setting, Active-online was not more effective than a nontailored website in increasing physical activity levels in volunteers from the general population. Further research may investigate ways of integrating Web-based physical activity interventions in a wider context, for example, primary care or workplace health promotion.

(J Med Internet Res 2009;11(3):e23) doi:10.2196/jmir.1179

KEYWORDS

Effectiveness; tailored intervention; adults; Internet; physical activity

Introduction

To reduce the burden of chronic disease and premature death due to an inactive lifestyle [1-3], interventions are needed that are effective in enhancing physical activity levels in the general population. In Switzerland, the health-enhancing physical activity (HEPA) recommendations advocate at least 30 minutes of moderate activity on most, preferably all, days of the week or at least 20 minutes of vigorous activity on three or more days of the week [4]. However, only 36% of the adult population in Switzerland meets either of these recommendations [5]. Thus, effective interventions reaching large numbers are required.

Computer-tailored interventions simulate a personal counseling situation by providing individual feedback based on the behavior, motivation, and attitudes of the user [6]. Tailored interventions of the first generation (using print materials for assessment and dissemination) have been effective in inducing behavior changes for smoking [7,8], nutrition [9,10], and physical activity [11-15]. Second-generation interventions use the advantages of the Internet—interactivity, availability at any time from any place, and immediate display of feedback—to potentially reach large populations at relatively low cost.

To date, studies investigating the effectiveness of second-generation Web-based tailored physical activity interventions have either been carried out in small confined populations [16-18], have not used truly tailored information but materials targeted to the stages of change [19,20], have looked at only short-term effects [21], or have been carried out in optimized and controlled settings such as computer labs [22]. Results from these studies were mixed [23,24]. Interventions shown to be effective in controlled settings may still be ineffective if delivered in an uncontrolled, real-life setting. The potential public health impact of Web-based computer-tailored interventions can only be estimated if their effectiveness is tested under real-life conditions. Thus, studies evaluating online physical activity interventions in real-life settings are now needed. To our knowledge, there is only one individually tailored Internet-based intervention targeting physical activity that has been evaluated in two samples of the general population in a real-life setting, showing mixed results [25,26].

Intensity of intervention use may be associated with induced physical activity changes [27]. However, little is known about the impact of frequency and duration of intervention use on the effectiveness of a Web-based tailored physical activity intervention. This may be an important issue for the interpretation of results from real-life effectiveness studies.

The focus of the present study was a Web-based tailored physical activity intervention that is freely accessible on the Internet [28]. Active-online was tested for its acceptability and feasibility before the final version went online in 2003 [29]. The main research questions were (1) How effective is Active-online, compared to a nontailored website with general information on physical activity and health, in increasing self-reported and objectively measured physical activity levels in the general population when delivered in a real-life setting? (2) Do respondents recruited for the randomized study differ from spontaneous users of Active-online, and how does effectiveness differ between these groups? (3) What is the impact of frequency and duration of use of Active-online on changes in physical activity behavior?

Methods

Study Design, Setting, and Participants

Participants for this Web-based study were recruited by advertisements in newspapers, in magazines, and on the Internet. They were invited to take part in a physical activity study and were given the link to the study website (with a domain name different from the one of the intervention). At the same time, spontaneous users were recruited directly from the Active-online website by redirecting them to the study website if they chose to participate in the study. The study was carried out in German, and recruitment lasted from May 1 to August 2, 2006. Based on sample size calculations assuming an increase in meeting the HEPA recommendations of 30% in the intervention group and 20% in the control group (alpha = .05, power = 0.8), 250 participants were required per group. Assuming a realistic loss-to-follow-up in a Web-based survey without face-to-face contact of about 50% over 1 year [30], this number doubled to 500 participants per group.

Interested individuals could access the study website from any computer with Internet access. Information regarding the study and all study questionnaires were provided there. Individuals completing the baseline questionnaire and leaving their email address were registered. Media-recruited participants were randomly allocated to either the intervention group (IG) or the control group (CG) and were forwarded to Active-online or the nontailored website, respectively. Spontaneous users (SU) were included as a separate study group but followed the same study protocol as the IG.

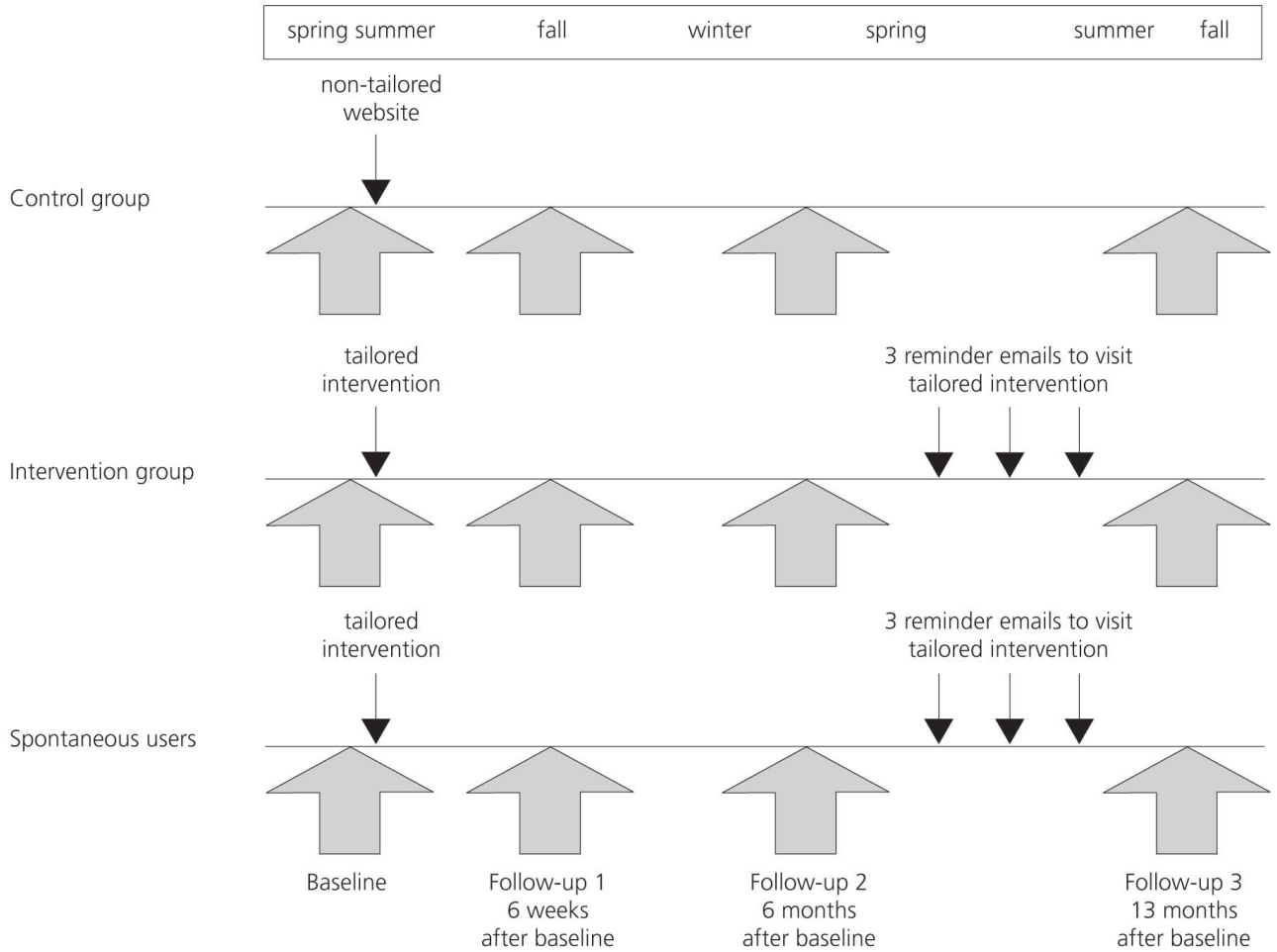
Respondents could volunteer to take part in accelerometer measurements via an additional Web page that they were routed to after the baseline questionnaire, depending on the availability of accelerometers. Volunteers were not forwarded directly to the intervention websites but were sent an accelerometer to obtain baseline measurements and had to return a separate written consent form. Only after the accelerometer was returned was an email sent out with a link to Active-online or the nontailored website.

Randomization was carried out using random numbers provided by the University of Geneva's online service [31] based on a physical quantum random number generator. Participants were not aware of the group they were randomized to. The study website with all the study questionnaires was kept strictly separate from Active-online, using two different domains, in

order to minimize the chance of controls getting involved with Active-online.

Email addresses were used to identify and contact participants at follow-up. All participants were followed up 6 weeks (FU1), 6 months (FU2), and 13 months (FU3) after the baseline assessment, receiving a maximum of three email invitations each time with a personal link referring them back to the study website. Those volunteers having participated in the accelerometer measures at baseline were asked to repeat accelerometer measures at each follow-up in addition to the online questionnaires. Individuals in the IG and SU additionally received three reminder emails with a personal link to Active-online between FU2 and FU3 at 9, 10, and 11 months after the baseline assessment, encouraging them to revisit Active-online. The study procedure is depicted in Figure 1 according to group. The study used an automated design with emails automatically timed to each participant's starting date. There were no face-to-face contacts. The study design had been tested in a feasibility study [32]. As incentives, two city bikes were being raffled among the participants who completed the study. The study was approved by the ethics commission of the canton of Berne, Switzerland. The trial was not registered as the funding agency (Swiss Federal Council of Sports ESK) did not request a trial registration and we were not aware when the study started in 2006 that web-based trials should be registered. In lieu of trial registration, we append the original application for funding, containing the protocol (Multimedia Appendix 1).

Figure 1. Study procedure for each group



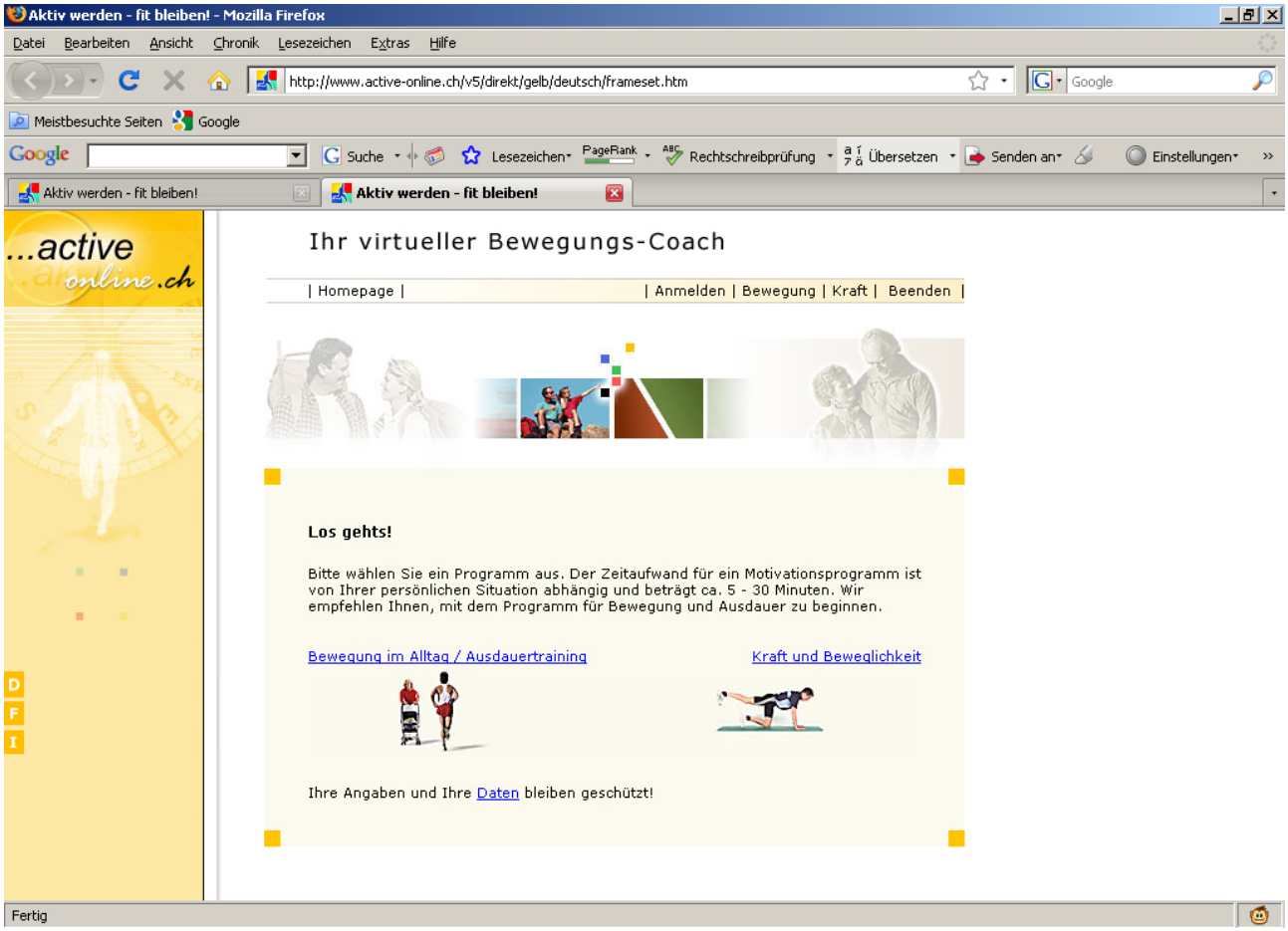
Tailored Intervention and Standard Website

Active-online is an interactive, individually tailored physical activity program targeting individuals aged 30 to 60 years. It has been freely available on the Internet since 2003. The aim of the program is to increase physical activity levels in users by offering individually tailored counseling and motivational feedback. The program was developed in German by an interdisciplinary team of experts in public health, sport sciences, psychology, design, and computer science, and then translated and culturally adapted for French and Italian audiences. The theoretical framework of the program is the transtheoretical model of behavior change [33]. Visitors may choose one of two tailored modules, either on everyday activities and endurance training or on strength and flexibility training. Figure 2 shows a screenshot of the Active-online page where one of the two modules can be selected. The first module offers a maximum of four tailored feedbacks using questionnaires on stages of change, decisional balance, processes of change, and self-efficacy. Stages of change are assessed according to a

seven-stage concept focusing on current behavior (moderate- and vigorous-intensity activity) as well as on intention to change [34]. The decisional balance and self-efficacy scales are based on instruments of Basler et al [35], and the processes of change scale on instruments of Marcus et al [36], and Nigg and Riebe [37]. Depending on their current stage of change, visitors are guided trough one, three, or four sections of the module. More information is available in Martin-Diener et al [34]. The module on strength and flexibility training offers a maximum of two tailored feedbacks based on questionnaires assessing five stages of change as well as attitudes and knowledge regarding strength training. The feedback for flexibility training is based on current behavior. Based on answers to these "diagnostic" questionnaires, short text segments are selected from the feedback library, compiled into unique individual feedback, and displayed immediately on screen. Feedback reports are available in a printer-friendly format. Additional support tools, such as strength and stretching exercise sheets, and organizational and motivational download forms, are provided.

Users may visit Active-online without registering, or they may register with their email address to obtain a password. Registered users have the possibility of following changes in their physical activity behavior when revisiting the website. They also receive reminder emails encouraging them to revisit Active-online.

Figure 2. Screenshot of the tailored intervention



Participants in the CG were forwarded to a nontailored website with general information on physical activity and health with no additional reminder emails. This was a static website with some tips on how to include more physical activity in daily life and some information regarding positive health effects of physical activity. Figure 3 shows a screenshot of the nontailored website.

Figure 3. Screenshot of the standard website for the CG

Measures

The online baseline questionnaire included questions on demographics, physical activity behavior, stage of change, self-efficacy, and general and mental health. Data are only presented for physical activity. Physical activity was assessed using a short questionnaire with four items on frequency and duration of moderate- and vigorous-intensity activity that is used in the official monitoring of physical activity in the Swiss population [38]. These questions allow the calculation of total minutes of moderate- and vigorous-intensity activity per week (total reported activity time) and the classification of participants according to the HEPA recommendations as outcome measures. Reported times exceeding 8 hours per day or 40 hours per week of moderate-intensity activity and 5 hours per day or 17.5 hours per week of vigorous-intensity activity were set to missing. Truncating these high values instead of setting them to missing did not change the results. Demographic variables included age, gender, living situation, highest education, nationality, smoking status, height, and weight. Body mass index (BMI) was calculated as weight (in kg) divided by height (in meters squared) and categorized as < 18.5, 18.5 to < 25, 25 to < 30, and ≥ 30 kg/m^2. The same questionnaire (except demographic variables) was used in the follow-up assessments.

Accelerometers (Actigraph models AM7164 and GT1M, formerly Computer Science and Applications, now Manufacturing Technology Inc, Fort Walton Beach, FL, USA) were used for objective physical activity assessment. The accelerometers have been validated in earlier studies [39,40]. Participants were asked to wear the accelerometer on their right hip during waking hours for a 7-day period at baseline and each follow-up. A minimum of 4 days with at least 10 hours per day of data recording, including one weekend day, were required to be included in the analysis. The data collected by the accelerometers are a series of counts integrating vertical acceleration over a specified time interval (epoch time). Epoch time was set to 1 minute. Cut-off points developed by Swartz et al were used to classify light (≤ 573 counts/minute), moderate (574-4944 counts/minute), and vigorous activities (≥ 4945 counts/minute) [41]. These cut-offs were chosen because they were derived using a wide range of lifestyle activities and may prove applicable for predicting time spent in different intensity categories during free-living activities [42]. Mean counts per minute over the recording period and minutes of moderate and vigorous activity per week according to accelerometer data (total accelerometry activity time, only bouts of at least 10 minutes) were calculated.

Data regarding the use of Active-online for the IG and SU were obtained from the Active-online user database. Each visit to the website was recorded, including start date and time, end date and time, number of pages viewed, etc. Participants were provided with a password to re-enter Active-online in order to track their repeated visits. Use of the nontailored website in the CG was not measured.

Statistical Analyses

Minutes of physical activity were positively skewed and were log-transformed for analysis. Chi-square tests for categorical variables and t-tests for continuous variables were used to compare responders and nonresponders and to compare differences between IG and CG and between IG and SU at baseline. In a preliminary analysis, paired t-tests and McNemar tests were applied, respectively, to compare changes in total activity time and changes in the proportion meeting the HEPA

recommendations between baseline and each FU and for each group separately. Mixed logistic and mixed linear models were used to simultaneously analyze the effects of time and group allocation on the proportion meeting the HEPA recommendations and on total activity time, respectively, including gender, age, BMI category, and stage of change at baseline as covariates in the adjusted model. Stage of change was included to account for baseline motivation to change. The inclusion of time-group interaction terms in mixed models allows identification of potential differences in changes between groups at any time point. Changes in total reported activity time were analyzed for all participants and separately for participants meeting and not meeting the HEPA recommendations at baseline, because the latter are those most in need of effective interventions to increase their physical activity behavior. Participants were analyzed as randomized.

The impact of the use of Active-online on changes in physical activity behavior in the IG and SU was analyzed with a linear regression model including the difference in total reported activity time between baseline and FU3 as the dependent variable and the minutes spent in the tailored intervention as the independent variable, including gender, age, BMI category, and stage of change at baseline as covariates in the adjusted model. STATA 9.2 (STATACorp LP, College Station, TX, USA) was used for all analyses.

Results

Participants

In total, 1919 respondents recruited via different media channels and 220 respondents recruited via the Active-online website started the baseline survey on our study website; 1401 and 168, respectively, finished the survey and were registered as participants (Figure 4). We excluded 38 respondents due to technical problems; 1369 were randomized into the IG (n = 681) or the CG (n = 688), and 162 were registered as SU (7.4% of all visits recorded in the Active-online user database during the recruitment period). The response is shown in Figure 4 according to group allocation and follow-up. No difference in response was seen at FU1. At FU2, response was significantly lower in the IG compared to CG, and in SU compared to IG. At FU3, response was significantly lower in both the IG and SU compared to CG, with no significant difference between IG and SU. Depending on availability of accelerometers, 326 participants (21.3%) had the choice to wear an accelerometer. Of those, 144 (44.2%) agreed to take part in the objective measures, corresponding to 9.4% of the total sample.

Figure 4. Participant flow: recruitment channels, randomization, baseline, and follow-up assessments

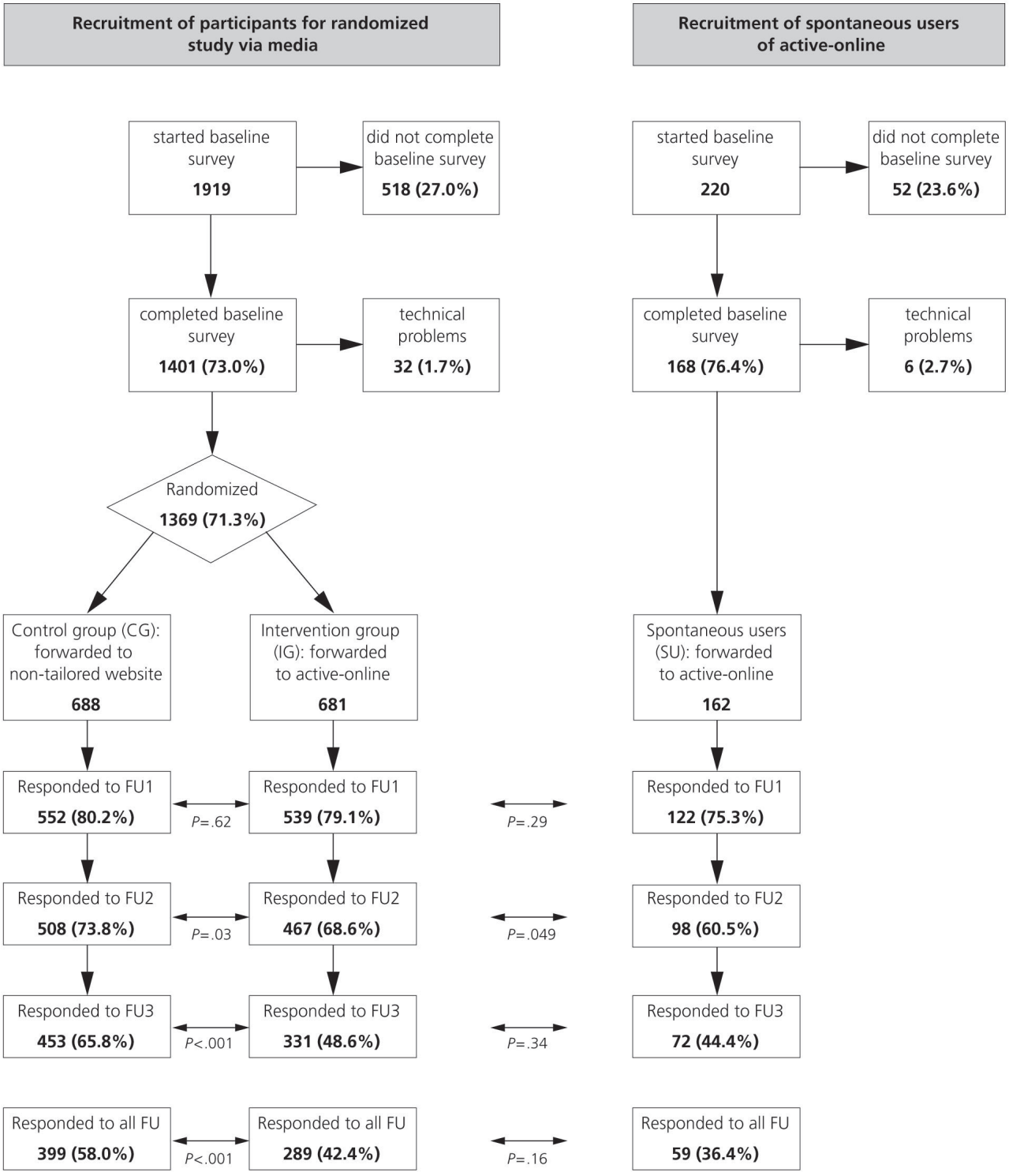


Table 1 displays the baseline characteristics of the total sample and of each group separately. There were no significant differences in demographic variables between the randomized groups (IG and CG). However, compared to the IG, SU were significantly younger (38.8 versus 44.2 years, P < .001), more likely to be smokers (23.5% versus 12.8%, P < .001), and less likely to be living with children (38.9% versus 53.5%, P < .001).

Table 1. Characteristics of participants at baseline according to group[a]

Self-Reported Measures	Total (n = 1531)	CG (n = 688)	IG (n = 681)	P(IG-CG)	SU (n = 162)	P(IG-SU)
Demographic variables						
Female (%)	74.9	75.9	74.7	.63	71.0	.33
Age, years	43.7 ± 13.1	44.2 ± 12.8	44.2 ± 13.3	.99	38.8 ± 13.0	< .001
Age groups (%)				.32		< .001
< 30 years	16.2	13.8	15.3		30.3	
30-60 years	72.9	75.6	72.1		64.8	
> 60 years	10.9	10.6	12.6		4.9	
Living with a partner (%)	70.0	70.4	70.8	.86	65.4	.18
Living with children (%)	53.2	56.3	53.5	.30	38.9	.001
Swiss nationality (%)	87.3	86.2	88.6	.19	86.4	.45
University degree (%)	24.9	25.2	24.1	.65	27.2	.41
Health-related variables						
Smokers (%)	13.1	10.9	12.8	.28	23.5	.001
BMI, kg/m^2	24.6 ± 4.6	24.5 ± 4.5	24.8 ± 4.6	.38	24.5 ± 4.6	.57
Overweight and obese (%)	39.3	38.3	41.1	.30	36.4	.28
Physical activity-related variables						
Meeting HEPA recommendations (%)	40.8	40.4	40.9	.84	42.2	.75
Total reported activity time, minutes/week	277 ± 253	276 ± 256	276 ± 258	.99	283 ± 222	.76
Objective Measures	Total (n = 133)	CG (n = 52)	IG (n = 62)	P(IG-CG)	SU (n = 19)	P(IG-SU)
Objective physical activity						
Mean counts per minute	451 ± 186	450 ± 176	457 ± 196	.85	436 ± 193	.69
Total accelerometry activity time, minutes/week	377 ± 214	383 ± 211	383 ± 227	.99	341 ± 183	.47

[a]Values are mean ± SD unless otherwise noted.

There were significant differences in some variables between participants who responded to each follow-up (responders) and those who did not respond to at least one follow-up (nonresponders). Nonresponders were slightly younger, less likely to be Swiss, more likely to be smokers at baseline, more likely to be overweight or obese, and less likely to meet the HEPA recommendations at baseline. The subgroup of participants with accelerometers (n = 144) were slightly older, more likely to live with children, and more likely to be overweight or obese than those not participating in the accelerometer part of the study.

Self-Reported Physical Activity

When including those participants with complete data for all four time points (n = 736), significant increases in the proportion of participants meeting the HEPA recommendations were observed in SU between baseline and FU1 (P = .045) and FU3 (P = .002). Nonsignificant increases between baseline and FU3 were seen in the IG and CG. Changes in total reported activity time per week between baseline and FU3 are depicted in Figure 5 according to group, for all participants with complete data (n = 736) and separately for those individuals meeting (n = 336) and not meeting (n = 400) the HEPA recommendations at baseline. When including only those participants who did not meet the HEPA recommendations at baseline, total reported activity time increased significantly in all groups. The increases observed in these insufficiently active individuals exceeded the increase observed in all participants; thus, a decrease in total reported activity time was found in those individuals meeting the HEPA recommendations at baseline. The decrease was significant in the CG.

Figure 5. Changes in total reported activity time (minutes/week) between baseline and FU3 according to group, for all participants with complete data (n = 736) and separately for those meeting (n = 336) and not meeting (n = 400) the HEPA recommendations at baseline

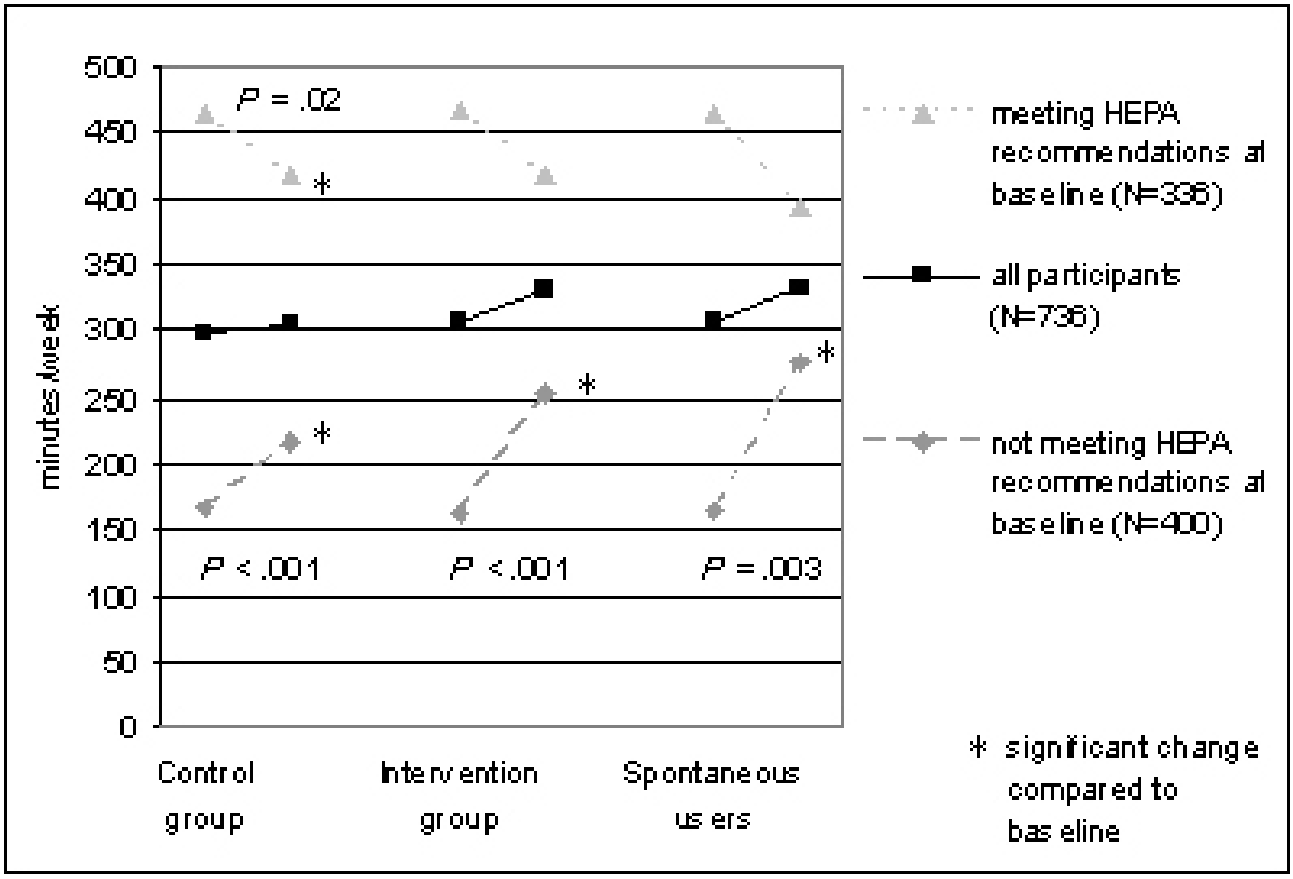


Table 2 shows the percent changes in the proportion meeting the HEPA recommendations and changes in total reported activity time between baseline and each follow-up according to group, including those participants who responded to the specific follow-up. A significant increase in individuals meeting the HEPA recommendations was observed in SU between baseline and FU3. Total reported activity time was generally lower at FU2; however, there were no significant changes at any follow-up or in any group.

Table 2. Percent changes in self-reported physical activity between baseline and each follow-up according to group[a]

	Baseline to FU1	*P*	Baseline to FU2	*P*	Baseline to FU3	*P*
Meeting HEPA recommendations						
CG	+2.2%	.27	−0.8%	.71	+3.8%	.12
IG	+2.3%	.27	−1.7%	.45	+4.0%	.21
SU	+7.4%	.11	−2.0%	.69	+18.3%	.005
Total reported activity time (minutes/week)						
CG	+4.8%	.13	−3.7%	.27	+4.5%	.25
IG	+4.7%	.14	−2.6%	.52	+3.4%	.51
SU	+9.6%	.11	+4.9%	.48	+15.4%	.19

[a]Results are based on participants with complete data at two time points (see Figure 4 for number of participants).

Table 3 shows the results from mixed logistic and mixed linear models for all participants, evaluating simultaneously the effect of time and group allocation. There were no differences between groups regarding the HEPA recommendations; however, a borderline significant increase in total reported activity time was found for SU compared to the CG. Irrespective of group allocation, participants were significantly more likely to meet the HEPA recommendations at FU3 compared to baseline, and a significant increase in total reported activity time was observed between baseline and FU1 as well as between baseline and FU3. There was a significant interaction between FU3 and SU in both the unadjusted (OR = 2.83, P = .03) and adjusted logistic model (OR = 2.95, P = .02), indicating that the proportion meeting the HEPA recommendations was significantly higher in SU

compared to the randomized groups at FU3. There were no interactions between time of follow-up and group allocation in the mixed linear model, indicating that there were no differences in total reported activity time between groups at any follow-up.

Table 3. Time and group parameters for changes in physical activity, based on mixed logistic and mixed linear models[a]

	Meeting HEPA Recommendations				Total Reported Activity Time (minutes/week)			
	Unadjusted		Adjusted		Unadjusted		Adjusted	
	OR	95% CI	OR	95% CI	Coeff	95% CI	Coeff	95% CI
Group								
IG	1.04	0.68-1.57	1.02	0.72-1.45	0.02	−0.10, 0.15	0.02	−0.08, 0.13
SU	1.15	0.59-2.24	1.08	0.62-1.89	0.16	−0.04, 0.35	0.17	0.000-0.35
Time								
FU1 (6 weeks)	1.34	0.96-1.85	1.30	0.93-1.82	0.15	0.06-0.23	0.14	0.05-0.22
FU2 (6 months)	1.04	0.75-1.46	1.01	0.71-1.42	0.02	−0.06, 0.11	0.02	−0.07, 0.10
FU3 (13 months)	1.49	1.05-2.11	1.47	1.03-2.09	0.19	0.10-0.28	0.19	0.10-0.28

[a]Basic unit is the CG at baseline. Adjusted models include gender, age, BMI category, and stage of change at baseline.

Objectively Measured Physical Activity

At baseline, 144 individuals (56 in CG, 68 in IG, and 20 in SU) wore an accelerometer, resulting in valid data for 133 individuals (92.4%). Valid accelerometer data were available for 117 individuals (88.0% of those with valid data at baseline) at FU1, for 114 individuals (85.7%) at FU2, and for 105 individuals (78.9%) at FU3; 93 participants (69.9%) had complete accelerometer data. There were no differences between groups.

Table 4 shows the percent changes in counts per minute and in total accelerometry activity time between baseline and each follow-up for each group separately. There were no significant changes observed in the IG. In the CG, activity levels decreased significantly between baseline and FU2 as well as between baseline and FU3. In SU, activity levels decreased significantly between baseline and FU2. Mixed linear models did not show any significant effects for time and group and no interaction effects.

Table 4. Percent changes in objective physical activity between baseline and each follow-up according to group[a]

	Baseline to FU1	*P*	Baseline to FU2	*P*	Baseline to FU3	*P*
Counts per minute						
CG	−4.8%	.19	−11.6%	.004	−8.1%	.03
IG	+2.5%	.52	−8.0%	.06	−1.2%	.83
SU	+1.5%	.83	−16.2%	.04	−2.1%	.81
Total accelerometry activity time (minutes/week)						
CG	−6.2%	.29	−17.5%	< .001	−10.3%	.03
IG	−1.5%	.82	−5.4%	.29	−1.2%	.87
SU	−5.1%	.72	−29.1%	.045	−16.0%	.16

[a]Results are based on participants with complete data at two time points.

Frequency and Duration of Use of Active-Online

In total, 2112 visits of IG and SU study participants (n = 843) were counted on Active-online, with a mean number of 2.5 (± 1.6) visits per person. The number of visits was described by a positively skewed distribution representing 50% with two or less visits on Active-online during the study period between baseline and FU3. On average, 46 pages were viewed per person, with a median of 31 pages.

In 1226 of all visits (58.0%), one of the two tailored modules was started. These 1226 visits can be attributed to 628 individuals (74.5% of all participants in IG and SU). The mean number of visits within a tailored module for these individuals was 1.9 (± 1.2). The mean and median time spent in the modules for participants who started a tailored module was 12 minutes and 9 minutes per visit, respectively, and 23 minutes and 15 minutes during the whole study period, respectively.

In 962 of all visits (45.5%), at least one tailored feedback was obtained in the module on everyday activities and endurance training, and in 460 of all visits (21.8%), at least one tailored feedback was obtained in the module on strength and flexibility training. There was no difference in the use of Active-online between the IG and SU.

In the CG, 62 of 453 participants responding to FU3 (13.7%) stated that they had heard about Active-online and had used it at least once during the preceding year.

Linear regression showed a weak but significant relation between total minutes spent within one of the tailored modules (IG participants and SU combined) and changes in total reported activity time between baseline and FU3 in the unadjusted model (coefficient = 1.13, 95% CI 0.09 - 2.17, P = .03), and a borderline significant relation in the model adjusted for age, gender, and BMI category (coefficient = 1.07, 95% CI 0.004 - 2.13, P = .049). When adding stage of change to the model, the relation was attenuated and no longer significant (coefficient = 0.58, 95% CI −0.43 to 1.59, P = .26), indicating that stage of change was associated with both changes in total reported activity time as well as time spent in the tailored modules. There was no interaction between stage of change and time spent in the tailored modules.

Discussion

Principle Results and Comparison With Prior Work

In the present study, there were significant increases in self-reported physical activity levels between baseline and the last follow-up after 13 months in all participants, but there were no significant differences between the randomized groups. More pronounced increases were found in SU of Active-online. However, these individuals were not randomized and thus cannot be directly compared with the randomized groups. Furthermore, SU willing to participate in the study may not be representative of all Active-online users since they were a self-selected sample and only represented 7.4% of all visits on Active-online during the recruitment period.

Self-reported changes in physical activity levels were not confirmed by objective measures. Differences between self-reported and objective measures may be due to the possibility that study participation influenced the perception of physical activity behavior and thus reporting of physical activity levels. A seasonal pattern [43], with lower activity levels in winter (FU2), was observed in both self-reported and objective physical activity data.

Results of other computer-tailored [10] and Web-based tailored [23,24] physical activity intervention studies have been mixed. The results in the present study are comparable with other studies investigating Web-based physical activity interventions. While some studies produced effective results in the short term [20,21] or when compared to a waiting list control group [26], others showed improvements in physical activity levels in both intervention and control conditions [17], like we did with regard to self-reported physical activity. A tailored intervention that has been effective when delivered on CD-ROM in a controlled setting after 6 months [44] and after 2 years [22] was not effective when delivered online in a real-life setting compared with online standard advice [25]. Similar to our study, Spittaels et al also found increases in self-reported physical activity levels in both intervention and control groups, and increases in physical activity levels were not confirmed by accelerometer data [25]. The present study adds evidence to the point that effectiveness of a Web-based physical activity intervention may be difficult to demonstrate when delivered in an uncontrolled setting.

As per the real-life setting, study participants were free to start and stop the intervention. In addition, the anonymous nature of the Internet and the wealth of available information may make it difficult to achieve sufficient levels of intervention use. On average, individuals in the IG and SU started a tailored module less than twice during the study period, accumulating a mean of 23 minutes in the tailored modules in total (12 minutes per visit). In a study assessing user attitudes toward a physical activity website, an average time of 7.1 minutes spent on the tailored intervention per visit and a total average of 356 minutes over 1 year was reported [45]. While the duration per visit was higher in our study sample, the total accumulated time spent on the intervention was 15 times higher in the other study. Leslie et al reported that participants who entered a tailored Web-based physical activity intervention spent, on average, 9 minutes per visit [46]. However, 152 participants produced 4114 visits on the website over 8 weeks, indicating that the accumulated exposure was clearly higher in that study than in our sample. Low exposure to intervention materials has been reported in other studies using objective data on website usage, indicating that achieving engagement in website-delivered physical activity interventions is challenging [23]. Moreover, one quarter of the participants in the IG and SU did not start a tailored module at all, and 13.7% of controls used Active-online independently of the study, suggesting some degree of contamination in the IG and CG. This may have reduced a potential effect but reflects the real-life delivery mode used in this study. While correlations between log-in frequency and weight change have been reported in a study focusing on a Web-based behavioral weight loss program [47], the role of frequency and duration of use of Active-online on changes in physical activity behavior could not be clarified in this study.

Because of the challenges that we face with stand-alone Web-based interventions that are freely accessible on the Internet, it may be more promising to embed a program like Active-online in a wider context of health promotion. Possibilities for better utilization of Active-online may be its application in a workplace setting, the “prescription” of Active-online to patients in primary care, or the inclusion of Active-online in a larger health promotion packet targeting different health issues, for example, in a community setting. Two studies that have evaluated Web-based tailored interventions in a primary care setting have reported increases in physical activity levels after 1 month [48] and after 6 weeks [49]. A study in two manufacturing and two office sites showed high levels of engagement in a Web-based and monitoring device–based physical activity and weight management program in a wide range of employees [50]. A computer-tailored (but not Web-based) intervention for nutrition and physical activity in a workplace setting demonstrated increases in the frequency of strengthening and flexibility exercise compared to a delayed group [11]. A Web-based workplace health promotion program targeting nutrition, stress, and physical activity did not outperform print materials used in the control group, even though improvements in some physical activity variables were reported in both groups [51]. Further research may investigate

possibilities of integrating Web-based interventions in a wider health promotion context. Marcus et al especially highlight the urgent need for research on Internet-based physical activity programs within the context of primary care [52].

Strengths and Limitations

A strength of the study was the delivery of the Web-based intervention under real-life conditions, not in a controlled setting. There were no face-to-face contacts or other factors that may increase compliance, because they do not represent realistic conditions for open-access Web-based interventions. Furthermore, objective physical activity assessment was used in a subsample of participants in addition to the questionnaires. We included SU of Active-online as an additional study arm. Frequency and duration of use of Active-online were monitored using objective data from the Active-online user database, making it possible to look at the relation between use of Active-online and physical activity changes. Other strengths are the long-term follow-up and the large number of participants included in the randomized study.

Several reasons may be responsible for the limited effectiveness of Active-online. The website was tested in 2003 and acceptability was generally good; participants especially liked the individual counseling, the pleasant tone, and the simple structure and design [29]. However, Internet technology is changing rapidly and Active-online may already be slightly out-of-date. Furthermore, Active-online is based on the transtheoretical model of behavior change, which was regarded as promising at the time when Active-online was developed, but has more recently been subject to some debate regarding its potential to change behavior [53,54]. In addition, baseline physical activity levels were already quite high in the study sample, with around 280 minutes total reported activity time per week; thus a ceiling effect may have occurred.

This study has several limitations. A rather low overall response of around 50%, as observed in other studies [30], was expected based on the experiences of the feasibility study [32] and was taken into account when calculating the sample size. High drop-out attrition has been recognized as a common problem in Internet-based studies [55]. Nonresponse in this study was differential between groups, with higher drop-outs in the IG and SU than in the CG. Differences between responders and nonresponders [30] and higher drop-out rates in intervention groups have been observed in other studies in the domain of physical activity [20] and nutrition [56,57]. The smaller number of recruited SU and the fact that they only represented 7.4% of all visits on Active-online during the recruitment period limits conclusions about the effectiveness of Active-online in this group. The time spent on Active-online recorded in the database may not represent the actual time spent interacting with the intervention, because Active-online may have been opened in the background while the user was browsing other websites opened simultaneously. On the other hand, Active-online users could print their feedback reports and read them offline. If study participants revisited Active-online without using their password, the estimated number of visits presented here may be conservative. Last, due to a technical problem, the reminder emails to revisit Active-online were not sent out according to the original schedule for registered users of Active-online at 2, 4, and 7 months.

Conclusions

The present study showed limited effectiveness of Active-online in a randomized sample of volunteers from the general adult population when offered as a stand-alone intervention delivered online under real-life conditions. Further research may investigate the potential of Web-based physical activity interventions integrated in a wider context, for example, primary care or workplace health promotion.

Acknowledgments

This study was funded by the Swiss Federal Council of Sports. The authors would like to thank Dr. Christian Schindler for his support with the statistical analyses and the colleagues from the University of Basel who have critically reviewed the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix

Application for funding

[PDF file (Adobe PDF), 114 KB - jmir_v11i3e23_app1.pdf]

References

1. ; Department of Health, Physical Activity, Health Improvement and Prevention. At least five a week: evidence on the impact of physical activity and its relationship to health. London, UK: Department of Health; 2004. URL: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4080994 [WebCite Cache ID 5c6Qz2MVd]
2. ; US Department of Health and Human Services. Physical activity and health: a report of the Surgeon General. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, The President's Council on Physical Fitness and Sports; 1996. URL: http://www.cdc.gov/nccdphp/sgr/pdf/sgrfull.pdf [WebCite Cache ID 5c6RK3IXQ]

3. Bouchard C, Blair SN, Haskell WL. Physical activity and health. Champaign, IL: Human Kinetics; 2007.
4. ; Swiss Federal Office of Sport, Swiss Federal Office of Public Health, Health Promotion Switzerland, Network HEPA Switzerland. Health-enhancing physical activity. A base document. Magglingen, Switzerland: Swiss Federal Office of Sport; 2006. URL: http://www.sport.admin.ch/hepa/downloads/Grundlagendoku_hepa2006_en.pdf [WebCite Cache ID 5c6XW7ayL]
5. Stamm H, Lamprecht M. Structural and cultural factors influencing physical activity in Switzerland. J Public Health 2005;13(4):203-211. [doi: 10.1007/s10389-005-0117-y]
6. Kreuter M, Farrell D, Olevitch L, Brennan L. Tailoring Health Messages. Customizing communication with computer technology. Mahwah, NJ: Erlbaum; 2000.
7. Strecher VJ. Computer-tailored smoking cessation materials: a review and discussion. Patient Educ Couns 1999 Feb;36(2):107-117. [Medline: 10223016] [doi: 10.1016/S0738-3991(98)00128-1]
8. Walters ST, Wright JA, Shegog R. A review of computer and Internet-based interventions for smoking behavior. Addict Behav 2006 Feb;31(2):264-277. [Medline: 15950392] [doi: 10.1016/j.addbeh.2005.05.002]
9. Brug J, Oenema A, Campbell M. Past, present, and future of computer-tailored nutrition education. Am J Clin Nutr 2003 Apr;77(4 Suppl):1028S-1034S [FREE Full text] [Medline: 12663313]
10. Kroeze W, Werkman A, Brug J. A systematic review of randomized trials on the effectiveness of computer-tailored education on physical activity and dietary behaviors. Ann Behav Med 2006 Jun;31(3):205-223. [Medline: 16700634] [doi: 10.1207/s15324796abm3103_2]
11. Campbell MK, Tessaro I, DeVellis B, Benedict S, Kelsey K, Belton L, et al. Effects of a tailored health promotion program for female blue-collar workers: health works for women. Prev Med 2002 Mar;34(3):313-323. [Medline: 11902848] [doi: 10.1006/pmed.2001.0988]
12. Bull FC, Jamrozik K, Blanksby BA. Tailored advice on exercise--does it make a difference? Am J Prev Med 1999 Apr;16(3):230-239. [Medline: 10198663] [doi: 10.1016/S0749-3797(98)00160-3]
13. Bull FC, Kreuter MW, Scharff DP. Effects of tailored, personalized and general health messages on physical activity. Patient Educ Couns 1999 Feb;36(2):181-192. [Medline: 10223022] [doi: 10.1016/S0738-3991(98)00134-7]
14. Marcus BH, Bock BC, Pinto BM, Forsyth LH, Roberts MB, Traficante RM. Efficacy of an individualized, motivationally-tailored physical activity intervention. Ann Behav Med 1998;20(3):174-180. [Medline: 9989324] [doi: 10.1007/BF02884958]
15. Bock BC, Marcus BH, Pinto BM, Forsyth LH. Maintenance of physical activity following an individualized motivationally tailored intervention. Ann Behav Med 2001;23(2):79-87. [Medline: 11394558] [doi: 10.1207/S15324796ABM2302_2]
16. Hageman PA, Walker SN, Pullen CH. Tailored versus standard internet-delivered interventions to promote physical activity in older women. J Geriatr Phys Ther 2005;28(1):28-33. [Medline: 16236225]
17. McKay HG, King D, Eakin EG, Seeley JR, Glasgow RE. The diabetes network internet-based physical activity intervention: a randomized pilot study. Diabetes Care 2001 Aug;24(8):1328-1334 [FREE Full text] [Medline: 11473065] [doi: 10.2337/diacare.24.8.1328]
18. Hurling R, Catt M, de Boni M, Fairley BW, Hurst T, Murray P, et al. Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. J Med Internet Res 2007;9(2):e7 [FREE Full text] [Medline: 17478409] [doi: 10.2196/jmir.9.2.e7]
19. Marshall AL, Leslie ER, Bauman AE, Marcus BH, Owen N. Print versus website physical activity programs: a randomized trial. Am J Prev Med 2003 Aug;25(2):88-94. [Medline: 12880874] [doi: 10.1016/S0749-3797(03)00111-9]
20. Napolitano MA, Fotheringham M, Tate D, Sciamanna C, Leslie E, Owen N, et al. Evaluation of an internet-based physical activity intervention: a preliminary investigation. Ann Behav Med 2003;25(2):92-99. [Medline: 12704010] [doi: 10.1207/S15324796ABM2502_04]
21. Oenema A, Brug J, Dijkstra A, de Weerdt I, de Vries H. Efficacy and use of an internet-delivered computer-tailored lifestyle intervention, targeting saturated fat intake, physical activity and smoking cessation: a randomized controlled trial. Ann Behav Med 2008 Apr;35(2):125-135. [Medline: 18363076] [doi: 10.1007/s12160-008-9023-1]
22. Vandelanotte C, De Bourdeaudhuij I, Brug J. Two-year follow-up of sequential and simultaneous interactive computer-tailored interventions for increasing physical activity and decreasing fat intake. Ann Behav Med 2007 Apr;33(2):213-219. [Medline: 17447874] [doi: 10.1007/BF02879903]
23. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [Medline: 17572313] [doi: 10.1016/j.amepre.2007.02.041]
24. van den Berg MH, Schoones JW, Vliet Vlieland TPM. Internet-based physical activity interventions: a systematic review of the literature. J Med Internet Res 2007;9(3):e26 [FREE Full text] [Medline: 17942388] [doi: 10.2196/jmir.9.3.e26]
25. Spittaels H, De Bourdeaudhuij I, Brug J, Vandelanotte C. Effectiveness of an online computer-tailored physical activity intervention in a real-life setting. Health Educ Res 2007 Jun;22(3):385-396 [FREE Full text] [Medline: 16971674] [doi: 10.1093/her/cyl096]
26. Spittaels H, De Bourdeaudhuij I, Vandelanotte C. Evaluation of a website-delivered computer-tailored intervention for increasing physical activity in the general population. Prev Med 2007 Mar;44(3):209-217. [Medline: 17197015] [doi: 10.1016/j.ypmed.2006.11.010]

27. Hillsdon M, Foster C, Thorogood M. Interventions for promoting physical activity. Cochrane Database Syst Rev 2005(1):CD003180. [Medline: 15674903] [doi: 10.1002/14651858.CD003180.pub2]
28. http://www.active-online.ch/. WebCite Cache not possible .
29. Martin-Diener E, Thuering N. www.aktiv-online.ch: Akzeptanz eines interaktiven Programms zur Bewegungsförderung [www.Active-online.ch: Acceptance of an interactive program promoting physical activity]. Schweiz Z Sportmed Sporttraumatol 2001;49(3):137 [FREE Full text] [WebCite Cache ID 5c6ek5fNe]
30. Glasgow RE, Nelson CC, Kearney KA, Reid R, Ritzwoller DP, Strecher VJ, et al. Reach, engagement, and retention in an Internet-based weight loss program in a multi-site randomized controlled trial. J Med Internet Res 2007;9(2):e11 [FREE Full text] [Medline: 17513282] [doi: 10.2196/jmir.9.2.e11]
31. Random Numbers Info. URL: http://www.randomnumbers.info/ [WebCite Cache ID 5c6f3EcNq]
32. Thüring N, Martin-Diener E, Martin B, Bauer G. Feasibility of a web-based randomised controlled trial for a tailored physical activity intervention. Schweiz Z Sportmed Sporttraumatol 2009;57(2):56-60.
33. Prochaska JO, DiClemente CC, Norcross JC. In search of how people change. Applications to addictive behaviors. Am Psychol 1992 Sep;47(9):1102-1114. [Medline: 1329589] [doi: 10.1037/0003-066X.47.9.1102]
34. Martin-Diener E, Thüring N, Melges T, Martin BW. The Stages of Change in three stage concepts and two modes of physical activity: a comparison of stage distributions and practical implications. Health Educ Res 2004 Aug;19(4):406-417 [FREE Full text] [Medline: 15155593] [doi: 10.1093/her/cyg051]
35. Basler HD, Jäkle C, Keller S, Baum E. Selbstwirksamkeit, Entscheidungsbalance und die Motivation zu sportlicher Aktivität [Self-efficacy, decisional balance, and the motivation to do vigorous exercise]. Zeitschrift für Differentielle und Diagnostische Psychologie 1999;20:203-216. [doi: 10.1024//0170-1789.20.3.203]
36. Marcus BH, Rossi JS, Selby VC, Niaura RS, Abrams DB. The stages and processes of exercise adoption and maintenance in a worksite sample. Health Psychol 1992;11(6):386-395. [Medline: 1286658] [doi: 10.1037/0278-6133.11.6.386]
37. Nigg CR, Riebe D. The Transtheoretical Model: research review of exercise behavior and older adults. In: Burbank P, Riebe D, editors. Promoting Exercise and Behavior Change in Older Adults: Interventions with the Transtheoretical Model. New York: Springer Publishing Company; 2002:147-180.
38. Martin B. Physical activity related attitudes, knowledge and behaviour in the Swiss population: comparison of the HEPA Surveys 2001 and 1999. Schweiz Schweiz Z Sportmed Sporttraumatol 2002;50(4):164-168 [FREE Full text]
39. 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] [doi: 10.1097/00005768-199805000-00021]
40. Abel MG, Hannon JC, Sell K, Lillie T, Conlin G, Anderson D. Validation of the Kenz Lifecorder EX and ActiGraph GT1M accelerometers for walking and running in adults. Appl Physiol Nutr Metab 2008 Dec;33(6):1155-1164. [Medline: 19088773] [doi: 10.1139/h08-103]
41. 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]
42. Strath SJ, Bassett DR, Swartz AM. Comparison of MTI accelerometer cut-points for predicting time spent in physical activity. Int J Sports Med 2003 May;24(4):298-303. [Medline: 12784173] [doi: 10.1055/s-2003-39504]
43. Tucker P, Gilliland J. The effect of season and weather on physical activity: a systematic review. Public Health 2007 Dec;121(12):909-922. [Medline: 17920646] [doi: 10.1016/j.puhe.2007.04.009]
44. Vandelanotte C, De Bourdeaudhuij I, Sallis JF, Spittaels H, Brug J. Efficacy of sequential or simultaneous interactive computer-tailored interventions for increasing physical activity and decreasing fat intake. Ann Behav Med 2005 Apr;29(2):138-146. [Medline: 15823787] [doi: 10.1207/s15324796abm2902_8]
45. Lewis B, Williams D, Dunsiger S, Sciamanna C, Whiteley J, Napolitano M, et al. User attitudes towards physical activity websites in a randomized controlled trial. Prev Med 2008 Nov;47(5):508-513. [Medline: 18773915] [doi: 10.1016/j.ypmed.2008.07.020]
46. Leslie E, Marshall AL, Owen N, Bauman A. Engagement and retention of participants in a physical activity website. Prev Med 2005 Jan;40(1):54-59. [Medline: 15530581] [doi: 10.1016/j.ypmed.2004.05.002]
47. Tate DF, Wing RR, Winett RA. Using Internet technology to deliver a behavioral weight loss program. JAMA 2001 Mar 7;285(9):1172-1177 [FREE Full text] [Medline: 11231746] [doi: 10.1001/jama.285.9.1172]
48. Woolf SH, Krist AH, Johnson RE, Wilson DB, Rothemich SF, Norman GJ, et al. A practice-sponsored Web site to help patients pursue healthy behaviors: an ACORN study. Ann Fam Med 2006;4(2):148-152 [FREE Full text] [Medline: 16569718] [doi: 10.1370/afm.522]
49. Kypri K, McAnally HM. Randomized controlled trial of a web-based primary care intervention for multiple health risk behaviors. Prev Med 2005;41(3-4):761-766. [Medline: 16120456] [doi: 10.1016/j.ypmed.2005.07.010]
50. Ware LJ, Hurling R, Bataveljic O, Fairley BW, Hurst TL, Murray P, et al. Rates and determinants of uptake and use of an internet physical activity and weight management program in office and manufacturing work sites in England: cohort study. J Med Internet Res 2008;10(4):e56 [FREE Full text] [Medline: 19117828] [doi: 10.2196/jmir.1108]
51. Cook RF, Billings DW, Hersch RK, Back AS, Hendrickson A. A field test of a web-based workplace health promotion program to improve dietary practices, reduce stress, and increase physical activity: randomized controlled trial. J Med Internet Res 2007;9(2):e17 [FREE Full text] [Medline: 17581811] [doi: 10.2196/jmir.9.2.e17]

52. Marcus BH, Ciccolo JT, Sciamanna CN. Using electronic/computer interventions to promote physical activity. Br J Sports Med 2009 Feb;43(2):102-105. [Medline: 19052143] [doi: 10.1136/bjsm.2008.053744]
53. Adams J, White M. Why don't stage-based activity promotion interventions work? Health Educ Res 2005 Apr;20(2):237-243 [FREE Full text] [Medline: 15253998] [doi: 10.1093/her/cyg105]
54. Brug J, Conner M, Harré N, Kremers S, McKellar S, Whitelaw S. The Transtheoretical Model and stages of change: a critique: observations by five commentators on the paper by Adams, J. and White, M. (2004) why don't stage-based activity promotion interventions work? Health Educ Res 2005 Apr;20(2):244-258 [FREE Full text] [Medline: 15618324] [doi: 10.1093/her/cyh005]
55. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [Medline: 15829473] [doi: 10.2196/jmir.7.1.e11]
56. Oenema A, Tan F, Brug J. Short-term efficacy of a web-based computer-tailored nutrition intervention: main effects and mediators. Ann Behav Med 2005 Feb;29(1):54-63. [Medline: 15677301] [doi: 10.1207/s15324796abm2901_8]
57. De Bourdeaudhuij I, Stevens V, Vandelanotte C, Brug J. Evaluation of an interactive computer-tailored nutrition intervention in a real-life setting. Ann Behav Med 2007 Feb;33(1):39-48. [Medline: 17291169] [doi: 10.1207/s15324796abm3301_5]

Abbreviations

BMI: body mass index
CG: control group
CI: confidence interval
FU: follow-up
HEPA: health-enhancing physical activity
IG: intervention group
OR: odds ratio
SU: spontaneous users

Edited by G Eysenbach; submitted 28.11.08; peer-reviewed by I DeBourdeaudhuij, W Brouwer; comments to author 19.01.09; revised version received 31.03.09; accepted 31.03.09; published 28.07.09.

Please cite as:
Wanner M, Martin-Diener E, Braun-Fahrländer C, Bauer G, Martin BW
Effectiveness of Active-Online, an Individually Tailored Physical Activity Intervention, in a Real-Life Setting: Randomized Controlled Trial
J Med Internet Res 2009;11(3):e23
URL: http://www.jmir.org/2009/3/e23/
doi:10.2196/jmir.1179
PMID:19666456

Original Paper

Feasibility and Effectiveness of Online Physical Activity Advice Based on a Personal Activity Monitor: Randomized Controlled Trial

Sander M Slootmaker[1], MSc; Mai J M Chinapaw[1], PhD; Albertine J Schuit[2,3], PhD; Jacob C Seidell[3], PhD; Willem Van Mechelen[1], MD, PhD

[1]EMGO Institute, Department of Public and Occupational Health, Body@Work Research Center Physical Activity, Work and Health, TNO-VU University Medical Center, Amsterdam, the Netherlands

[2]National Institute for Public Health and the Environment, Bilthoven, the Netherlands

[3]Institute of Health Science, Faculty of Earth and Life Sciences, VU University, Amsterdam, the Netherlands

Corresponding Author:
Mai J M Chinapaw, PhD
Van der Boechorststraat 7
1081 BT Amsterdam
The Netherlands
Phone: +31 20 4448203
Fax: +31 20 4448387
Email: m.chinapaw@vumc.nl

Abstract

Background: Inactive people are often not aware of the fact that they are insufficiently active. Providing insight into their actual physical activity (PA) levels may raise awareness and could, in combination with tailored PA advice, stimulate a physically active lifestyle.

Objective: This study evaluated the feasibility and effectiveness of a 3-month intervention in which Dutch office workers were provided with a personal activity monitor (PAM) coupled to simple and concise Web-based tailored PA advice (PAM COACH).

Method: Participants were randomly assigned to the 3-month PAM intervention (n = 51) or received a single written information brochure with brief general PA recommendations (n = 51). Study outcome measures were changes in PA (recall of minutes per week spent on PA, as measured by the Activity Questionnaire for Adolescents and Adults), determinants of PA, aerobic fitness, and body composition. Follow-up measurements were performed immediately after the 3-month intervention and at 8-months, 5 months after the end of the 3-month intervention period.

Results: A total of 102 workers, 23 to 39 years old, completed the baseline measurement at the worksite. 48 completed the 3-month follow up and 38 the 8-month follow-up in the intervention group, 50 completed the 3-month follow up and 42 the 8-month follow up in the control group. 35 out of 48 (73%) participants in the PAM intervention group reported wearing the PAM regularly, and the PAM COACH was used almost once a week; 24 out of 46 (52%) PAM users set a personal goal, and 33 (72%) entered their favorite activities on the website. Main reasons for not using these items were lack of interest or not being able to find the item on the website. The majority of PAM users (34 out of 46, 74%) read the advice, of whom 14 (39%) found it unappealing. After the 3-month intervention, no significant intervention effect was observed (adjusted difference in min/week) for sedentary behavior (β = 10, 95% CI = −435 to 455), light-intensity PA (β = −129, 95% CI = −337 to 79), moderate-intensity PA (β = −13, 95% CI = −89 to 63), vigorous-intensity PA (β= −6, 95% CI = −75 to 62), and moderate- to vigorous-intensity PA (β = −23, 95% CI = −121 to 76). No significant intervention effect was observed in the PA outcomes at the 8-month follow-up. For the determinants of PA, aerobic fitness, and body composition, no statistically significant intervention effect was observed in the total study population immediately after the 3-month intervention or the 8-month follow-up.

Conclusions: The intervention appeared to be easily applicable to real-life settings. The intervention was ineffective in improving PA behavior or its determinants in healthy office workers. More attention should have been given to the quality and appropriateness of the tailored advice.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN): 93896459; http://www.controlled-trials.com/ISRCTN93896459/ (Archived by WebCite at http://www.webcitation.org/5iR3mf7ex)

(J Med Internet Res 2009;11(3):e27) doi:10.2196/jmir.1139

Introduction

According to the current Dutch Public Health physical activity (PA) recommendation, [1] adults should accumulate at least 30 minutes of moderate-intensity PA on at least 5 days of the week, or a minimum of 20 minutes of vigorous-intensity aerobic PA on 3 days of the week to promote and maintain health. Surveillance data from the Netherlands [2] and the United States [3] have shown consistent but low adherence (55% and 45%, respectively) to this recommendation among adults in general but have also shown large differences among subpopulations [4,5]. It is suggested that adults in full-time employment or those going through life events such as marriage and having children are more at risk of becoming physically inactive due to increased commitments [6].

Inactive subjects are often not aware of the fact that they are insufficiently active. This was recently shown in a survey of 2600 Dutch adults [7]. No less than 60% of the people who did not meet the recommendation believed that they were sufficiently active. The use of a PA monitor (eg, pedometer, accelerometer) that continuously registers and displays the actual PA level of the user may raise awareness and could thus overcome the problem of poor self-evaluation [8]. Hence, objective instant feedback by a PA monitor could positively affect the PA level in inactive subjects [8,9]. Hultquist et al [10] found that sedentary women who were given pedometers and who were instructed to walk 10,000 steps a day walked almost 2000 steps per day more than women who were instructed to go for a brisk 30-minute walk each day. PA monitors can be worn without major inconvenience, require little effort of the user, and are compatible with most daily activities, making them a practical and socially acceptable measure of PA [11].

Internet-based self-management interventions for PA have been shown to have potential [12-15] because they can reach large numbers of at-risk participants in a variety of settings at any time and location. There is evidence that health-related behavior is more affected by a tailored approach than by general health promotion activities [16,17]. Computer-tailored PA promotion programs are relatively new [18]. They provide respondents with individually adapted feedback about their current PA level and additionally provide individualized suggestions to change sedentary behavior and to promote daily PA. To date, little evidence is available on the feasibility and effectiveness of Internet-based tailored interventions coupled to an activity monitor [19,20].

The PAM concept (PAM BV, Doorwerth, the Netherlands) combines the use of a personal activity monitor (PAM) with simple and concise Web-based tailored PA advice (PAM COACH). The PAM (model AM101, PAM BV, Doorwerth, the Netherlands) is an uni-axial accelerometer in the vertical direction that can be easily attached to a belt. The validity of the PAM accelerometer has been tested in a laboratory setting and has shown results similar to the MTI Actigraph for estimating energy expenditure in walking and stair walking [21]. The PAM produces a single index score that accumulates during the day and is a proxy measure of total daily PA. The PAM shows the PAM score continuously on its display. Via a docking station, which must be connected to a computer with an Internet connection, the user can upload his or her personal PAM scores through PAM software to the PAM COACH website at any time throughout the day. On the PAM COACH website, users can interactively plan and evaluate their own activity advice based on their actual PAM scores and their PA goals and preferences.

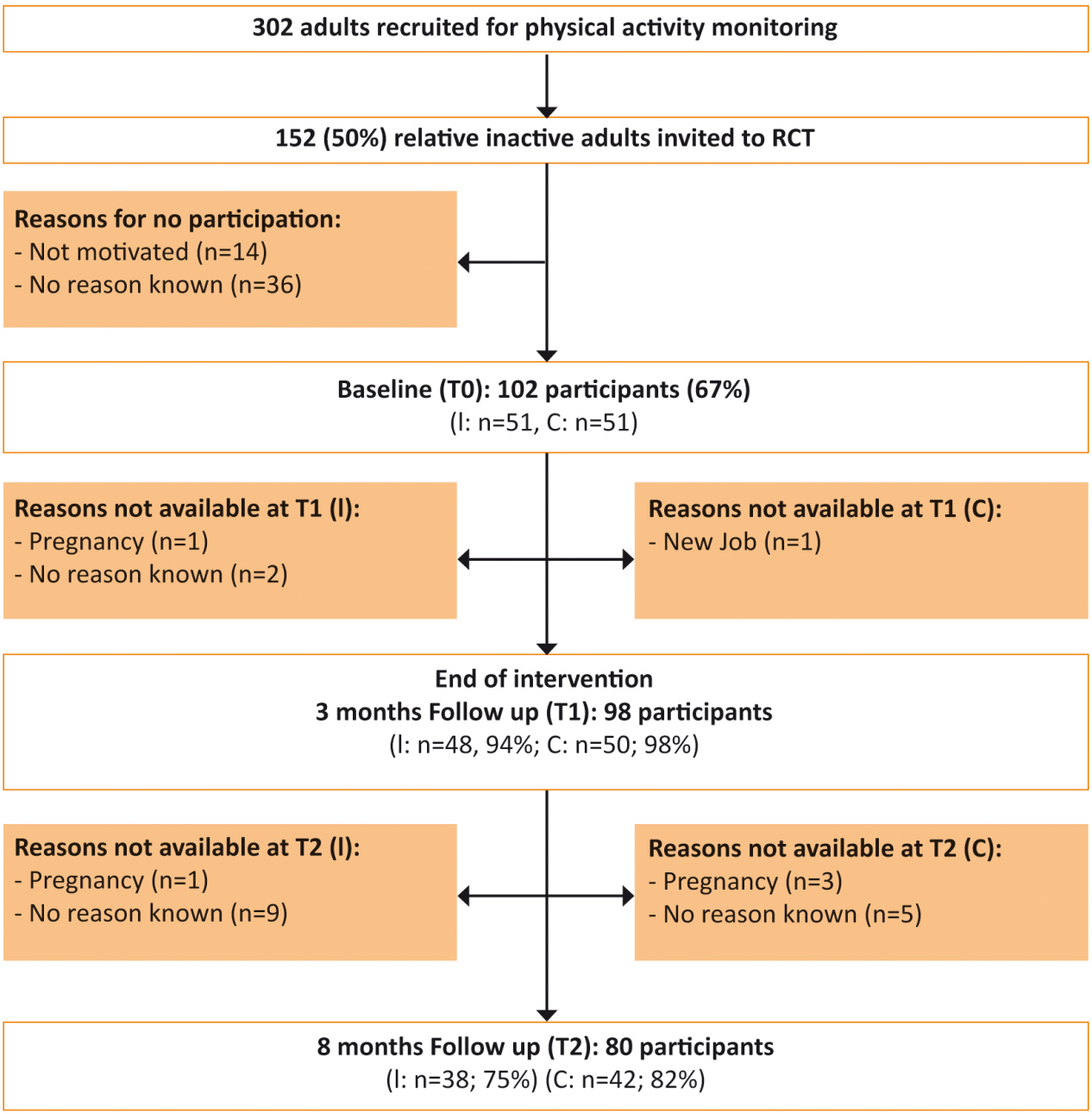
The objective of this study was to evaluate the feasibility and effectiveness of providing a PAM coupled to simple and concise Web-based personalized PA advice on the daily PA level of young Dutch inactive office workers in a randomized controlled trial (RCT). In addition, the effects on determinants of PA, aerobic fitness, and body composition were examined. We hypothesized that the use of a PAM combined with individually tailored PA advice would increase awareness and subsequent PA levels of inactive office workers.

Methods

Study Design and Population

This RCT is part of the PAM project, which is described extensively elsewhere [22]. Mainly office workers from 20 to 40 years old, all apparently healthy, with differing levels of education were recruited from eight worksites in the surrounding areas of Amsterdam, the Netherlands. For all worksites, the same recruitment protocol was used. In the recruitment procedure, we informed the participants about the beneficial health outcomes of regular PA (ie, increased cardiovascular health, and reduction of the risk of overweight, type 2 diabetes, depression, and some types of cancer) by providing a brochure and through individual communication. Inclusion criteria were ability to walk without aid, Dutch speaking, and not being pregnant. First, PA levels were monitored for 2 weeks by means of a PAM and a PA questionnaire. Based on these 2 weeks, the study population (N = 302) was divided into “active” (most active 50% of the population) and “inactive” (least active 50% of the population) groups. The inactive adults were invited to participate in the RCT and were told that they were eligible because of their low level of PA. To be able to detect a between-group difference of 20% in PA level (80% probability and a significance level of .05), two groups of 50 participants were required. Of the 152 invited adults, 102 (67% response rate) completed the baseline measurement and were randomly assigned to either the intervention (n = 51) or the control group (n = 51). Randomization to the intervention or control group was performed at the individual level by choosing sealed envelopes after the baseline measurements. The flow of subjects through the RCT and the distribution of nonresponders are shown in Figure 1. This study was approved by the Medical Ethics Committee of VU University Medical Center and was conducted between September 2004 and November 2005. All participants gave their informed consent.

Figure 1. Flowchart of the intervention (I) and control (C) subjects in the RCT. Note that for analysis of the primary outcome measure at t1, 2 out of 48 (4%) and 1 out of 50 (2%) participants in the intervention and control group respectively, were excluded for analysis due to impossible values for physical activity (not shown in figure, see text under "Primary Outcome Measure").



Intervention

After randomization, participants in both the intervention and control groups were advised to increase their PA levels. The control group received a single written information brochure with brief general PA recommendations. This print brochure is published by the Netherlands Heart Foundation and contains brief general information on the health benefits of PA and the PA recommendation. Everyone can obtain the brochure free of charge.

The intervention group received the PAM and was provided with Web-based tailored PA advice (PAM COACH) [23] for a 3-month period. Figure 2 and Table 1 show the sitemap of the PAM COACH website.

Figure 2. Functionalities of the PAM COACH website

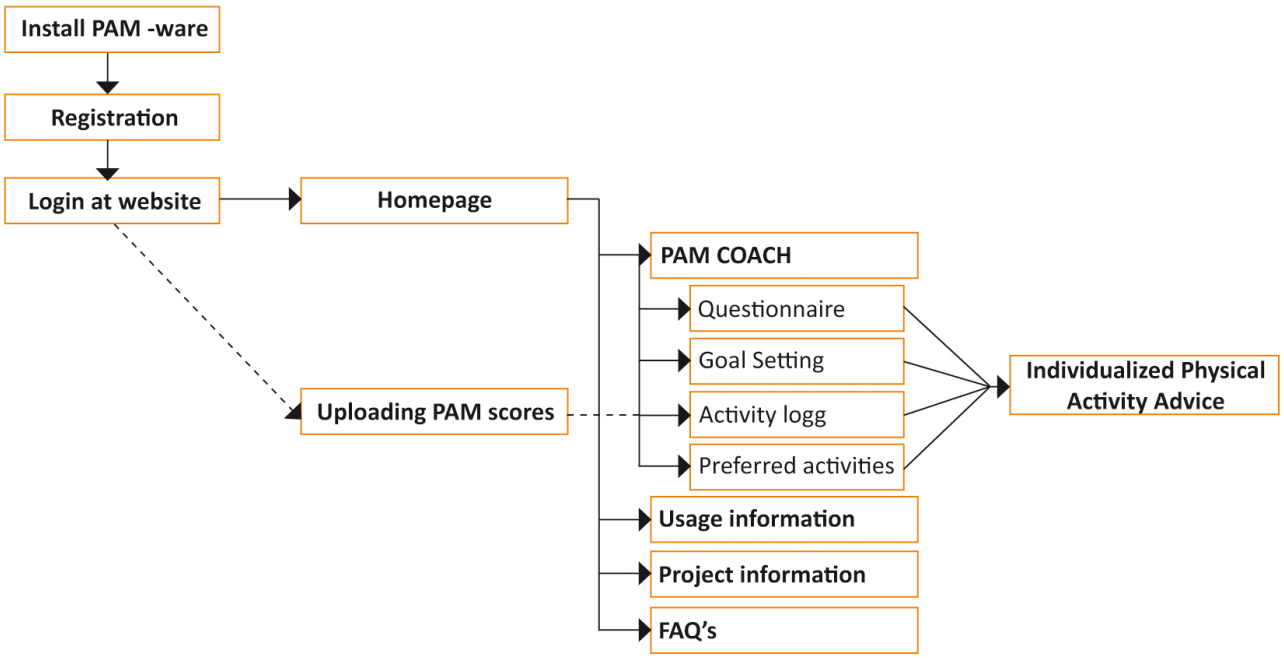
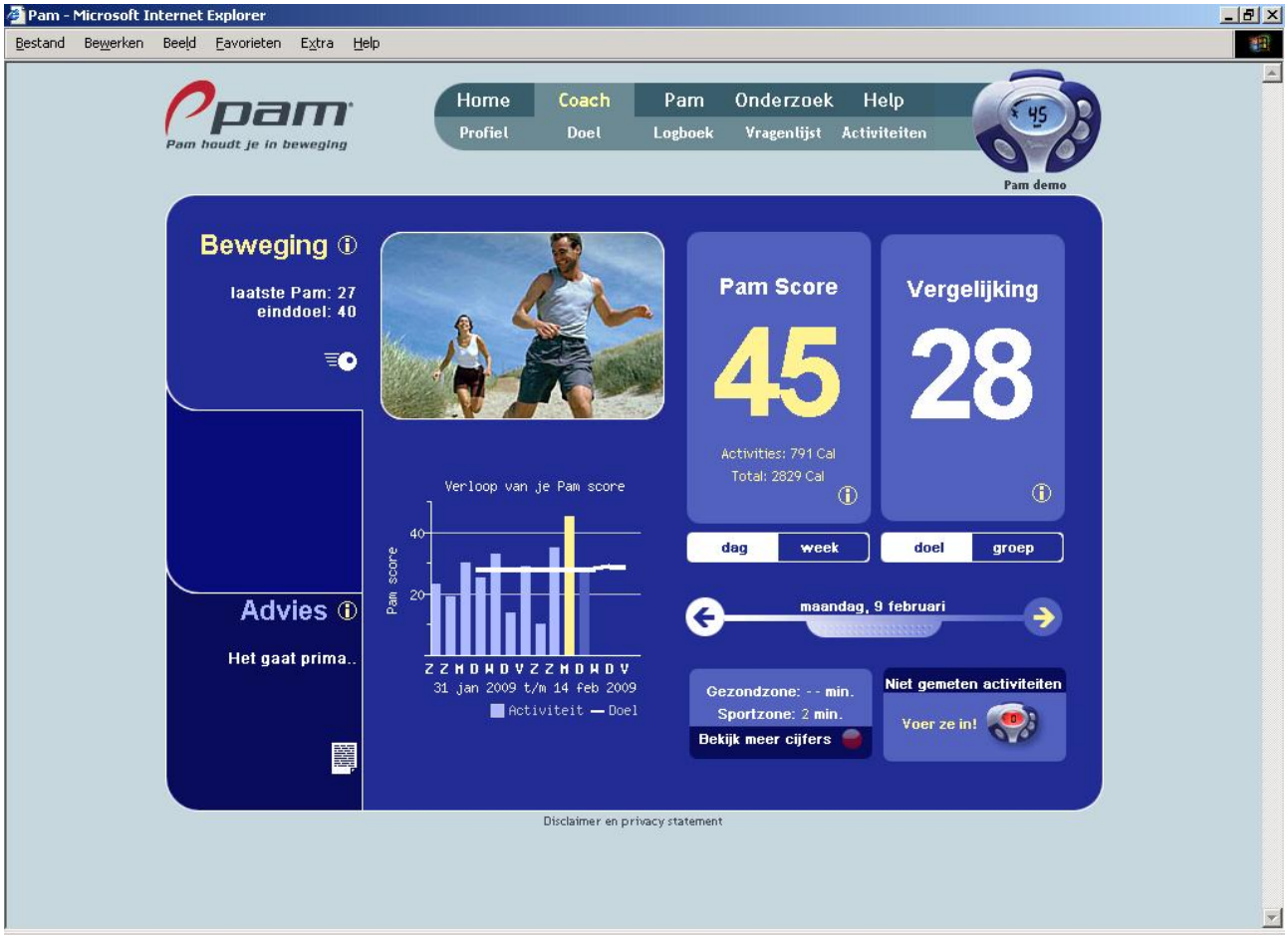


Table 1. Contents of the PAM COACH website

Website Section	Description
Home page	Presentation of the latest PAM week score, hyperlink to complete advice and motivational PA tips.
Goal setting	Setting the PAM goal score, indicated by the deficiency in minutes per day for their preferred activities.
Activity log	Presentation of all uploaded PAM scores (per day/week/month).
Questionnaire	Twelve questions on perceived physical activity barriers (yes or no).
Preferred physical activities	Categories: transport, school activities, in and around the house, individual and team sports.
Individualized PA advice	Translation of PAM goal score in the deficiency of minutes per day for their preferred physical activities. Feedback on their answers to the questionnaire. Stimulating feedback. Comparison of users' PAM score to their peers in the intervention group.
Usage information	Information about the use of the PAM and PAM COACH website, including a demonstration.
Project information	The aim of the project and contact information.
FAQs	Answers to frequently asked questions about the (use of) PAM and PAM COACH website.

First, the participants had to install the PAM software on their computer in order to use the PAM reader. When reading the PAM, the participant is automatically directed to the PAM COACH website. The user can register on the PAM COACH website by filling out a form with personal data (ie, username and password) and answering 12 questions on perceived PA barriers. Upon entry to the PAM COACH website, the user formulates a PAM goal score for the 3-month intervention period. Based on the user's uploaded PAM score for the first week, the PAM COACH assigns a lower goal that increases daily until the PAM goal score is reached at the end of the intervention period. The PAM goal score can be changed by the user throughout the intervention. On every subsequent log-in, the PAM COACH website presents all the uploaded PAM scores and coupled PAM goals in orderly graphs per week or month (Figure 3).

Figure 3. Screenshot of the PAM COACH website



The uploaded PAM scores are automatically accompanied by tailored PA advice on the computer screen as well as motivational tips for increasing PA (n = 21). The advice includes information on how to reach the PAM goal, which is based on (1) the user's preferred activities (eg, an extra 60 minutes walking, or 25 minutes running, or 20 minutes playing squash daily), and (2) user-perceived PA barriers. Furthermore, the feedback is given in a stimulating way (eg, "You are doing better, but to reach your goal you have to do more"). If the PAM is not worn during certain activities (eg, swimming), these minutes can be added manually to the PAM score on the website. Apart from the short feedback on the PAM COACH website, the users can easily monitor their progress in daily PA by reading their PAM score directly on the display of the PAM. The participants received written and verbal instructions and practical demonstrations on how to wear the PAM and how to use the PAM COACH website (ie, setting a personal goal and favorite activities). Participants were instructed to register and upload PAM data in the first week of the intervention, to check if the system worked properly. After that, the participants could make use of the PAM and PAM COACH website as much as they wanted. At least one computer with PAM software and Internet access was available at all worksites except one, where ICT policies did not allow PAM software on the network. Participants from this worksite accessed the PAM COACH website at home only.

Measurements

All measurements took place at the worksite during working hours, at baseline, immediately after the 3-month intervention, and at an 8-month follow-up, which corresponds to 5 months after the end of the 3-months intervention period. The measurements are described in detail elsewhere [22]. Gender, age, and education level were obtained at baseline. Education level was categorized into low and high. High education level comprised higher vocational education or a university degree. All other levels of education were defined as low.

Primary Outcome Measure

The Activity Questionnaire for Adolescents and Adults (AQuAA) was used to assess the amount of minutes per week spent on light-intensity (2-4 metabolic equivalents, METs), moderate-intensity (4-6.5 METs), and vigorous-intensity (> 6.5 METs) PA, as well as time spent sedentary (< 2 METs), such as watching TV and using the computer. The AQuAA refers to activities in the past week (7-day recall). Participants filled in the questionnaire while supervised by a research assistant. The research assistant checked the questionnaires when they were returned. Based on the assumption that one sleeps 8 hours per day, sixteen hours (960 minutes) was considered the maximum amount of time per day a person can spend on PA. At 3-months follow up, 2 out of 48 (4%) and 1 out of 50 (2%) participants in the intervention and control group respectively, were excluded because they exceeded this maximum time.

Secondary Outcome Measures

Determinants of Physical Activity

A short questionnaire was developed to assess behavioral intention to participate in sports more often. Intention to exercise was assessed by a single question: "Do you intend to participate more frequently in sports over the next three months?" For the determinants of attitude, social influences, and self-efficacy expectations and personal barriers toward sport (Cronbach alpha: .49, .85, and .78, respectively), a selection of two or three relevant questions was made, based on previous studies [24-27]. Answering formats were 5-point Likert scales (very low to very high). Per determinant, multiple items were converted into summary scores.

Awareness of complying with the Public Health PA recommendations was assessed by self-reported answers to the following questions: "On how many days of the week did you spent at least 30 minutes of moderate activity" and "Do you think you spend enough time participating in sports?" (yes or no). According to a method described by Ronda et al [7], respondents were allocated to four categories of awareness (underestimators, overestimators, realists adequate, or realists inadequate) based on their self-rated compliance with the PA recommendations and the results of the PA questionnaire. In the analyses, awareness was dichotomized in nonrealists (underestimators and overestimators) and realists (realists adequate and realists inadequate). Participants were classified as complying with the recommendation when they reported at least 150 minutes of moderate- to vigorous-intensity activity per week. Subjects' knowledge of the PA recommendation was tested by the question "How much time per day do you have to spend on PA to stay healthy?"

Aerobic Fitness

A submaximal test, the Chester Step Test (CST) [28], was used to predict maximal aerobic capacity. The CST consists of five increasing paces of stepping on and off a bench. A step height of 30 cm was used for active participants and 20 cm for inactive participants. The CST starts at the relatively slow pace of 15 steps per minute and increases every 2 minutes to 20, 25, 30, and 35 steps per minute. Throughout the test, the heart rate (HR) is monitored. After each stage, the subject is asked to rate his or her perceived exertion on the Borg scale, which is a 15-point numerical rating scale ranging from 6 (very very light) to 20 (exhaustion). The test is terminated when the subject's HR reaches 80% of the age-estimated maximal HR (ie, 220 minus age) or when the subject rates 14 on the Borg scale. Prior to this test, the subjects were screened with the Physical Activity Readiness Questionnaire (PAR-Q) [29]. Based on this screening, five subjects were excluded. Maximal aerobic capacity (VO_2max) was predicted by the provided CST calculator (ASSIST Creative Resources Limited, Wrexham, England). This is based on the extrapolation of a line of best fit, which passes through the submaximal HR responses for each stepping stage up to a level that equals the participant's age-estimated maximal HR.

Body Composition

Standard procedures were used to measure body weight, body height, waist and hip circumference, and the thickness of four skin folds (biceps, triceps, subscapular, and suprailiac). Body weight was measured in light clothing without shoes. Body mass index (BMI) was calculated by dividing the weight (kg) by height squared (m^2). Percentage body fat was estimated from the sum of the four skin fold thickness measurements according to an age- and gender-specific method by Durnin and Womersley [30]. Before the baseline measurement, intrarater and interrater reliability for all four skin folds were determined. Intrarater reliability and interrater reliability (intraclass correlation coefficient, ICC) varied between .83 and .98.

Process Measures

After the 3-month intervention, PAM users were asked to evaluate (1) the PAM (ie, appreciation of the PAM score, frequency of wearing the PAM), (2) the PAM COACH website (ie, appreciation of the website, use of activity goal, and favorite activities), and (3) the tailored advice (ie, reading and appreciation of the advice). The uploaded PAM scores and the log-in frequency to the PAM COACH website were registered for each participant in the intervention group.

Data Analysis

To compare baseline values, the chi-square test was used for gender, education, and awareness distributions. Nonparametric testing (Mann-Whitney U test) was used for PA data. Independent samples t-test was used to analyze all other demographic variables, determinants of PA, aerobic fitness, and body composition. The effect of the intervention was estimated based on the intention-to-treat principle, including all participants who had attended at least one follow-up measurement. Logistic regression analysis was used for the dichotomous outcome measure, awareness of meeting the PA, and sports recommendation (0 = nonrealists, 1 = realists). For all other outcome measures, standard linear regression analysis was used to test the differences between intervention and control groups at follow-up. The follow-up measurements were defined as dependent variables. Baseline values of the particular dependent variable were always included as covariates. The parameters of interest are the regression coefficients (β), indicating the effect of the intervention of interest compared to the control group. In a second analysis, gender (0 = male, 1 = female), age (continuous), education (0 = low, 1 = high), and BMI (continuous; not added in the analyses for body composition outcome measures) were considered as possible confounders or effect modifiers. If the interaction between group allocation and the variable concerned had a p-value below .10, then subgroup analyses were performed. Additionally, we adjusted for program adherence on the outcomes by performing regression analyses among intervention participants, including the log-in frequency. Analyses were performed using SPSS version 14.0 (SPSS Inc, Chicago, IL, USA).

Results

Study Population and Baseline Measurements

Table 2 shows baseline characteristics of both the intervention and control groups; 61 out of 102 participants were female (60%), and 66 out of 102 (65%) had a high level of education. Participants in the intervention and control groups were comparable except for age (intervention group was, on average, 1.3 years older; P = .05) and intent to participate in sports at baseline (higher in control group; P = .01). The majority of participants (68 out of 102, 67%) met the PA recommendation at baseline, with 78% (n = 32) of the men and 59% (n = 36) of the women meeting the recommendation.

Table 2. Baseline characteristics of total sample, PAM intervention group, and control group

Characteristic	Total (n = 102)	PAM (n = 51)	Control (n = 51)
Demographics			
Mean age in years ± SD	31.8 ± 3.5	32.5 ± 3.4	31.2 ± 3.5[a]
Female (%)	60	61	59
Highly educated (%)	65	63	67
Familiar with PA recommendations (%)	63	59	67
Compliance with PA recommendations (%)	67	69	65
Determinants of participating in sports (mean ± SD)[b]			
Attitude	4.25 ± 0.69	4.30 ± 0.75	4.25 ± 0.64
Self-efficacy	3.34 ± 0.71	3.22 ± 0.70	3.45 ± 0.71
Intention	3.44 ± 1.23	3.10 ± 1.20	3.78 ± 1.19[a]
Awareness of compliance with recommendations (%)			
Realist inadequate	23	23	23
Underestimator	27	27	27
Overestimator	10	11	9
Realist adequate	40	39	41
Awareness of sports participation (%)			
Realist inadequate	62	59	65
Underestimator	3	6	0
Overestimator	27	31	23
Realist adequate	8	4	12
Aerobic fitness (mean ± SD)			
VO_2max (mL O_2/kg/min)	41.4 ± 7.5	41.7 ± 8.4	41.2 ± 6.7
Body composition (mean ± SD)			
Weight (kg)	77.7 ± 14.6	79.0 ± 15.6	76.5 ± 13.6
BMI (kg/m^2)	25.2 ± 4.1	25.9 ± 4.5	24.4 ± 3.5
Sum of skin folds (mm)	65.3 ± 31.6	69.4 ± 36.2	61.2 ± 25.8
Body fat (%)	27.1 ± 7.6	27.9 ± 8.0	26.4 ± 7.2
Waist circumference (cm)	85.4 ± 11.6	86.4 ± 11.9	84.5 ± 11.4

[a]Difference at baseline between intervention and control group (P< .05).

[b]Assessed on a 5-point Likert scale.

Primary Outcome Measure

In the total study sample, our 3-month intervention did not significantly affect PA levels (Table 3). However, because of effect modification by education, we conducted subgroup analyses. These analyses showed that the 3-month intervention resulted in a relative lowering of light-intensity PA (2-4 METs) among higher-educated participants (adjusted difference between intervention and control group in min/week, β = −349, 95% CI = −632 to −66, P = .02). This effect was not sustained at the 8-month follow-up, 5 months after the intervention. A higher

adherence to the program did not result in increased levels of PA (data not shown).

Table 3. Median PA scores and mean difference in PA and sedentary time between PAM intervention group and control group at baseline (n=51 in each group), 3 months (control: n=49; intervention n=46), and 8 months (control: n=42; intervention n=38)

Outcome Measure (min/week)	PAM	Control	Crude Difference[b]	Adjusted Difference[c]
	Median (IQR[a])	Median (IQR[a])	β (95% CI)	β (95% CI)
Sedentary time				
Baseline	3390 (2580; 3810)	3375 (2870; 3855)	–	–
3 months	3400 (2850; 3840)	2470 (2495; 3941)	101 (−338; 540)	10 (−435; 455)
8 months	2925 (2358; 4206)	3342 (2741; 3998)	−174 (−721; 374)	−267 (−803; 268)
Light-intensity PA				
Baseline	630 (480; 1320)	720 (450; 1220)	–	–
3 months	636 (345; 950)	678 (408; 1320)	−84 (−290.9; 123.3)	−129 (−337; 79)
8 months	500 (326; 994)	593 (323; 1020)	−18 (−220.6; 185.1)	−2.0 (−210; 206)
Moderate-intensity PA				
Baseline	90 (5; 240)	120 (10; 203)	–	–
3 months	75 (20; 180)	90 (8; 240)	−22 (−96; 53)	−13.0 (−89; 63)
8 months	120 (19; 241)	90 (8; 278)	97 (−47; 241)	103 (−42; 248)
Vigorous-intensity PA				
Baseline	170 (60; 315)	120 (30; 240)	–	–
3 months	80 (0; 210)	113 (41; 290)	−4 (−71; 63)	−6 (−75; 62)
8 months	120 (30;259)	115 (30; 303)	−17 (−97; 62)	−28 (−110; 54)
Moderate- to vigorous-intensity PA				
Baseline	320 (120; 510)	240 (75; 443)	–	–
3 months	197 (100; 480)	281 (150; 488)	−27 (−123; 68)	−23 (−121; 76)
8 months	223 (150; 548)	263 (143; 420)	81 (−109; 272)	74 (−119; 267)

[a]Interquartile range between 25th and 75th quartile.

[b]Baseline values of the particular dependent variable were always included as covariate.

[c]Adjusted for gender, age, education, and BMI at baseline.

Secondary Outcome Measures

Determinants of Physical Activity

For the determinants of PA, no statistically significant intervention effect was observed in the total study sample (Table 4); however, an intervention effect was observed in subgroups of BMI. The proportion of subjects being aware of their adherence to the sports recommendation increased among overweight participants in the intervention group. The adjusted odds ratio (OR) between intervention (n = 16) and control group (n = 21) was 16.4 (95% CI = 1.3 to 214, P = .02). This significant effect was not sustained at the 5-month follow-up after the intervention.

Aerobic Fitness

No statistically significant intervention effect was observed on aerobic fitness.

Body Composition

No statistically significant intervention effect was observed on body composition in the total study population. However, subgroup analyses showed a decrease in body weight among low-educated intervention participants compared to their peers in the control group (adjusted difference, β = −1.6 kg, 95% CI = −2.8 to −0.4, P = .01). This difference was still observable at the 8-month follow-up, 5 months after the intervention (adjusted difference, β = −2.1 kg, 95% CI = −4.4 to 0.3, P = .08).

Table 4. Effectiveness of the 3-month PAM intervention on determinants of PA, aerobic fitness, and body composition: results of regression analyses

Outcome Measure	Crude Difference[a]	Adjusted Difference[b]
	β (95% CI)	β (95% CI)
Determinants of playing sports (5-point Likert scale)		
Attitude	−0.18 (−0.40; 0.04)	−0.20 (−0.43; 0.02)
Social influence	−0.01 (−0.31; 0.28)	−0.05 (−0.36; 0.24)
Self-efficacy	0.64 (−0.14; 0.27)	0.05 (−0.15; 0.26)
Intention	0.27 (−0.20; 0.74)	0.30 (−0.20; 0.80)
Aerobic fitness		
VO_2max (mL O_2/kg/min)	1.28 (−1.34; 3.90)	1.82 (−0.73; 4.39)
Body composition		
Weight (kg)	−0.27 (−1.12; 0.57)	−0.36 (−1.23; 0.49)
Sum of skin folds (mm)	1.49 (−4.38; 7.38)	1.34 (−4.62; 7.30)
Waist circumference (cm)	−0.51 (−1.85; 0.82)	−0.73 (−2.10; 0.63)
	OR (95% CI)	OR (95% CI)
Awareness (%)[c]		
Compliance with PA recommendations	1.45 (0.62; 3.37)	1.33 (0.54; 3.27)
Sports participation	0.81 (0.31; 2.11)	0.62 (0.22; 1.75)

[a]Baseline values of the particular dependent variable were always included as covariate.

[b]Adjusted for gender, age, education, and BMI at baseline. BMI at baseline was not added as confounder in the analyses for the body composition outcome measures.

[c]Awareness was analyzed with logistic regression (nonrealists = 0, realists = 1).

Process Measures

Of the PAM users, 35 out of 48 (73%) reported to have worn the PAM “regularly” to “often” (Table 5). This finding was supported by the log-in frequency (almost once a week) of the PAM data to the PAM COACH website. Just over half of the PAM users (24 out of 46, 52%) set a personal goal, and 33 users (72%) entered their favorite activities on the website. Main reasons for not using these items were lack of interest and not being able to find them on the website. The tailored advice was read by 34 out of 46 (74%) PAM users, of whom 14 did not find the advice appealing. Main reasons were as follows: the advice was not personal or specific enough (n = 9), the advice was not applicable to their daily situation (n = 6), little variety in the advice (n = 3). Overall, the participants rated the PAM and the PAM COACH website as sufficient.

Table 5. Process evaluation data of the PAM accelerometer and the PAM COACH website

Variable	No.	Mean ± SD
Log-in frequency to the PAM COACH website		0.9 ± 0.6
1st month of intervention	47	3.8 ± 2.5
2nd month of intervention	47	3.6 ± 2.6
3rd month of intervention	47	3.4 ± 3.6
Mean uploaded PAM score to the website	26	
1st month of intervention		18.4 ± 7.8
2nd month of intervention		16.7 ± 7.5
3rd month of intervention		17.8 ± 7.6
Appreciation of PAM score[a]	47	6.4 ± 2.1
Appreciation of PAM COACH website[a]	47	6.5 ± 1.9
	No.	%[b]
Wore the PAM accelerometer	48	
Never		2
Hardly ever		10
Sometimes		15
Regularly		38
Often		35
Set personal PAM goal on website	46	52
Entered favorite activities on website	46	72
Read personalized advice on website	46	74
Found advice on website appealing	36	39

[a]On a scale of 1 (very negative) to 10 (very positive).

[b]Percentages are based on self-report.

Discussion

This study investigated the feasibility and effectiveness of providing a PAM in combination with simple and concise tailored PA advice delivered through the Internet. The primary aim of the intervention was to improve daily PA, but we also examined effects on secondary outcomes such as determinants of PA, aerobic fitness, and body composition. According to the intention-to-treat analysis, the PAM intervention did not result in increased PA levels of young Dutch office workers, nor did it improve any of the secondary outcomes. These results may partly be due to the fact that only 39% (n = 14) of the users found their PA advice appealing.

The use of a personal website seemed to be applicable at every worksite with an Internet connection as well as being a suitable mode of conducting PA interventions among young employees. Yet, our intervention seemed ineffective at promoting PA in the total study population. This is in contrast to previous controlled interventions [10,19,20,31-33], which showed that a PA monitor helped sedentary participants to set goals and motivated them to increase their PA. However, these studies were not designed as RCTs and included mainly overweight participants or patients with type 2 diabetes.

Although our study was not designed for subgroup analyses, we conducted them after observing significant effect modification. Among low-educated intervention subjects, we observed a decrease in body weight of 1.6 kg. This effect is considerable after 3 months and is clinically relevant. Moreover, the proportion of realists increased among overweight subjects, which makes this concept of self-monitoring and Web-based feedback interesting for future research in these specific target groups.

Limitations

The results of this study must be interpreted in light of its limitations. First, our primary outcome and some of our secondary outcomes were based on self-report and therefore prone to misreporting. However, since we looked at changes in PA behavior, at least the bias associated with systematic errors is cancelled out. Nevertheless, we compared the self-reported total PA data with the objective uploaded PAM data among participants of the intervention group (data not shown). The PAM data confirm the decline in total PA as assessed by the self-report (change in median min/week: −147) after the 3-month intervention.

Second, our control group is not a truly non-intervention group because they received an information leaflet on PA. However,

we do not expect changes in PA by providing such brochures only. Furthermore, our findings may reflect ceiling effects associated with a relatively active sample at baseline. Although the study was aimed at inactive employees, 35 out of 51 (69%) participants in the PAM intervention group and 33 out of 51 (65%) participants in the control group already met the PA recommendation at baseline. Even though this information was based on self-report, it seems likely that the participants selected for the RCT were in general more active and health conscious than the general Dutch population. This is supported by the facts that the percentage of subjects who were acquainted with the PA recommendation was high in both groups, and the percentage of subjects who were aware (realists) of their compliance with the PA recommendation and their regular participation in sports was high in both groups. The fact that we did not collect information about the participants' willingness to become physically active can be considered as a limitation. In addition, our study results are mainly applicable to people who are employed at a workplace that allows personal Internet use, which limits the generalizability of our study.

Finally, the practical advice given on the website was partly based on the objectively monitored PAM score. Accelerometers are insensitive to certain types of movements, in particular, nonambulatory physical activities with arm and or limb movements, such as cycling and weightlifting. This limitation of the accelerometer may have reduced the accuracy and relevance of the advice given by the PAM COACH website, particularly for subjects who cycle a lot, which is common in the Netherlands. Although activities that are not accurately measured by the PAM can be included manually on the PAM COACH website, a study has shown that recipients of negative or unexpected feedback responded by doubting the accuracy and credibility of the feedback information [34]. This phenomenon may have discouraged our participants from achieving their personal PA goal.

Strengths

Strengths of the study are its design, the easy to implement intervention, and the low dropout rate. This RCT was set up as a short-term minimal intervention strategy in order to make it easily applicable in real-life settings. During the intervention, PAM users received short personalized PA advice together with supportive practical advice to reach their personal PA goal. In order to reach this PA goal, the activity preferences of the user were taken into account so that PA could be more easily implemented into their daily life. After registering, the user could decide when and how often to log in to the PAM COACH website. In spite of the minimal contact during the intervention between the researcher and the participant, adherence was moderate to high. PAM users logged in 10 times on average during the 3-month RCT, which is almost once a week and is comparable with frequencies of website log-ins from previous studies (range 0.7 to 1.5 times per week) [35,36]. Moreover, during the intervention we observed a low dropout, only 3 (6%) and 1 (2%) out of the 51 participants for the intervention and control group, respectively. The recurrent visits and low dropout during the intervention suggest that participants were interested in new information and were acquainted with the technology.

The appreciation of the intervention materials differed largely among participants in the intervention group; most participants expected more varied and concrete advice and found the advice not applicable to their daily life. This occurred in spite of our aim to tailor the PA advice to a certain extent based on the users' actual PAM score in relation to their PAM goal. We strived for simple and concise PA advice and a variety of motivational tips on the PAM COACH website.

Conclusions

To conclude, we hypothesized that the combination of wearing a PAM combined with tailored PA advice delivered through the Internet would be potentially successful in increasing awareness of personal activity levels and actual PA levels. However, we did not observe any significant effect on awareness, PA level, determinants of PA, aerobic fitness, or body composition among the total group of young healthy Dutch employees. This may be explained by the fact that we conducted a minimal intervention in a study population that largely (67%) met the PA recommendations at baseline. Moreover, a large part of the intervention population did not find the advice appealing. Hence, the results of the present study do not give cause for wider implementation of this minimal intervention among healthy adults. Since we observed a tendency for a positive intervention effect on body weight among low-educated adults, more research may be necessary to investigate the effectiveness of this type of intervention among people who are overweight or of low socioeconomic status. In this, attention should be given to the quality and appropriateness of the tailored advice.

Acknowledgments

This study was funded by a grant from The Netherlands Organization for Health Research and Development (ZonMw; Grant number: 2100.0093).

Conflicts of Interest

None declared.

References

1. 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. [Medline: 17762377] [doi: 10.1249/mss.0b013e3180616b27]

2. Statistics Netherlands (CBS). Physical activity according to age specific guideline URL: http://statline.cbs.nl [accessed 2008-07-24] [WebCite Cache ID 5eAV1DILV]
3. ; Centers for Disease Control and Prevention (CDC). Adult participation in recommended levels of physical activity--United States, 2001 and 2003. MMWR Morb Mortal Wkly Rep 2005 Dec 2;54(47):1208-1212 [FREE Full text] [Medline: 16319815]
4. King AC, Blair SN, Bild DE, Dishman RK, Dubbert PM, Marcus BH, et al. Determinants of physical activity and interventions in adults. Med Sci Sports Exerc 1992;24(Suppl):S221-S236. [Medline: 1625548]
5. Owen N, Bauman A. The descriptive epidemiology of a sedentary lifestyle in adult Australians. Int J Epidemiol 1992 Apr;21(2):305-310. [Medline: 1428485] [doi: 10.1093/ije/21.2.305]
6. Brown WJ, Trost SG. Life transitions and changing physical activity patterns in young women. Am J Prev Med 2003 Aug;25(2):140-143. [Medline: 12880882] [doi: 10.1016/S0749-3797(03)00119-3]
7. Ronda G, Van Assema P, Brug J. Stages of change, psychological factors and awareness of physical activity levels in The Netherlands. Health Promot Int 2001 Dec;16(4):305-314 [FREE Full text] [Medline: 11733449] [doi: 10.1093/heapro/16.4.305]
8. Rooney B, Smalley K, Larson J, Havens S. Is knowing enough? Increasing physical activity by wearing a pedometer. WMJ 2003;102(4):31-36. [Medline: 12967019]
9. Paschali AA, Goodrick GK, Kalantzi-Azizi A, Papadatou D, Balasubramanyam A. Accelerometer feedback to promote physical activity in adults with type 2 diabetes: a pilot study. Percept Mot Skills 2005 Feb;100(1):61-68. [Medline: 15773694] [doi: 10.2466/PMS.100.1.61-68]
10. Hultquist CN, Albright C, Thompson DL. Comparison of walking recommendations in previously inactive women. Med Sci Sports Exerc 2005 Apr;37(4):676-683. [Medline: 15809569] [doi: 10.1249/01.MSS.0000158993.39760.1B]
11. 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] [doi: 10.1249/01.mss.0000185657.86065.98]
12. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [Medline: 17572313] [doi: 10.1016/j.amepre.2007.02.041]
13. Norman GJ, Zabinski MF, Adams MA, Rosenberg DE, Yaroch AL, Atienza AA. A review of eHealth interventions for physical activity and dietary behavior change. Am J Prev Med 2007 Oct;33(4):336-345 [FREE Full text] [Medline: 17888860] [doi: 10.1016/j.amepre.2007.05.007]
14. Kroeze W, Werkman A, Brug J. A systematic review of randomized trials on the effectiveness of computer-tailored education on physical activity and dietary behaviors. Ann Behav Med 2006 Jun;31(3):205-223. [Medline: 16700634] [doi: 10.1207/s15324796abm3103_2]
15. Van den Berg MH, Schoones JW, Vliet Vlieland TPM. Internet-based physical activity interventions: a systematic review of the literature. J Med Internet Res 2007;9(3):e26 [FREE Full text] [Medline: 17942388] [doi: 10.2196/jmir.9.3.e26]
16. Brug J, Campbell M, Van Assema P. The application and impact of computer-generated personalized nutrition education: a review of the literature. Patient Educ Couns 1999 Feb;36(2):145-156. [Medline: 10223019] [doi: 10.1016/S0738-3991(98)00131-1]
17. Skinner CS, Campbell MK, Rimer BK, Curry S, Prochaska JO. How effective is tailored print communication? Ann Behav Med 1999;21(4):290-298. [Medline: 10721435] [doi: 10.1007/BF02895960]
18. de Vries H, Brug J. Computer-tailored interventions motivating people to adopt health promoting behaviours: introduction to a new approach. Patient Educ Couns 1999 Feb;36(2):99-105. [Medline: 10223015] [doi: 10.1016/S0738-3991(98)00127-X]
19. Hurling R, Catt M, Boni MD, Fairley BW, Hurst T, Murray P, et al. Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. J Med Internet Res 2007;9(2):e7 [FREE Full text] [Medline: 17478409] [doi: 10.2196/jmir.9.2.e7]
20. Richardson CR, Mehari KS, McIntyre LG, Janney AW, Fortlage LA, Sen A, et al. A randomized trial comparing structured and lifestyle goals in an internet-mediated walking program for people with type 2 diabetes. Int J Behav Nutr Phys Act 2007;4(1):59 [FREE Full text] [Medline: 18021411] [doi: 10.1186/1479-5868-4-59]
21. Slootmaker SM, Chinapaw MJM, Schuit AJ, Van Mechelen W, Koppes LL. Concurrent validity of the PAM accelerometer relative to the MTI Actigraph using oxygen consumption as a reference. Scand J Med Sci Sports 2009 Feb;19(1):36-43. [Medline: 18266793]
22. Slootmaker SM, Chinapaw MJ, Schuit AJ, Seidell JC, Van Mechelen W. Promoting physical activity using an activity monitor and a tailored web-based advice: design of a randomized controlled trial [ISRCTN93896459]. BMC Public Health 2005;5(1):134 [FREE Full text] [Medline: 16356182] [doi: 10.1186/1471-2458-5-134]
23. PAM COACH website. URL: http://www.pam.com/ [accessed 2009-01-31] [WebCite Cache ID 5eAT5xpup]
24. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice-Hall; 1986.
25. Van Sluijs EM, Van Poppel MN, Twisk JW, Chinapaw MJ, Calfas KJ, Van Mechelen W. Effect of a tailored physical activity intervention delivered in general practice settings: results of a randomized controlled trial. Am J Public Health 2005 Oct;95(10):1825-1831 [FREE Full text] [Medline: 16186461] [doi: 10.2105/AJPH.2004.044537]
26. Sallis JF, Pinski RB, Grossman RM, Patterson TL, Nader PR. The development of self-efficacy scales for health related diet and exercise behaviors. Health Educ Res 1988;3(3):283-292. [doi: 10.1093/her/3.3.283]

27. Sallis JF, Grossman RM, Pinski RB, Patterson TL, Nader PR. The development of scales to measure social support for diet and exercise behaviors. Prev Med 1987 Nov;16(6):825-836. [Medline: 3432232] [doi: 10.1016/0091-7435(87)90022-3]
28. Buckley JP, Sim J, Eston RG, Hession R, Fox R. Reliability and validity of measures taken during the Chester step test to predict aerobic power and to prescribe aerobic exercise. Br J Sports Med 2004 Apr;38(2):197-205 [FREE Full text] [Medline: 15039259] [doi: 10.1136/bjsm.2003.005389]
29. Shephard RJ, Thomas S, Weller I. The Canadian Home Fitness Test. 1991 update. Sports Med 1991 Jun;11(6):358-366. [Medline: 1925183] [doi: 10.2165/00007256-199111060-00002]
30. Durnin JV, Womersley J. Body fat assessed from total body density and its estimation from skinfold thickness: measurements on 481 men and women aged from 16 to 72 years. Br J Nutr 1974 Jul;32(1):77-97. [Medline: 4843734] [doi: 10.1079/BJN19740060]
31. Tudor-Locke C, Bell RC, Myers AM, Harris SB, Ecclestone NA, Lauzon N, et al. Controlled outcome evaluation of the First Step Program: a daily physical activity intervention for individuals with type II diabetes. Int J Obes Relat Metab Disord 2004 Jan;28(1):113-119. [Medline: 14569279] [doi: 10.1038/sj.ijo.0802485]
32. Carr LJ, Bartee RT, Dorozynski C, Broomfield JF, Smith ML, Smith DT. Internet-delivered behavior change program increases physical activity and improves cardiometabolic disease risk factors in sedentary adults: results of a randomized controlled trial. Prev Med 2008 May;46(5):431-438. [Medline: 18207228] [doi: 10.1016/j.ypmed.2007.12.005]
33. Chan CB, Ryan DA, Tudor-Locke C. Health benefits of a pedometer-based physical activity intervention in sedentary workers. Prev Med 2004 Dec;39(6):1215-1222. [Medline: 15539058] [doi: 10.1016/j.ypmed.2004.04.053]
34. Bowen DJ, Fries E, Hopp HP. Effects of dietary fat feedback on behavioral and psychological variables. J Behav Med 1994 Dec;17(6):589-604. [Medline: 7739032] [doi: 10.1007/BF01857599]
35. McKay HG, King D, Eakin EG, Seeley JR, Glasgow RE. The diabetes network internet-based physical activity intervention: a randomized pilot study. Diabetes Care 2001 Aug;24(8):1328-1334 [FREE Full text] [Medline: 11473065] [doi: 10.2337/diacare.24.8.1328]
36. Tate DF, Wing RR, Winett RA. Using Internet technology to deliver a behavioral weight loss program. JAMA 2001 Mar 7;285(9):1172-1177 [FREE Full text] [Medline: 11231746] [doi: 10.1001/jama.285.9.1172]

Abbreviations

AQuAA: Activity Questionnaire for Adolescents and Adults
BMI: body mass index
CST: Chester Step Test
HR: heart rate
ICT: information and communication technologies
MET: metabolic equivalent
PA: physical activity
PAM: personal activity monitor
RCT: randomized controlled trial

Edited by T Houston; submitted 25.09.08; peer-reviewed by J Kabukoba; comments to author 11.12.08; revised version received 24.02.09; accepted 03.03.09; published 29.07.09.

Please cite as:
Slootmaker SM, Chinapaw MJM, Schuit AJ, Seidell JC, Van Mechelen W
Feasibility and Effectiveness of Online Physical Activity Advice Based on a Personal Activity Monitor: Randomized Controlled Trial
J Med Internet Res 2009;11(3):e27
URL: http://www.jmir.org/2009/3/e27/
doi:10.2196/jmir.1139
PMID:19674956

Original Paper

Mobile and Fixed Computer Use by Doctors and Nurses on Hospital Wards: Multi-method Study on the Relationships Between Clinician Role, Clinical Task, and Device Choice

Pia Andersen[1], BSc; Anne-Mette Lindgaard[1], BSc; Mirela Prgomet[2], BAppSc(HIM); Nerida Creswick[2], PhD; Johanna I Westbrook[2], PhD

[1]Medical Informatics Group, Department of Health Science and Technology, Aalborg University, Denmark

[2]Health Informatics Research and Evaluation Unit, University of Sydney, Australia

Corresponding Author:
Johanna I Westbrook, PhD
University of Sydney
Director, Health Informatics Research & Evaluation Unit
Faculty of Health Sciences
PO Box 170
Lidcombe 1825
Australia
Phone: +61 2 9351 9677
Fax: +61 2 9351 9676
Email: J.Westbrook@usyd.edu.au

Abstract

Background: Selecting the right mix of stationary and mobile computing devices is a significant challenge for system planners and implementers. There is very limited research evidence upon which to base such decisions.

Objective: We aimed to investigate the relationships between clinician role, clinical task, and selection of a computer hardware device in hospital wards.

Methods: Twenty-seven nurses and eight doctors were observed for a total of 80 hours as they used a range of computing devices to access a computerized provider order entry system on two wards at a major Sydney teaching hospital. Observers used a checklist to record the clinical tasks completed, devices used, and location of the activities. Field notes were also documented during observations. Semi-structured interviews were conducted after observation sessions. Assessment of the physical attributes of three devices—stationary PCs, computers on wheels (COWs) and tablet PCs—was made. Two types of COWs were available on the wards: generic COWs (laptops mounted on trolleys) and ergonomic COWs (an integrated computer and cart device). Heuristic evaluation of the user interfaces was also carried out.

Results: The majority (93.1%) of observed nursing tasks were conducted using generic COWs. Most nursing tasks were performed in patients' rooms (57%) or in the corridors (36%), with a small percentage at a patient's bedside (5%). Most nursing tasks related to the preparation and administration of drugs. Doctors on ward rounds conducted 57.3% of observed clinical tasks on generic COWs and 35.9% on tablet PCs. On rounds, 56% of doctors' tasks were performed in the corridors, 29% in patients' rooms, and 3% at the bedside. Doctors not on a ward round conducted 93.6% of tasks using stationary PCs, most often within the doctors' office. Nurses and doctors were observed performing workarounds, such as transcribing medication orders from the computer to paper.

Conclusions: The choice of device was related to clinical role, nature of the clinical task, degree of mobility required, including where task completion occurs, and device design. Nurses' work, and clinical tasks performed by doctors during ward rounds, require highly mobile computer devices. Nurses and doctors on ward rounds showed a strong preference for generic COWs over all other devices. Tablet PCs were selected by doctors for only a small proportion of clinical tasks. Even when using mobile devices clinicians completed a very low proportion of observed tasks at the bedside. The design of the devices and ward space configurations place limitations on how and where devices are used and on the mobility of clinical work. In such circumstances, clinicians will initiate workarounds to compensate. In selecting hardware devices, consideration should be given to who will be using the devices, the nature of their work, and the physical layout of the ward.

(J Med Internet Res 2009;11(3):e32) doi:10.2196/jmir.1221

KEYWORDS

Study; multi-method study; observational study; mobility; mobile computers; computers; computer hardware; medical order entry systems; computerized physician order entry system; computerized provider order entry (CPOE)

Introduction

The use of information and communication technologies (ICT) in the health care sector has become widespread in several countries [1], and governments around the world continue to invest in the implementation of ICT systems [2]. These clinical systems comprise a variety of functions, including medication management, order entry, results viewing, clinical documentation, and decision support capability [3]. Introduction of ICT in a hospital affects the operation of the organization, health care delivery, and patient outcomes while offering potential benefits in improved patient safety, reduced hospital costs, and increased efficiency and effectiveness of medical care [4-6].

Not all ICT systems are successfully implemented. A key factor for success is the extent to which systems integrate with clinical workflow [7,8]. Clinical work is characterized by a complex mixture of routine and unexpected events and involves close collaboration among practitioners [9]. Furthermore, clinical work is highly mobile. Health professionals move frequently among wards, clinics, offices, and other locations and require information at each of these locations [10,11]. Thus, new ICT systems must, among other factors, complement the mobile, collaborative nature of medical work in order to succeed [9].

A core component of system implementation is the selection of hardware. Early clinical system implementations relied upon replacing paper-based records with information accessible via stationary personal computers (PCs). Stationary PCs allow easy storage, searching, retrieval, and sharing of information [12]; however, they constrain work to a fixed location [13]. The design of new and more mobile hardware devices has increased the range of options available to health care organizations, and mobile devices have been advocated [14] as a means of providing practitioners with access to patient and clinical information at the point of care [15]. Subsequently, the integration and use of both stationary and mobile computing technologies within health care have been promoted as the approach most likely to achieve the greatest results for clinicians and their patients [13,16].

Systematic reviews of the use of handheld devices within medicine have reported benefits in allowing easy access to information, decision support, and improved communication, but such reviews have also identified barriers to their implementation and use [17-21]. Very few studies have made comparisons between stationary and mobile devices, investigated how clinical staff select a device when multiple devices are available, or determined the locations in which clinicians choose to use mobile devices. Thus, this limited research base means that determining the right mix of stationary and mobile devices is a significant challenge for system planners and implementers. Little is known about the degree to which different hardware devices are capable of adequately supporting the complex and often collaborative nature of work in a hospital environment. In order to obtain the benefits ICT systems offer, it is necessary to gain insights into the way users interact with different hardware devices and the impact these systems have on work practices. The extent or ways in which different hardware devices may be more effective for different professionals or clinical tasks has rarely been investigated.

We conducted a study to investigate the relationships between clinician role, clinical task, and selection of hardware device. Our aim was to answer two central questions: (1) which device is used by whom, where, for which clinical task, and in collaboration with whom? and (2) what impact does the design of the device have on its use on hospital wards?

Methods

Study Design

We utilized a multi-method approach which included: (1) direct observations, (2) interviews, (3) an assessment of the physical attributes of available hardware devices, and (4) heuristic evaluation of the user interfaces. Advocates of a multi-method design support its use as a means of gaining clearer understanding and insights into the impact of technology on health care services [22].

Setting

The study was conducted in two geriatric wards of a Sydney metropolitan teaching hospital. Each ward had 26 beds and was at maximum occupancy throughout the majority of the study period. The Cerner Millennium PowerChart system (Cerner, Kansas City, MO), which comprises computerized test ordering, results viewing, and electronic medication management, was available from stationary PCs, COWs, and tablet PCs across both study wards. While computerized test ordering and results viewing have been used for several years in the two wards, the electronic medication management system had been implemented and used since November 2007 in ward A and July 2008 in ward B.

Computer Hardware Devices and Ward Layouts

The two wards were located next to one another, and their layout and characteristics were much alike (Figure 1). Each ward had seven stationary PCs. Three were located in the doctors' offices, one in the corridor, one in the medication room, and two in the central workstation (nurses' and clerks' station). The two wards also had stationary PCs in the Nursing Unit Manager's office; however, these computers were not included in the study as they were not used for clinical tasks. Ward A and ward B had five and six COWs, respectively. In ward A, four of the COWs were generic COWs (laptops mounted on trolleys—three Acer Travelmate 5620 and one Acer Travelmate 7720), and one was an ergonomic COW (an integrated computer and cart device

specifically designed to make it easy to move—Dell Latitude D620 and StyleView® Notebook Cart) (Figure 2). In ward B, the six COWs were all generic COWs (Acer Travelmate 7720). The COWs were generally distributed throughout the corridor area. Each of the wards also had two tablet PCs (two Motion Computing C5 and two Motion Computing LE1700). In ward A, the tablet PCs were placed in the medication room, while in ward B they were kept at the central workstation. All devices ran Windows XP as their operating system.

Figure 1. Overview of the placement of devices in ward A and ward B

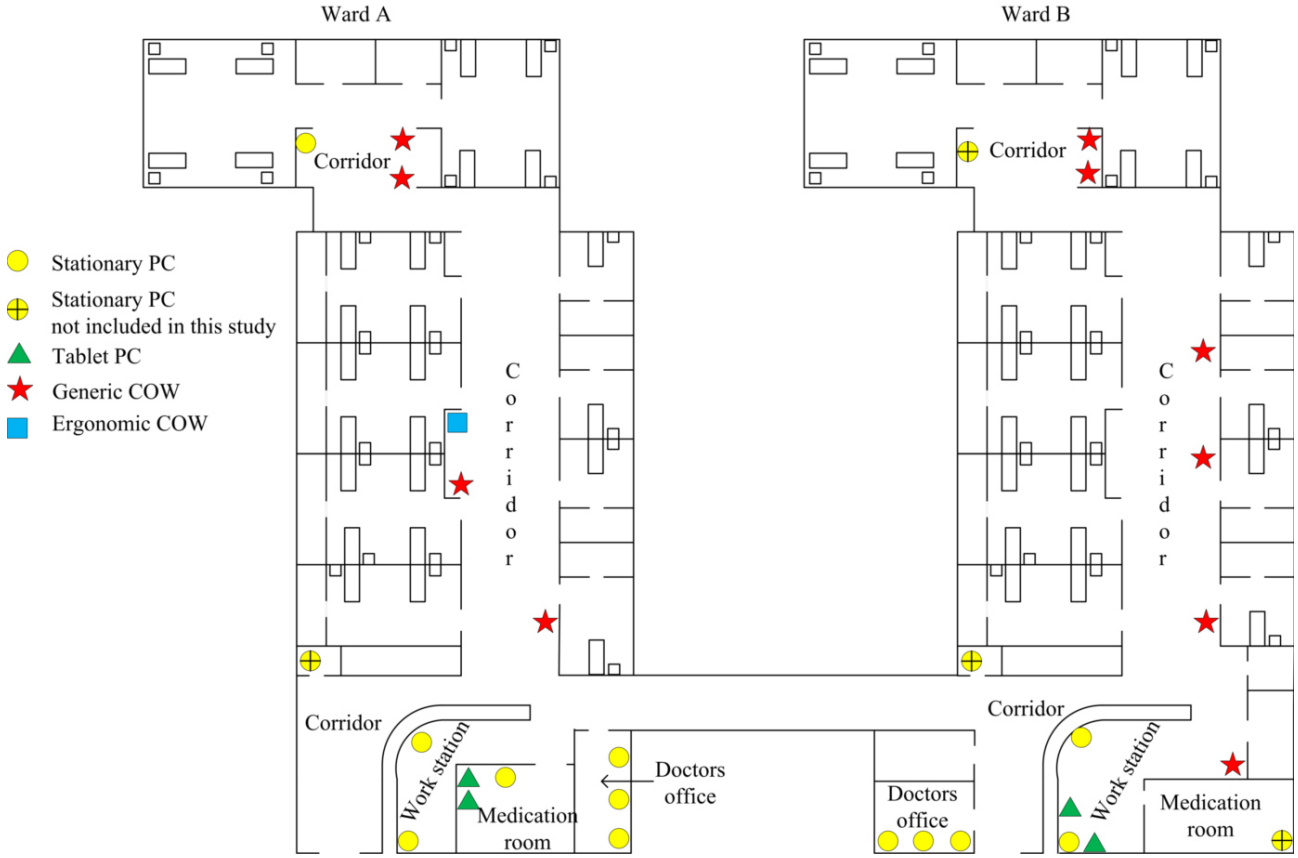


Figure 2. Devices available in the two wards (A: tablet PC ward A, B: tablet PC ward B, C: generic COW, D: ergonomic COW)

Sample

The observational study was conducted between October 13 and 31, 2008 for a total of 80 hours and 1 minute spread across 144 observation sessions. The sample consisted of 27 nurses (13 from ward A and 14 from ward B) and 8 doctors who worked across both wards. This represents approximately 50% of the nursing and medical staff allocated to these wards. Nurses had an average work experience of 2.7 years in the respective wards, and the average experience of doctors was 2.9 years in the study wards. Participants were provided with an information letter outlining the study, and each consenting participant was assigned a unique identifying number. When the researchers arrived on the ward, participants using the PowerChart system were identified and one was selected for observation. Each participant was observed for a maximum of two consecutive hours and a maximum of six hours in total for the duration of the study. A convenience sample of ten nurses and two doctors who participated in the observational study also participated in brief interviews.

Data Collection Procedures

Observational Study

The observational study was conducted over a three-week period in October 2008. The study participants were observed while undertaking clinical tasks during their day-to-day routine that required interaction with a stationary PC, one of two styles of computers on wheels (COWs), or a tablet PC.

We developed an observational data collection form with a list of clinical tasks based upon previous work (Table 1) [23,24]. This form was used to record: the participant-ID; the ward; the date; the start and end time of each observation session; clinical tasks performed; hardware device used (stationary PC, generic COW, ergonomic COW or tablet PC); the location of task completion (doctors' office, corridor, bedside, medication room,

workstation, or patient room); whether collaboration occurred during the task (with a doctor, nurse or other health care personnel); and whether the observed participant was on a ward round. The location was defined as at the "bedside" if the observed participant used one of the devices directly next to the bedside of the patient being treated. Where the device was placed next to one patient's bedside while treatment was provided to another patient within the room, the location was recorded as "patient room". Collaboration between the observed participant and another health professional was recorded if they looked at the computer screen at the same time and conducted a clinical task together (eg, clinicians discussing a particular prescription in the medication chart or a nurse observed asking another nurse to document witnessing a drug to be administered). During the observation sessions, additional factors potentially impacting on device selection were also noted, for example, when a computer crashed. The average length of the observation sessions was 33 minutes. All observational sessions were conducted by AL and PA.

Table 1. Clinical task classification

Clinical task	Definition
Review chart	Reviewing the medical chart, allergy, height, weight, health insurance, or other patient information.
Prepare drug	Reading medication order to select drug.
Administer drug	Recording that a drug has been administrated.
Unchart drug	Removing documentation of drug administration (eg, because it was recorded prior to actual administration and the drug did not end up being given to the patient).
Witnessing drug	Confirming that preparation of a restricted drug, or another special drug, has been checked.
Ordering drug/test	Writing up a new medication order or ordering a test.
Changing/canceling drug	Changing drug order details or canceling a drug order.
Viewing results	Looking at a patient's test results.
Generating discharge	Writing a discharge summary.
Reorder drug	Extending the length of a drug order.
Drug information	Looking at drug information.
Documenting allergies	Documentation of a patient's allergies.
Documenting others	Documentation of other information relevant to medication orders (eg, height, weight, blood glucose).
Copying medication order	Copying medication orders from paper to PowerChart.

Researchers were trained in the PowerChart system to allow them to become acquainted with the system and to be able to distinguish accurately among the different clinical tasks performed using the system. A pilot study, consisting of approximately 17 hours of observation, was undertaken between October 7 and 13, 2008 to assess the design of the data collection form. During the first session of the pilot study, it was discovered that the doctors and nurses changed clinical tasks rapidly, which was not supported by the initial design of the form. The pilot study also revealed that the list of defined clinical tasks did not cover all tasks doctors and nurses conducted in the PowerChart system, and some of the defined tasks were difficult for the researchers to distinguish. This led to the addition of new tasks (eg, "unchart drug") and combination of other tasks (eg, entering a new drug and ordering a test were combined in the task "ordering drug/test"). After modifying and testing the new data collection form, inter-observer reliability testing between the researchers was conducted by the two observers concurrently, but independently, observing nurses and doctors for a total of four hours and comparing their results. Formal data collection began once an overall agreement of over 85% was achieved between the data collectors. Throughout the study, data were collected on weekdays between the hours of 7 am and 4 pm. Each researcher spent half of the study period independently observing in one ward and then rotated to observe in the alternate ward. After each observation session, the manually collected data were transcribed onto a computer. Field notes were documented by the researchers to provide contextual information about, for example, physical space limitations which impacted upon the way a device was used.

Interviews

Interviews were conducted during the last week of data collection. Based on the prior observations, the nurses and doctors were asked about their individual device preference in a given context (eg, when giving medication or going on ward rounds) and their reasons for choosing a particular device. Participants were asked the following questions: When conducting a task in relation to medication, results viewing, or ordering tests, which of the hardware devices do you generally use and why?; Is there a certain device you prefer and why?; Do you ever use one of the other devices for certain tasks? If yes, what made you choose to use the other device? If no, has somebody shown you how to use the device? If yes, why have you decided not to use that device? The data collected in the interviews were recorded on paper and later transcribed to digital format. The interviews were conducted by the same researchers who undertook the observations.

Physical Attributes Assessment

Examination of the physical attributes of each hardware device (such as the size of the screen and boot-up time) was performed to further identify factors that may impact on device selection. These were conducted by two researchers (AL and PA). A form was developed to evaluate the physical attributes of each device. Weighing the generic and ergonomic COWs was not possible; therefore, only the tablet PCs were weighed. The system boot-up times were measured, as were the start and restart times. The start time was calculated from the time the start button on the device was pressed until the programs had finished loading (indicated by the hourglass disappearing). The time it took to restart the device was measured from pressing "ok" on the Windows shut-down screen until the device had finished reloading all programs. Additionally, the time it took to start PowerChart was measured. First, the time from clicking on the PowerChart icon on the desktop until the login screen appeared was measured. Second, after typing in a username and password, the time from pressing "ok" on the log in screen until PowerChart had completely loaded was measured. These times were measured both immediately after the computer had been started or restarted (one time) and after PowerChart had been previously opened (measured three times). Each device within the wards was evaluated individually.

Heuristic Evaluation

The heuristic evaluation was undertaken to identify design factors which may influence device selection and use. The evaluation was completed independently by two researchers (AL and PA) for each defined clinical task and conducted on the basis of a set of 10 recognized usability principles [25]. Following independent assessment by each investigator, results were compared with regard to how clinical tasks were completed, taking into account the 10 usability principles to gain consensus. As the same clinical system was accessible across the various computing devices and the appearance of the system was identical, the particular focus was on identifying how the screen size of each device may have affected the aesthetics and usability of the device, subsequently influencing device selection. The evaluation was conducted on a stationary PC with a 17″ screen followed by a tablet PC with a 10.5″ screen, with an emphasis on user-interface design issues that might be present in smaller devices.

Data Analysis

Observational Study and Interviews

Observational data were highly structured as described above, and data analysis was undertaken in SPSS version 16.0 (SPSS Inc, Chicago, IL). Descriptive statistics were calculated, including the proportions of clinical tasks undertaken using the different devices, device use by clinical group, and the physical locations where devices were used. Interviews and field notes were analyzed separately by the researchers and the main themes identified. This involved, for example, all responses to specific interview questions being independently reviewed by two researchers, and the most frequently recurring issues raised by respondents were identified. This allowed identification of the factors which doctors and nurses most frequently raised as reasons for why they selected or did not select a specific device. Given the structured nature of the questions, there was little disagreement between reviewers. When disagreement arose, it was resolved through consensus and discussion with the other co-authors. Field notes were analyzed in terms of identification of the range of issues raised with greater emphasis given to those factors which occurred on multiple occasions. For re-occurring issues, these were counted to provide an indication of their frequency.

Physical Attributes Assessment and Heuristic Evaluation

Physical and heuristic assessment data were tabulated to allow comparisons between devices on a range of attributes. The physical attributes data were used to calculate the mean and standard deviation boot-up times for each device.

The study was approved by the University of Sydney Human Research Ethics and the hospital Human Research Ethics Committees.

Results

Observational Study Results

A total of 144 observation sessions was conducted over a period of 80 hours and one minute during which 2158 clinical tasks were observed. Nurses were observed in 103 (71.5%) of the observation sessions, and their tasks accounted for 81.1% of the tasks recorded. Doctors were observed in 41 (28.5%) observation sessions, and their tasks made up 18.9% of the tasks observed. Of the observed tasks performed by doctors, 57.5% were collected during ward rounds.

The number of clinical tasks performed per individual ranged from 3 (0.1%) to 227 (10.5%) for nurses and 7 (0.3%) to 112 (5.2 %) for doctors. Observations were fairly evenly distributed across the two wards, with 71 sessions conducted in ward A, in which 42.6% of the tasks were collected, and 68 sessions in ward B, accounting for 57.4% of tasks. Five observation sessions occurred in both wards, where doctors walked from one ward to the other and used a hardware device in both wards.

Observed Use of Hardware Device by Clinical Task and Location

The majority of observed tasks (82.3%) were undertaken using the generic COW, with 9.6% conducted on a stationary PC, 5.1% on a tablet PC, and 3.0% on the ergonomic COW. Most clinical tasks (49.1%) were completed in a patient's room, followed by the corridor (35.2%), the doctors' office (8.1%), a patient's bedside (4.3%), the central workstation (3.0%), and the medication room (0.2%). The most frequently observed tasks were administering drugs (31.4%), preparing drugs (27.9%), and reviewing a patient's chart (15.4%). The frequency of the other observed tasks ranged from 0.1% to 7.5%.

Observation of Nurses' Use of Hardware Devices on Wards

Nurses undertook 93.1% of observed tasks using the generic COWs. All except one of the 27 observed nurses used a generic COW during the study period. Ten nurses also used a stationary

PC, four nurses used an ergonomic COW, and one nurse used a tablet PC during the study period.

The generic COWs, ergonomic COW, and the tablet PCs were most frequently used by nurses in the patients' rooms (56.7%) and in the corridors (35.8%), mostly to administer and prepare drugs. If a generic COW and the ergonomic COW were available and placed next to each other, it was often observed that nurses chose the generic COW. When nurses used the stationary PCs, they primarily used the ones at the workstations, followed by those located in the corridors, to witness and prepare drugs (Table 2). Less than 2% of nurses' tasks were undertaken using stationary PCs. None of the nurses were observed using a stationary PC in the doctors' office or in the medication room. It was observed on one occasion that a generic COW was wheeled into the medication room, when a stationary PC within the room was out of order.

Table 2. Clinical tasks conducted by nurses by device and location of use (N = 1751 tasks)

Location	Clinical task	Tablet PC	Generic COW	Ergonomic COW	Stationary PC	Total (% of total tasks)
Bedside						
	Review chart	1	5	-	-	6
	Prepare drug	1	28	1	-	30
	Administer drug	3	30	2	-	35
	Witnessing drug	1	6	1	-	8
	Ordering drug/test	-	1	-	-	1
	Drug information	-	1	-	-	1
	Viewing results	-	1	-	-	1
	Documenting others	-	3	1	-	4
	Total	6	75	5	-	86 (4.9%)
Patient's room						
	Review chart	-	65	4	-	69
	Prepare drug	4	348	10	-	362
	Administer drug	6	396	11	-	413
	Unchart drug	-	7	1	-	8
	Witnessing drug	3	62	3	-	68
	Ordering drug/test	-	14	2	-	16
	Changing/canceling drug	-	1	-	-	1
	Reorder drug	-	1	-	-	1
	Drug information	-	12	1	-	13
	Viewing results	-	2	-	-	2
	Documenting others	1	38	-	-	39
	Total	14	946	32	-	992 (56.7%)
Medication room						
	Prepare drug	-	2	-	-	2
	Administer drug	-	2	-	-	2
	Ordering drug/test	-	1	-	-	1
	Total	-	5	-	-	5 (0.3%)
Workstation						
	Review chart	-	1	-	1	2
	Prepare drug	-	1	-	5	6
	Administer drug	-	8	-	6	14
	Witnessing drug	-	8	-	10	18
	Total	-	18	-	22	40 (2.3%)
Corridor						
	Review chart	2	96	10	2	110
	Prepare drug	2	187	9	3	201
	Administer drug	1	203	6	1	211
	Unchart drug	-	6	-	-	6
	Witnessing drug	2	63	2	-	67
	Ordering drug/test	-	9	-	-	9

Location	Clinical task	Tablet PC	Generic COW	Ergonomic COW	Stationary PC	Total (% of total tasks)
	Reorder drug	-	1	-	1	2
	Drug information	-	7	-	-	7
	Viewing results	-	1	-	-	1
	Documenting others	-	11	1	-	12
	Other task	-	1	-	-	1
	Total	7	585	28	7	627 (35.8%)
Multiple locations						
	Prepare drug	-	1	-	-	1
	Total	-	1	-	-	1 (0.06%)
Total		27	1630	65	29	1751 (100%)

Observation of Doctors' Use of Hardware Devices on Ward Rounds and Outside Ward Rounds

Eight doctors were observed at least once on a ward round during which they primarily used a generic COW (Table 3). Four doctors were observed using a stationary PC, and two doctors used a tablet PC, while none of them used an ergonomic COW on ward rounds during the observation period.

Of the observed tasks conducted on ward rounds, the majority (57.3%) were completed using a generic COW, while 35.9% were completed using a tablet PC. These two devices were most frequently used in the corridor, followed by the patients' rooms, to review charts and view results. Use of a tablet PC at the workstation was observed more often than the use of a stationary PC (Table 3). Of the 234 clinical tasks completed during ward rounds, 57.3% were undertaken in the corridors, 29.1% in the patients' rooms, and 3% at a patient's bedside.

Seven doctors were observed at least once while not on a ward round. On each of these occasions they were observed using a stationary PC, with one doctor also using a generic COW. Doctors were not observed using tablet PCs or the ergonomic COW when not on ward rounds.

These seven doctors (not on ward rounds) conducted the vast majority (93.6%) of tasks using stationary PCs to view results and review charts, and most often they did this within the doctors' office (Table 3). Generating discharge summaries was primarily conducted when doctors were not on ward rounds and only using stationary PCs.

Table 3. Clinical tasks conducted by doctors on ward rounds and not on ward rounds by device and location of use (n = 407 tasks)

		Device			
Location	Clinical task	Tablet PC	Generic COW	Stationary PC	Total (% of total tasks)
Ward round					
Bedside					
	Review chart	2	-	-	2
	Changing/canceling drug	1	-	-	1
	Viewing results	4	-	-	4
	Total	7	-	-	7 (1.7%)
Patient's room					
	Review chart	10	15	-	25
	Ordering drug/test	9	5	-	14
	Changing/canceling drug	2	4	-	6
	Viewing results	10	13	-	23
	Total	31	37	-	68 (16.7%)
Doctors' office					
	Review chart	-	-	3	3
	Ordering drug/test	-	-	4	4
	Changing/canceling drug	-	-	2	2
	Generating discharge	-	-	1	1
	Viewing results	-	-	3	3
	Total	-	-	13	13 (3.2%)
Workstation					
	Review chart	4	-	1	5
	Ordering drug/test	1	-	1	2
	Viewing results	6	-	-	6
	Other task	-	-	1	1
	Total	11	-	3	14 (3.4%)
Corridor					
	Review chart	16	38	-	54
	Administer drug	-	1	-	1
	Ordering drug/test	4	12	-	16
	Changing/canceling drug	6	15	-	21
	Drug information	-	1	-	1
	Viewing results	9	30	-	39
	Total	35	97	-	132 (32.4%)
Total		84	134	16	234 (57.5%)
Not on Ward Round					
Doctors' office					
	Review chart	-	-	51	51
	Administer drug	-	-	2	2
	Ordering drug/test	-	-	22	22
	Changing/canceling drug	-	-	5	5

Location	Clinical task	Device: Tablet PC	Generic COW	Stationary PC	Total (% of total tasks)
	Generating discharge	-	-	19	19
	Reorder drug	-	-	1	1
	Viewing results	-	-	61	61
	Other task	-	-	1	1
	Total	-	-	162	162 (39.8%)
Workstation					
	Review chart	-	5	-	5
	Ordering drug/test	-	2	-	2
	Viewing results	-	4	-	4
	Total	-	11	-	11 (2.7%)
Total		-	11	162	173 (42.5%)

Collaborative Activities by Nurses and Doctors Using Hardware Devices

Nurses collaborated with other health professionals on 197 clinical tasks (11.3%) undertaken while using a computer device. Nurses mostly collaborated with other nurses (58.4%) and student nurses (40.0%) using generic COWs (91.9%). Collaboration centered around the witnessing (33.5%, n = 66), preparation (27.4%, n = 54) and administration (24.9%, n = 49) of drugs. Nurses were observed collaborating with a doctor using generic COWs to review charts (2 tasks) and prepare drugs (1 task).

Doctors on ward rounds conducted 88 clinical tasks (37.6%) in collaboration with another health professional. On ward rounds, doctors collaborated most often with other doctors (68.2%) and medical students (26.1%), and this occurred most frequently using a generic COW (75.0%). Doctors most often collaborated on ward rounds to review charts (42.0%, n = 37) and view results (39.8%, n = 35). Doctors not on ward rounds conducted 12 clinical tasks (6.9%) in collaboration with another health professional and mostly collaborated with other doctors (58.3%) on the stationary PCs (100.0%) and primarily when viewing results (50%, n = 6). Doctors were not observed collaborating with nurses at any time when using one of the computing devices.

Contextual Factors Identified From the Field Notes

Two main themes from the field notes were identified: device mobility and availability. In relation to mobility, nine nurses placed a drawer containing a patient's medications on top of the trolley of the generic COW. This was observed in 14 nurse observation sessions (13.6%). In this hospital, most patients' medications are locked in their individual bedside tables. Thus, we observed nurses placing these medication drawers (one patient at a time) onto the generic COWs in both the patient's room and also when the trolley was situated in the corridors. The nurses went back and forth between the device and each patient's bedside to administer the medication.

During six observation sessions, five nurses and one doctor were observed writing details about a patient's drug order on a napkin or a piece of paper. Most frequently this action was observed in patients' rooms when using a generic COW. The nurses used the paper note to take with them to the medication room in order to prepare a medication which was not available in a patient's bedside drawer. It was also observed that the generic COW was placed between a patient's bed and a wall on several occasions which left clinicians with very little space for movement.

The field notes identified that availability of the hardware devices was also an issue. For example, in 15 sessions (10.4%), problems with the generic COWs occurred primarily because the device ran out of power due to the fact it had not been recharged or because it did not work when first attempted by the observed participant. Similar problems were observed with the ergonomic COW during four sessions.

Results of Interviews With Nurses and Doctors

Ten of the 27 observed nurses and two of the eight observed doctors were interviewed regarding their choice of device for completing tasks in relation to medications management, results viewing, or test ordering. All interviewed nurses preferred to use a generic COW over a tablet PC when giving medication, and three of the nurses reported that they had never used a tablet PC. The two doctors had both used the tablet PC and generic COW on ward rounds, but one preferred the tablet PC while the other preferred the generic COW.

From the responses to the interview questions, which focused on why doctors and nurses selected or did not select specific devices, six main themes were identified. These were: availability, speed, mobility, device design, knowledge about the device, and problems. Eight participants (seven nurses and one doctor) stated that one of the main reasons for using the generic COW as opposed to the tablet PC or stationary PC was the design of the trolley, which had space for the storage of medication, paper charts and/or other equipment. Despite comments that the need to recharge the batteries of generic

COWs was a limitation and restricted mobility, a larger number of users stated that the advantage of the generic COWs over the other devices was better mobility because they can be used everywhere and because COWs allow items required for task completion to be stored on them. In relation to the use of the tablet PCs, clinicians reported difficulties with the stylus, including that it made task completion slow. Other limitations included small screen size and that it was awkward to carry the tablet PCs while walking.

Nine participants (seven nurses and two doctors) stated that they had used the tablet PCs in their daily routine. Of the participants that had used a tablet PC, two had received training while four reported having never received training; in addition, three didn't comment on whether they had received training. Five nurses only used the tablet PCs if no generic COWs were available or the generic COW had crashed. Three nurses stated they had not used the tablet PCs, and also mentioned that they had not received any training in the use of the tablet PC. Four nurses reported that they also used stationary PCs when preparing or witnessing drugs in the medication room as it is at the point of need.

Results of Assessment of the Physical Attributes of Hardware Devices

The evaluation of the physical attributes revealed that none of the devices automatically logged off the PowerChart system or closed down after a period of inactivity. Additionally, it was possible to log on to the PowerChart system from more than one device at the same time. All devices were connected to the Internet through the hospital's wireless network.

There were different types of stationary PCs in the two wards. Six of these had a 15″ screen, four had a 17″ screen, and two had a 19″ screen (Table 4). In ward B, the stationary PC in the medication room and the one at the end of the corridor were out of order throughout the data collection period, and thus the attributes of these devices were not included in the study.

Table 4. Physical attributes of the different devices

	Stationary PCs	Generic COW	Ergonomic COW	Tablet PC
Number of devices	12	10	1	4
Screen size diagonal (inch)	15", 17", 19"	17"	14"	10.5", 12"
Weight (kg)				1.6, 1.9
Dimensions		Trolley	Trolley	Device
Height min/max (cm)	-	73/105 - 77/100	70/110	25 - 26
Width min/max (cm)	-	92/121 - 96/126	53	29.5 - 25.5
Depth (cm)	-	37 - 45	67	2.5 - 2.5
Boot-up times (min)				
Mean restart time (SD)	2.54 (1.33)	2.39 (0.46)	2.13 (-)	3.51 (0.41)
Mean start time (SD)	2.16 (1.35)	2.15 (0.50)	1.30 (-)	3.11 (1.12)
Time from desktop to log on screen (sec)				
Mean 1st time (SD)	20.81 (6.62)	17.23 (2.50)	26.16 (-)	20.68 (7.42)
Mean subsequent times (SD)	16.12 (2.66)	16.05 (1.12)	18.21 (2.94)	15.06 (1.45)
Time from log on screen to system start up (sec)				
Mean 1st time (SD)	3.56 (0.57)	3.99 (0.57)	3.37 (-)	3.84 (0.6)
Mean subsequent times (SD)	3.5 (0.45)	3.67 (0.48)	5.30 (2.64)	3.6 (0.51)

All generic COWs had a 17″ screen laptop placed on top of a trolley that was adjustable in height (from floor to the keyboard) and width (Figure 2, C: generic COW). There were two types of trolleys and each ward had one type. The ergonomic COW had a 14″ screen integrated onto a cart, which was only adjustable in height (Figure 2, D: ergonomic COW).

There were two types of tablet PCs available in the wards (Figure 2, A: tablet PC ward A and Figure 2, B: tablet PC ward B). The tablet PCs in ward B (Figure 2, B: tablet PC ward B) were designed with a handle to facilitate transportation, which was done by hand. The lowermost part of the screen on both types of tablet PCs contained an onscreen keyboard where the user, by means of a stylus, interacted with the PowerChart system.

Results of Heuristic Evaluation

The main user interface was composed of a topmost header containing the name of the patient and the patient's medical record number. The remaining part of the user interface comprised tabs, from where it was possible to access the functionality defined for each clinical task (eg, the medication administration record [MAR] tab contained detailed information about the patient's medication). In some of the tasks, an additional window appeared. For example, drug information could be reached by right clicking on an icon near a specific drug in the MAR tab and choosing "reference manual". A new window then appeared which contained information about the drug.

The main difference between using the PowerChart system on a 17″ screen and a 10.5″ screen on the tablet PC was that the system was adjusted to a smaller screen size, and as a result, the font size and the amount of patient data shown on the screen decreased. This resulted in the user having to scroll more with the stylus on a tablet PC to obtain an overview of the information presented on the screen. Additional usability problems were discovered in the task "drug information". In the pop-up window containing drug information an "ok" button was placed in the bottom right corner, which was the only button available for closing the window and returning to the main user interface. When the drug information window was opened on a tablet PC, the lower part of the window was hidden behind the onscreen keyboard and the window needed to be dragged to a higher position before it was possible to close it. Additionally, it was only possible to move or reduce the size of the window by dragging in the frame of the window with the stylus, which was inconvenient. This usability problem was not present on the evaluated stationary PC.

Discussion

This is one of the first studies to investigate computer hardware selection and to quantify the frequency of use by clinician role, clinical task, and location of use. We found that the choice of device was related to clinician role, clinical task, degree of mobility required, and device design. Nurses' work, and tasks performed by doctors during ward rounds, requires highly mobile computer devices. Nurses and doctors on ward rounds showed a strong preference for generic COWs over all other devices. While the greater availability of the generic COWs would account for some of their high utilization, preference for the generic COWs was both observed by the researchers and reported by the clinicians. The ergonomic COW was used occasionally but often only when the generic COWs were unavailable. Rarely did nurses elect to use a tablet PC or a stationary PC.

Nurses' Use of Computer Devices

The nurses indicated that they preferred the generic COW due to its availability, mobility, and design. When not in use the generic COWs were placed along the corridor and were easily accessible to the nurses. The nurses also complimented the mobility of the device, in that it was easy to move around and could be used almost everywhere. The design of the generic COW, which allowed users to store medications and charts, was highly appreciated by the majority of nurses interviewed.

For nurses a high proportion of clinical tasks were completed in the patients' rooms and in the corridors. Space in the patient rooms was a problem when using the generic COW and appeared to be a significant factor preventing nurses from using the COWs at the patients' bedsides. During the observations, we noted that when the generic COW was placed near a patient's bed, it was often difficult for nurses to access the patient's medication drawer. This resulted in the practice of taking the patient's drawer out of the cabinet and placing it on top of the COW. This was a workaround in response to the space limitations imposed by the use of the COWs.

Although the concept of the ergonomic and generic COWs was identical (ie, a laptop placed on a trolley), our study showed large differences in user satisfaction with the two versions of the COW. The main disadvantage of the ergonomic COW was that the available table space was in front of the screen. When nurses placed items on the table space, these items obscured the screen. Additionally, the ergonomic COW had a 14″ screen, whereas the generic COWs had a 17″ screen. The ergonomic COW was also reported as being "hard to push around". These findings clearly indicate that the design of mobile devices impacts their use, which has also been found in different settings [26]. Krogh et al [27] compared a generic COW with a tablet PC, in relation to supporting pharmacists' clinical documentation. Based on subjective evaluations, the authors showed that the pharmacists preferred the tablet PC over the generic COW. The tablet PC was favored due to the design because pharmacists found the device easy to manoeuvre during rounds and the input system (handwriting-to-text functionality) user-friendly and simple.

Only one nurse in our study was observed using the tablet PC while giving medication. This nurse placed the tablet PC on an unused bedside table and rolled this around in the patient room almost as a substitute for a COW. Nurses reported a lack of obvious places to set down the tablet PC while providing direct patient care as a limitation. This problem was also identified by Bogossian et al [28], who found that the portability of tablet PCs was not viewed favorably. The authors reported that the tablet PC was placed at a central point and nurses kept coming back to it since it was inconvenient to carry around. In these situations, the generic COW, with the available trolley space, supports work routines in a better way and is one of the central reasons why the majority of nurses chose to use the COWs in preference to the tablet PCs.

Doctors' Use of Computer Devices

Doctors also preferred the generic COWs for highly mobile tasks such as conducting ward rounds. The generic COWs were preferred particularly on account of the trolley, with table and storage space, and the larger screen size which easily allowed more than one person to view the screen. The doctors also utilized the tablet PCs on many occasions; however, only the tablet PCs from ward B were used. This suggests that the design of the tablet PCs on ward B was preferable to the design of those on ward A. Tablet PCs on ward B weighed less and had a handle which made it easy to transport. Interestingly, when doctors used the COWs and tablet PCs the majority of this use occurred in the corridors, with one-third occurring in the patients' rooms, and only a small proportion at the bedside. The benefit of mobile devices at the patient's bedside was articulated by one doctor who commented that *"you can order and change medications and tests while you talk to and about the patient"*. In a study by Reuss et al [11], where clinical work routine was investigated, physicians reported that, if they did not enter information into the system right away, they felt like they had to process the job twice. Given that these devices are designed to enable use and provision of information directly at the point of care, it is surprising that they were not used at these locations in more instances. Unlike doctors on ward rounds, doctors not on ward rounds were generally not required to examine patients

and found it more convenient to use the system on a stationary PC.

Physical Attributes of Computing Devices Influencing Their Selection and Use

The stylus and the small screen size of the tablet PCs caused usability problems where scrolling was necessary due to the screen size but cumbersome to perform with the stylus. These factors also contributed to nurses selecting devices other than tablet PCs. Similar problems have been reported by doctors using laptops [29], and these factors have also been identified in other clinical settings [30]. Lack and type of training could also explain the low rates of use of the tablet PCs by nurses, given that poor user training may relate to poor technology acceptance [31]. Nurses received formal training in the PowerChart system away from the wards on stationary PCs, which are more comparable to laptops than tablet PCs, since navigation on tablet PCs is done using a stylus. Although tablet PCs had the longest boot-up times, this wait was not reported as a problem by clinicians and did not appear to have an influence on device selection.

The Lack of Mobility of "Mobile" Computing Devices

Although many of the devices are intended to be mobile, it is difficult for them to match the mobility of paper [28]. Five nurses and one doctor were observed transcribing details about medication onto a napkin or paper. The "copy" of the medication chart was then used by nurses in the medication room to prepare drugs for one or several patients at a time, despite the availability of a stationary PC in the medication room. Copying details from the computerized medication chart is another example of a workaround resulting from the hardware devices failing to support adequately clinical workflow [32]. This workaround is inexpedient and can lead to errors [33]. This finding suggests that despite the necessary information being available on the stationary PC located in the medication room and users being able to log in to the PowerChart system at two different computers simultaneously, nurses find it is easier to copy information onto paper and, thus, the system is not being used as intended.

Collaboration and Device Selection

Nurses, doctors on ward rounds, and doctors not on ward rounds collaborated mostly on the device each group preferred. Both groups of doctors collaborated when viewing results, whereas nurses mostly collaborated in order to witness the preparation of drugs. This indicates that collaboration is not the primary reason for choice of device. The study also showed that doctors mostly collaborated with other doctors, and nurses with other nurses, which is consistent with previous findings [24,34].

Implications of the Results for Health Care Planners

The results provide some key information which may assist health planners in other hospitals to plan and evaluate their computer hardware device needs. We found that the number of COWs (five in one ward and six in the other) in each 26-bed ward appeared adequate for the tasks required. However, the number of clinical functions to be undertaken on computer devices should be considered. In our study, paper records were still used to record clinical progress notes. If this function was computerized, demand for computers would be expected to increase. Physical space limitations on a ward, including access to a power source, need to be considered when deciding upon the number of COWs which can be accommodated. This assessment of space should consider how COWs will move within patient and treatment rooms on the ward. Important features of COWs identified include the capacity to transport items such as paper charts and medications, and easy manoeuvrability. Demand for COWs was greatest during ward rounds; thus, the frequency and timing of ward rounds is a further consideration in determining the number of COWs required.

We found that despite the availability of two tablet PCs on both wards they were substantially underutilized. Interviews with nurses revealed that lack of training, specifically in the use of tablet PCs, may be a reason for nurses' reluctance to use them more frequently. System training in hospitals usually occurs on stationary PCs with the assumption that skills will be transferable for use on other devices. Our results suggest this may not be the case, and utilization of tablet PCs might increase with specific training. Tablet PCs with handles were more frequently used, and this finding suggests that this is an important design feature to consider during hardware selection.

While there is a constant emphasis on allowing clinicians to access information at a patient's bedside, we found that only a relatively small proportion of clinical tasks were completed at this location with more activity occurring in the corridors. This result reinforces the need for mobile devices and suggests that stationary PCs at each patient's bedside may not be a solution to clinical information needs.

Limitations

Our findings only relate to device selection in two wards in one hospital between the weekday hours of 7 am and 4 pm, and they may not be generalizable to other settings or times. The type of medication distribution system, namely the storage of most medicines in patients' bedside drawers, may have influenced device preferences and, subsequently, the results. We were only able to look at devices available in the study wards and, thus, hardware devices such as personal digital assistants (PDAs) were not considered. Our hospital was still using paper-based records for some functions such as clinical progress notes. If a fully computerized patient record was in use, the need for trolley space to place paper charts would be reduced, and there would be an increased demand for access to hardware devices. Increasing the use of tablet PCs might be one solution to this demand, given that they make a reduced claim on the limited physical space available in most wards. We were only able to interview a small proportion of clinical staff to obtain data to inform the observational study results. We relied upon a convenience sample, and we cannot be sure of the representativeness of their views. However, the responses of these participants in relation to preference for specific hardware devices were consistent with the overall observational findings. We did not interview staff immediately after each observation session because this was deemed as potentially disruptive to clinical care on the ward. Additionally, only two assessors

completed the heuristic evaluation, instead of the recommended minimum of three.

Data were entered into the tablet using a stylus to select from dropdown menus, press buttons, and enter text using a graphical keyboard interface on the tablet. Handwriting recognition may have been part of the tablet's functionality, but the clinical information systems accessed using the tablet did not allow handwriting recognition. Thus, we were not able to assess this feature. The provision of specific training for nurses in the use of tablet PCs may increase their appeal.

Conclusions

Nurses' medication-related tasks require high levels of mobility, and computer devices need to support this mobility. Nurses move from patient to patient and back and forth between patient rooms and the medication room. All nurses preferred the generic COW independent of the clinical task conducted, particularly as it allowed them to carry other items such as paper records at the same time. However, we found evidence that mobile devices sometimes limit nurses' mobility. In response, nurses may initiate workarounds, such as transcribing information from computer to paper rather than carrying a computer device or logging onto an available stationary PC in another location.

Doctors' choice of device is dependent on whether or not they are on a ward round. On ward rounds, doctors move between patients and back and forth between the patient room and the corridor and, thus, require a mobile device. While mobile devices are designed to allow greater provision of information at the point of care, we found over half of observed ward round tasks were performed in corridors and away from patients' rooms. The results indicate that a doctor's choice of mobile device is individual and dependent on device design. Doctors not on ward rounds tend to conduct clinical tasks in the same spot, and their device of choice is the stationary PC in an office. Tablet PCs with handles were preferred to those without.

In selecting hardware devices, consideration should be given to those who will be using the system, and the nature and location of their clinical tasks, including whether ward rounds are a frequent occurrence. Furthermore, the extent to which the physical layout of a ward will accommodate different types of stationary and mobile computing devices should be considered. Devices which allow clinicians to provide care close to a patient's bedside, but which are also easily manoeuvred to other ward locations, may reduce the initiation of potentially unsafe workaround practices.

Acknowledgments

This research was supported by an ARC Linkage grant (LP 0989144). NC was supported by a Faculty of Health Sciences post-doctoral fellowship. PA and AL were on an exchange program at the Health Informatics Research & Evaluation Unit, Faculty of Health Sciences, University of Sydney, Australia at the time of conducting the research. We thank the staff at the study hospital for their participation.

Conflicts of Interest

None declared.

References

1. Berg M, Winthereik BR. Waiting for Godot: episodes from the history of patient records. In: Berg M, editor. Health Information Management: Integrating Information Technology in Health Care Work. London, UK: Routledge; 2004:11-44.
2. Anderson GF, Frogner BK, Johns RA, Reinhardt UE. Health care spending and use of information technology in OECD countries. Health Aff (Millwood) 2006;25(3):819-831. [Medline: 16684749] [doi: 10.1377/hlthaff.25.3.819]
3. Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. Int J Med Inform 2008 May;77(5):291-304. [Medline: 17951106] [doi: 10.1016/j.ijmedinf.2007.09.001]
4. Nightingale PG, Adu D, Richards NT, Peters M. Implementation of rules based computerised bedside prescribing and administration: intervention study. BMJ 2000 Mar 18;320(7237):750-753 [FREE Full text] [Medline: 10720357] [doi: 10.1136/bmj.320.7237.750]
5. Richardson WC, Berwick DM, Bisgard JC. Using information technology. In: Briere R, editor. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press; 2001:164-180.
6. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006;144(10):742-752 [FREE Full text] [Medline: 16702590]
7. Bardram JE. I love the system — I just don't use it. In: Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work: The Integration Challenge (Phoenix, Arizona, USA). New York, NY: ACM Press; 1997 Presented at: p. 251-260.
8. Han YY, Carcillo JA, Venkataraman ST, Clark RSB, Watson RS, Nguyen TC, et al. Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. Pediatrics 2005 Dec;116(6):1506-1512 [FREE Full text] [Medline: 16322178] [doi: 10.1542/peds.2005-1287]

9. Pratt W, Reddy MC, McDonald DW, Tarczy-Hornoch P, Gennari JH. Incorporating ideas from computer-supported cooperative work. J Biomed Inform 2004 Apr;37(2):128-137. [Medline: 15120659] [doi: 10.1016/j.jbi.2004.04.001]
10. Reddy MC, Dourish P, Pratt W. Coordinating heterogeneous work: information and representation in medical care. In: Printz W, Jarke M, Rogers Y, Schmidt K, Wulf V, editors. Proceedings of the Seventh European Conference on Computer Supported Cooperative Work. Bonn, Germany: Kluwer Academic Publishers; 2001 Presented at: p. 239-258.
11. Reuss E, Menozzi M, Büchi M, Koller J, Krueger H. Information access at the point of care: what can we learn for designing a mobile CPR system? Int J Med Inform 2004 May;73(4):363-369. [Medline: 15135755] [doi: 10.1016/j.ijmedinf.2004.02.003]
12. Sellen AJ, Harper RHR. The Myth of the Paperless Office. Cambridge, MA: MIT Press; 2002.
13. Luff P, Heath C. Mobility in collaboration. In: Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work (Seattle, Washington, USA). New York, NY: ACM Press; 1998 Presented at: p. 305-314.
14. Lapinsky SE, Holt D, Hallett D, Abdolell M, Adhikari NKJ. Survey of information technology in Intensive Care Units in Ontario, Canada. BMC Med Inform Decis Mak 2008;8(1):5 [FREE Full text] [Medline: 18218117] [doi: 10.1186/1472-6947-8-5]
15. Lu YC, Xiao Y, Sears A, Jacko JA. A review and a framework of handheld computer adoption in healthcare. Int J Med Inform 2005 Jun;74(5):409-422. [Medline: 15893264] [doi: 10.1016/j.ijmedinf.2005.03.001]
16. Ammenwerth E, Buchauer A, Bludau B, Haux R. Mobile information and communication tools in the hospital. Int J Med Inform 2000 Jan;57(1):21-40. [Medline: 10708253] [doi: 10.1016/S1386-5056(99)00056-8]
17. Baumgart DC. Personal digital assistants in health care: experienced clinicians in the palm of your hand? Lancet 2005 Oct 1;366(9492):1210-1222. [Medline: 16198770] [doi: 10.1016/S0140-6736(05)67484-3]
18. Fischer S, Stewart TE, Mehta S, Wax R, Lapinsky SE. Handheld computing in medicine. J Am Med Inform Assoc 2003;10(2):139-149 [FREE Full text] [Medline: 12595403] [doi: 10.1197/jamia.M1180]
19. Lindquist AM, Johansson PE, Petersson GI, Saveman BI, Nilsson GC. The use of the Personal Digital Assistant (PDA) among personnel and students in health care: a review. J Med Internet Res 2008;10(4):e31 [FREE Full text] [Medline: 18957381] [doi: 10.2196/jmir.1038]
20. Lu YC, Xiao Y, Sears A, Jacko JA. A review and a framework of handheld computer adoption in healthcare. Int J Med Inform 2005 Jun;74(5):409-422. [Medline: 15893264] [doi: 10.1016/j.ijmedinf.2005.03.001]
21. Martins HM, Jones MR. What's so different about mobile information communication technologies (MICTs) for clinical work practices? A review of selected pilot studies. Health Inform J 2005;11(2):123-134 [FREE Full text] [doi: 10.1177/1460458205052361]
22. Westbrook JI, Braithwaite J, Georgiou A, Ampt A, Creswick N, Coiera E, et al. Multimethod evaluation of information and communication technologies in health in the context of wicked problems and sociotechnical theory. J Am Med Inform Assoc 2007;14(6):746-755 [FREE Full text] [Medline: 17712083] [doi: 10.1197/jamia.M2462]
23. Westbrook JI, Ampt A, Williamson M, Nguyen K, Kearney L. Methods for measuring the impact of health information technologies on clinicians' patterns of work and communication. Stud Health Technol Inform 2007;129(Pt 2):1083-1087. [Medline: 17911882]
24. Westbrook JI, Ampt A, Kearney L, Rob MI. All in a day's work: an observational study to quantify how and with whom doctors on hospital wards spend their time. Med J Aust 2008 May 5;188(9):506-509 [FREE Full text] [Medline: 18459920]
25. Nielsen J. Ten usability heuristics. useit.com: usable information technology. 2005. URL: http://www.useit.com/papers/heuristic/heuristic_list.html [accessed 2008 Dec 4] [WebCite Cache ID 5iC2Cs1Ro]
26. Han SH, Hong SW. A systematic approach for coupling user satisfaction with product design. Ergonomics 2003;46(13-14):1441-1461. [Medline: 14612330] [doi: 10.1080/00140130310001610928]
27. Krogh PR, Rough S, Thomley S. Comparison of two personal-computer-based mobile devices to support pharmacists' clinical documentation. Am J Health Syst Pharm 2008 Jan 15;65(2):154-157. [Medline: 18192262] [doi: 10.2146/ajhp070177]
28. Bogossian FE, Kellett SEM, Mason B. The use of tablet PCs to access an electronic portfolio in the clinical setting: a pilot study using undergraduate nursing students. Nurse Educ Today 2009 Feb;29(2):246-253. [Medline: 18945525] [doi: 10.1016/j.nedt.2008.09.001]
29. Lium JT, Tjora A, Faxvaag A. No paper, but the same routines: a qualitative exploration of experiences in two Norwegian hospitals deprived of the paper based medical record. BMC Med Inform Decis Mak 2008;8(1):2 [FREE Full text] [Medline: 18186935] [doi: 10.1186/1472-6947-8-2]
30. Silvey GM, Lobach DF, Macri JM, Hunt M, Kacmaz RO, Lee PP. User interface considerations for collecting data at the point of care in the tablet PC computing environment. AMIA Annu Symp Proc 2006:1096 [FREE Full text] [Medline: 17238715]
31. Ammenwerth E, Gräber S, Herrmann G, Bürkle T, König J. Evaluation of health information systems-problems and challenges. Int J Med Inform 2003 Sep;71(2-3):125-135. [Medline: 14519405] [doi: 10.1016/S1386-5056(03)00131-X]
32. Vogelsmeier AA, Halbesleben JRB, Scott-Cawiezell JR. Technology implementation and workarounds in the nursing home. J Am Med Inform Assoc 2008;15(1):114-119 [FREE Full text] [Medline: 17947626] [doi: 10.1197/jamia.M2378]
33. Khajouei R, Jaspers MWM. CPOE system design aspects and their qualitative effect on usability. Stud Health Technol Inform 2008;136:309-314. [Medline: 18487749]

34. Creswick N, Westbrook JI. Social network analysis of medication advice-seeking interactions among staff in an Australian hospital. Int J Med Inform 2008 Nov 11. [Medline: 19008147] [doi: 10.1016/j.ijmedinf.2008.08.005]

Abbreviations

COW: computer on wheels
CPOE: computerized provider order entry
ICT: information and communication technologies
MAR: medication administration record
PC: personal computer
PDA: personal digital assistant

Edited by G Eysenbach; submitted 19.02.09; peer-reviewed by S Koch, R La Tella; comments to author 21.04.09; revised version received 08.05.09; accepted 22.06.09; published 04.08.09.

Please cite as:
Andersen P, Lindgaard AM, Prgomet M, Creswick N, Westbrook JI
Mobile and Fixed Computer Use by Doctors and Nurses on Hospital Wards: Multi-method Study on the Relationships Between Clinician Role, Clinical Task, and Device Choice
J Med Internet Res 2009;11(3):e32
URL: http://www.jmir.org/2009/3/e32/
doi:10.2196/jmir.1221
PMID:19674959

Original Paper

Potential Benefits and Harms of a Peer Support Social Network Service on the Internet for People With Depressive Tendencies: Qualitative Content Analysis and Social Network Analysis

Yoshimitsu Takahashi[1,2], MS; Chiyoko Uchida[3], MD, PhD; Koichi Miyaki[1,4], MD, PhD; Michi Sakai[2,5], MPH; Takuro Shimbo[2], MD; Takeo Nakayama[1], MD, PhD

[1]Department of Health Informatics, Kyoto University School of Public Health, Kyoto, Japan

[2]Department of Clinical Research and Informatics, Research Institute, International Medical Center of Japan, Tokyo, Japan

[3]University Health Center Ibaraki University, Ibaraki, Japan

[4]Department of Neurology, School of Medicine, Keio University, Tokyo, Japan

[5]Department of Epidemiology and Healthcare Research, Kyoto University School of Public Health, Kyoto, Japan

Corresponding Author:
Takeo Nakayama, MD, PhD
Kyoto University School of Public Health
Yoshida Konoe Sakyo
Kyoto 606-8501
Japan
Phone: +81 75 753 4488
Fax: +81 75 753 4497
Email: nakayama@pbh.med.kyoto-u.ac.jp

Abstract

Background: Internet peer support groups for depression are becoming popular and could be affected by an increasing number of social network services (SNSs). However, little is known about participant characteristics, social relationships in SNSs, and the reasons for usage. In addition, the effects of SNS participation on people with depression are rather unknown.

Objective: The aim was to explore the potential benefits and harms of an SNS for depression based on a concurrent triangulation design of mixed methods strategy, including qualitative content analysis and social network analysis.

Methods: A cross-sectional Internet survey of participants, which involved the collection of SNS log files and a questionnaire, was conducted in an SNS for people with self-reported depressive tendencies in Japan in 2007. Quantitative data, which included user demographics, depressive state, and assessment of the SNS (positive vs not positive), were statistically analyzed. Descriptive contents of responses to open-ended questions concerning advantages and disadvantages of SNS participation were analyzed using the inductive approach of qualitative content analysis. Contents were organized into codes, concepts, categories, and a storyline based on the grounded theory approach. Social relationships, derived from data of "friends," were analyzed using social network analysis, in which network measures and the extent of interpersonal association were calculated based on the social network theory. Each analysis and integration of results were performed through a concurrent triangulation design of mixed methods strategy.

Results: There were 105 participants. Median age was 36 years, and 51% (36/71) were male. There were 37 valid respondents; their number of friends and frequency of accessing the SNS were significantly higher than for invalid/nonrespondents (P = .008 and P = .003). Among respondents, 90% (28/31) were mildly, moderately, or severely depressed. Assessment of the SNS was performed by determining the access frequency of the SNS and the number of friends. Qualitative content analysis indicated that user-selectable peer support could be passive, active, and/or interactive based on anonymity or ease of use, and there was the potential harm of a downward depressive spiral triggered by aggravated psychological burden. Social network analysis revealed that users communicated one-on-one with each other or in small groups (five people or less). A downward depressive spiral was related to friends who were moderately or severely depressed and friends with negative assessment of the SNS.

Conclusions: An SNS for people with depressive tendencies provides various opportunities to obtain support that meets users' needs. To avoid a downward depressive spiral, we recommend that participants do not use SNSs when they feel that the SNS is

not user-selectable, when they get egocentric comments, when friends have a negative assessment of the SNS, or when they have additional psychological burden.

(J Med Internet Res 2009;11(3):e29) doi:10.2196/jmir.1142

KEYWORDS

Social network service (SNS); Internet; depression; peer support; social network; qualitative method; content analysis; mixed methods

Introduction

Mental health topics are especially popular on the Internet, and there are high levels of untreated and undiagnosed depression in users of Internet depression communities [1]. Depression is an important global public health issue, and an effective approach to prevention may involve the targeted screening of people with chronic diseases as well as those with social isolation [2-5]. Peer support—providing support based on mutual counseling and the sharing of information and experience—is becoming an increasingly important strategy in health care environments that face shrinking financial and human resources [6,7]. Although many Internet-based support groups have emerged, a systematic review did not find conclusive evidence concerning the health benefits of virtual communities and peer-to-peer online support [8]. The benefits of Internet-based support groups for depression have been assessed in two independent studies [9,10].

We have also witnessed a new revolution in the field of communication through the Internet, called Web 2.0 [11]. Social network services (SNSs), such as MySpace or Facebook, are Internet-based Web 2.0 applications that allow the building of online social networks where individuals can share interests and activities. SNS users have increased in number all over the world. Moreover, SNSs for specific health-related purposes have emerged (eg, for quitting smoking or for people with cancer or a mental health disorder). A few studies have examined the risks of MySpace for adolescents [12-15]. However, no SNS-related studies on depression or peer support existed in PubMed as of February 2009. SNSs have thus far been used without knowledge of their benefits and harms.

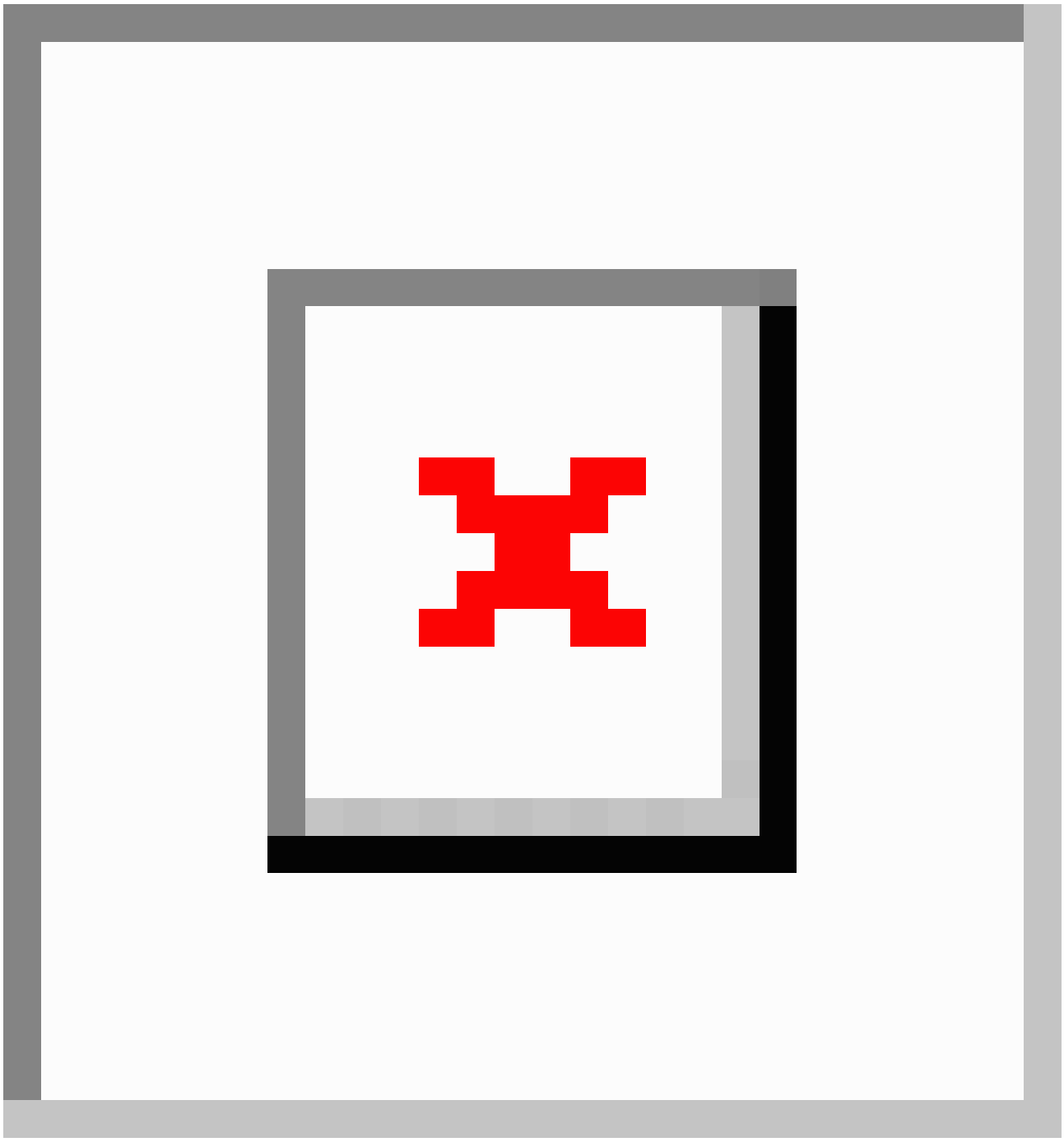
These social backgrounds have prompted the hypothesis that SNS users with depressive tendencies may wish to seek out their peers and information on depression from their peers. Inherent in this, however, are the detrimental effects of Internet addiction [16] and psychological distress [17]. Some recent studies have shown that obesity, smoking, and happiness can spread through social networks in what is termed a “network phenomenon” [18-20]. This may suggest that negative effects of SNSs can be exaggerated by Internet social networks. Therefore, we sought to address the following questions: (1) What are the characteristics of SNS participants? (2) What kinds of social relationships are there in SNSs? and (3) Why and how do participants use SNSs? These questions are both quantitative and qualitative. Hence, we quantitatively examined characteristics and social relationships and qualitatively examined reasons for usage and the process of SNS usage. Finally, this study aimed to explore the potential benefits and harms of SNSs for people with depressive tendencies in Japan through a concurrent triangulation design of mixed methods strategy (combining quantitative and qualitative methods) and to produce converging findings from both methods [21-24].

Methods

Study Setting

An SNS was launched by an individual in April 2006 for people in Japan with self-reported depressive tendencies (Figure 1). Registrants in the SNS described themselves as (1) people who were / had been depressed, (2) people who had family or a friend who was depressed, (3) people who had concerns about depression, or (4) psychiatrists or clinical psychotherapists. They had voluntarily registered in the SNS, which was free of charge and had standard SNS functions. Generally, SNSs in Japan have the following functions: an “invitation” function, a “footprints” function, a “privacy controls” function, and a function to access the SNS from cell phones.

Figure 1. Screenshot of the SNS home page for people with depressive tendencies (Friends: online friends whose profiles are featured as links on one's own profile. Communities: online communities for people with similar interests or activities. Messages: online messages, such as Web-based email. Information: displays information from the administrator. Invitation function: users who want to participate need an invitation from a participant for registration. Footprints function: participants can ascertain and access the history of another participant. Privacy controls function: participants can choose who can view their profile or contact them.)



Study Design and Participants

The study was an observational cross-sectional study in which we conducted two different surveys. In the first survey, the administrator extracted SNS log files, which are electronic records from the SNS database. In the second survey, an Internet questionnaire survey of participants was conducted during March and April 2007. After developing a Secure Sockets Layer compatible website and pretesting for usability and technical functionality, unique IDs and passwords were individually sent to participants through "messages" in the SNS (a nonopen, password-protected survey). Research methods have been presented in compliance with the checklist for reporting results of Internet e-surveys (CHERRIES) [25].

We addressed potential sources of bias. Given that participants were SNS users, we tried to contact them only through the SNS to avoid causing changes in the participants' behavior. In order to evaluate the reliability of the Internet survey, we compared responses on the questionnaire to data from SNS log files regarding age, gender, and frequency of access. Participants

were required to list their diagnoses and medications, and these data were confirmed by a psychiatrist (C Uchida).

Measures

SNS log files included access logs and data in "profiles," "blogs," "communities," and "friends" (see Figure 1). "Friends" included data of the participant's network of "friends" (see the Glossary in Multimedia Appendix 1). The questionnaire for the Internet survey included a total of 67 items related to characteristics, activities, and outcomes of four Web pages. Items related to characteristics included age, gender, time of Internet use, job status, whether or not the person lived alone, diagnosis of mood disorders, and assessment of mood states by the Zung Self-Rating Depression Scale and a numerical rating scale. The depression scale ranged from 20 (no depression) to 80 (major depression) and was categorized as follows: not depressed (≤ 39), mildly depressed (40-47), moderately depressed (48-55), and severely depressed (≥ 56) [26,27]. Mood states were measured by a numerical scale that ranged from 0 (good mood) to 10 (bad mood) [28] in the following three situations: (1) during normal time (not using the SNS or Internet), (2) while using the Internet, and (3) while using the SNS. Items related to activities included profile display, frequency of accessing the SNS, frequency of updating blogs (three times or more per week vs less than three times per week), number of communities, and number of friends. Items related to outcomes in the assessment of the SNS were evaluated by the question, "Do you feel your sense of illness management for depression improved compared to before participating in the SNS?" Response choices were "much more," "more," "no change," or "less." "Much more" and "more" were classified as "positive assessment," and the other responses were classified as "not positive assessment." Moreover, the descriptive open-ended question "What are the advantages and disadvantages of participating in the SNS?" was included to explore potential effects of the SNS. Answers to the Internet questionnaire were checked for consistency and saved page by page. All saved data, both complete and incomplete, were analyzed.

Framework of Analytic Methodology

We used the concurrent triangulation design of mixed methods strategy to analyze both quantitative and qualitative data with qualitative priority [23,24]. First, we examined participant characteristics using statistical analyses (see Step 1 below). Second, we examined reasons for SNS usage and the process of SNS usage with qualitative content analysis [29,30] (see Step 2). Third, we examined the relationships between participants using social network analysis [31-34] (see Step 3). After quantitative and qualitative research questions were examined, these results were integrated based on the mutual validation model, which regards the search for convergent findings deemed to be validity indicators as the most important purpose of triangulation [23]. We explored potential benefits and harms using qualitative results, while we inferred the extent of the benefits and harms using quantitative results. Discrepancies in the results were interpreted and discussed at the discussion session. Multimedia Appendix 2 provides a summary of the framework of analytic methodology.

Step 1: Statistical Analyses

To compare variables between groups, we used the Fisher exact test for two categorical variables, Pearson chi-square test for three or more categorical variables, and Student *t* test for continuous variables. The Mann-Whitney U test was used for age, communities, friends, and centrality given that the distribution was estimated to be skewed. Valid respondents, who answered any item in the questionnaire, were compared to invalid/nonrespondents, and people with a "positive" assessment of the SNS were compared to people with a "not positive" assessment. All statistical analyses were performed using SPSS 15.0J (SPSS Inc, Chicago, IL, USA). All P values were 2-sided, with $P < .05$ considered statistically significant.

Step 2: Qualitative Content Analysis

Descriptive contents of responses to open-ended questions concerning advantages and disadvantages of SNS participation were analyzed using the inductive approach of qualitative content analysis [29,30]. In content analysis, it is assumed that words and phrases that are mentioned often reflect important concerns [35]. However, contents can involve multiple meanings and be latent as well as manifest [29]. In order to achieve trustworthiness [30], contents were inductively organized into codes, concepts, categories, and a storyline based on the grounded theory approach, a commonly used systematic qualitative research method to generate theories regarding social phenomena [21,22,36].

The analysis was performed by multidisciplinary members: a nurse, a pharmacist, and two public health researchers (Y Takahashi and M Sakai). All responses were read and interpreted repeatedly. After discussing the meanings of responses, a coding frame was developed and sentences were coded for analysis. If new codes emerged, the coding frame was changed and sentences were re-read according to the new structure. This constant comparison process was also used to develop concepts, which were then conceptualized into broad categories after further discussion. After discussing the relationship of codes, concepts, and categories, we generated a storyline. We used ATLAS.ti 5.2 (Scientific Software Development GmbH, Berlin, Germany) for data analysis.

Step 3: Social Network Analysis

Data on social relationships among participants were analyzed based on the social network theory [34], in which people were defined as nodes, and relationships were defined as linkages among nodes [31-34]. Social network analysis is the study of social structure that provides a means to quantitatively explore social relationships between people. Commonly used by sociologists, its use in health-related fields is increasing as an effective approach for research centered on describing, exploring, and understanding the relational aspects of health [18-20,32,33]. Network measures and the extent of interpersonal association were calculated based on the theory. The overall social network was graphed with the Kamada-Kawai algorithm using Pajek 1.20 software (University of Ljubljana, Slovenia) from data in "friends" [18-20], according to the sociocentric network approach [31,32]. To identify characteristics of an individual within a network, centrality was measured by (1)

degree, (2) closeness, and (3) betweenness [31-33]. These measures of centrality identify the most prominent individual. Degree refers to the sum of individuals who are linked together. Closeness refers to the distance between individuals. Betweenness refers to the number of times an individual connects pairs of others; people with high betweenness centrality are able to play a gatekeeper role, controlling the flow of resources in the network. To examine characteristics of a group, cliques were counted that included three or more individuals connected by all possible connections. These network measures were analyzed using UCINET 6.1 (Analytic Technologies, Lexington, KY, USA).

In order to evaluate the extent of interpersonal association as a whole (whether people with similar attributes tended to be connected with each other or not), we counted numbers of connections sorted by type, including assessment of the SNS (positive vs not positive), depressive state (moderately or severely depressed vs not moderately or severely depressed), and frequency of accessing the SNS (three times or more per week vs less than three times per week). Then we computed odds ratios and 95% confidence intervals. For example, assessment of the SNS was defined as follows: P was a person with a "positive" assessment of the SNS; p was a friend with "positive" assessment; N was a person with "not positive" assessment; n was a friend with "not positive" assessment; p_P was the number of friends with "positive" assessment for people with a "positive" assessment. The odds ratio for assessment of the SNS was computed by (p_P / n_P) / (p_N / n_N).

Ethical Considerations

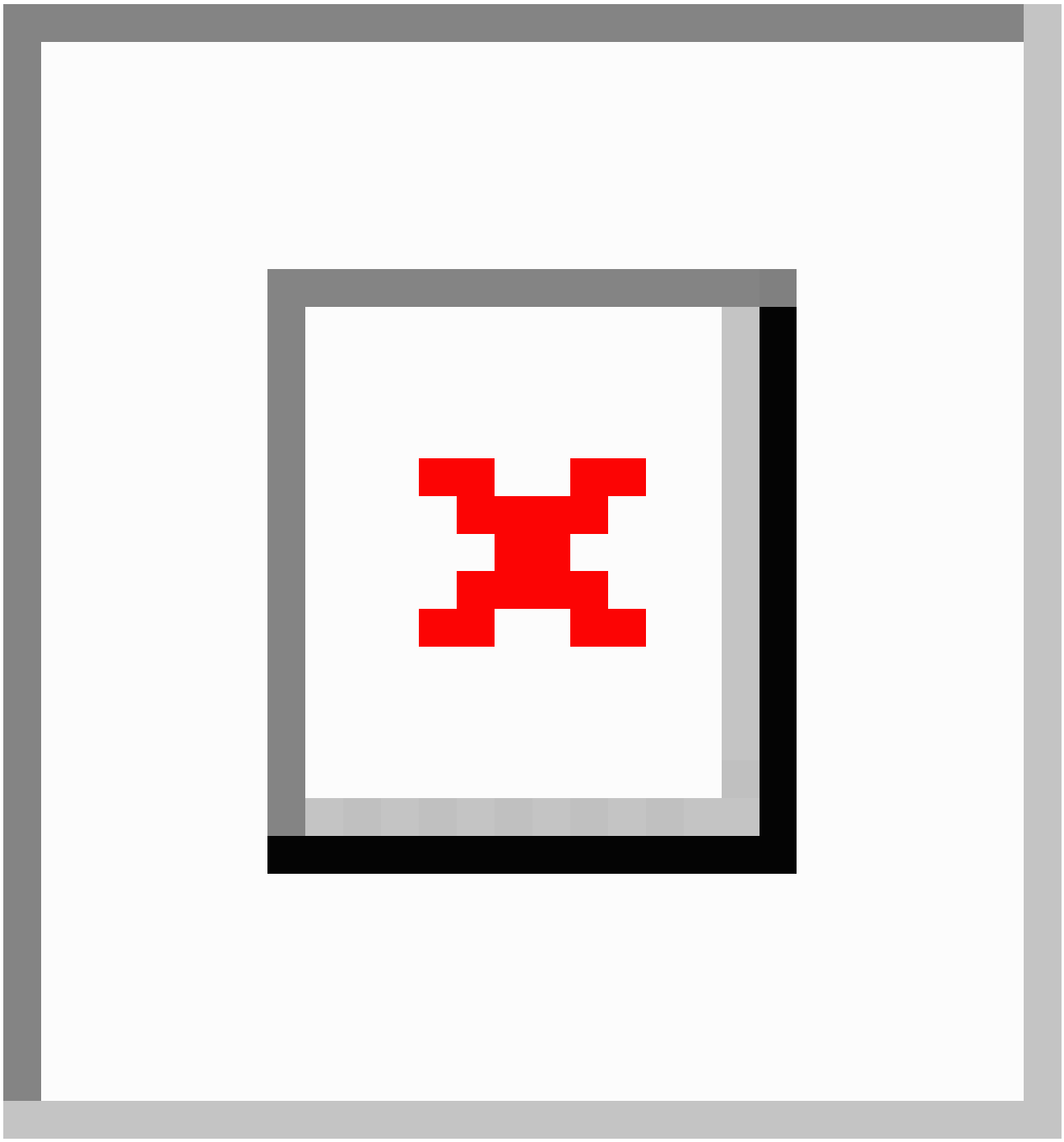
We prepared a site that explains the study [37]. We also declared that we collected the SNS log files without personal information and provided an opportunity for refusal (opt-out recruitment). Informed consent was requested from all participants on the first page of the questionnaire (opt-in recruitment). Quantitative data without personal information in SNS log files were collected and analyzed. The questionnaire and SNS log files were anonymized in a linkable fashion. The study protocol was approved by the Ethics Committee of Kyoto University Faculty of Medicine (No. E254, January 12, 2007).

Results

Statistical Analyses

Of the total registrants (N = 116), eight people withdrew. Three people were excluded because one was the administrator and two had registered within the week before collection of data from the SNS log files. Participants (N = 105) were the subjects for the SNS log file analysis. Among the 40 people who responded to the questionnaire, three people were excluded because no item was answered. Valid respondents (N = 37) were the subjects for the questionnaire analysis (Figure 2).

Figure 2. Flow diagram of subjects surveyed and analyzed: “Participants” were subjects of the SNS log file analysis; “Valid respondents” were subjects of the Internet questionnaire analysis; “Withdrew” refers to people who had deleted their accounts themselves; “Exclusion” refers to people who were excluded (one was the administrator, and two had registered within the week before collection of data from the SNS log files); “Respondents” were people who provided informed consent; “Valid respondents” were people who answered any item of the questionnaire; “Invalid respondents” were people who did not answer any item, although they provided informed consent. † is subjects of the SNS log file analysis. ‡ is subjects of the Internet questionnaire survey analysis.



Participants (N = 105) had the following characteristics (Table 1): median age was 36 years (range 21-57), 36/71 (51%) were male, and 47/102 (46%) logged in three times or more per week. Frequency of accessing the SNS by personal computer or cell phone, frequency of updating blogs, and “friends” of valid respondents were significantly higher compared to invalid/nonrespondents (P = .003, P = .02, and P = .008, respectively).

Table 1. Characteristics of participants (N = 105)

	n[a]	Total (N = 105)	Response to the Questionnaire		*P*
			Valid Respondents (N = 37)	Invalid/Nonrespondents[b] (N = 68)	
Characteristic					
Age in years, median (range)	105	36 (21-57)	37 (21-52)	33 (22-57)	.09[c]
Male, n (%)	71	36 (51)	16 (43)	20 (59)	.24[d]
Profile,[e] n (%)	105	97 (92)	34 (92)	63 (93)	.99[d]
Accessing the SNS,[f] n (%)					
By personal computer or cell phone	102	47 (46)	24 (67)	23 (35)	.003[d]
By cell phone	102	6 (6)	1 (3)	5 (8)	.99[d]
Updating blogs,[f] n (%)	102	16 (16)	11 (31)	5 (8)	.02[d]
Communities, median (range)	105	3 (2-15)	2 (2-12)	3 (2-15)	.051[c]
Friends, median (range)	105	2 (0-42)	7 (0-27)	2 (0-42)	.008[c]
Relationship (social network analysis)					
Centrality, median (range)					
Degree	105	2 (0-42)	7 (0-27)	2 (0-42)	.008[c]
Closeness	105	2011 (1925-2177)	1992 (1948-2177)	2011 (1925-2095)	.02[c]
Betweenness	105	1.1 (0-1718)	27.2 (0-960)	0 (0-1718)	.002[c]

[a] Several items included missing data.

[b] Invalid respondents (N = 3); nonrespondents (N = 65).

[c] Mann-Whitney test.

[d] Fisher exact test.

[e] Number of people who had written their profile.

[f] Three times or more per week.

Table 2 shows characteristics of valid respondents (N = 37). Median age was 37 years (range 21-52), and 16/37 people (43%) were male. Moreover, 32/35 people (91%) could be diagnosed with mood disorders, 28/31 (90%) were mildly, moderately, or severely depressed, and 19/37 (54%) had a "positive" assessment of the SNS. The frequency of accessing the SNS by personal computer or cell phone and "friends" of people with a "positive" assessment of the SNS were significantly higher than for people with a "not positive" assessment (*P* = .02 and *P* = .01, respectively). When comparing the mood state (measured by the numerical scale) while using the SNS and during normal time, people with a "positive" assessment of the SNS showed a greater improvement in their mood than people with a "not positive" assessment (*P* = .07).

Table 2. Characteristics and outcomes of valid respondents (N = 37)

	n[a]	Total	Assessment of the SNS[b]			*P* [c]
		(N = 37)	Positive (N = 19)	Not Positive (N = 16)	No Assessment (N = 2)	
Characteristic						
Age, median (range)	37	37 (21-52)	37 (26-51)	32 (21-52)	44.5 (40-49)	.30[d]
Male, n (%)	37	16 (43)	8 (42)	7 (44)	1 (50)	.99[e]
Internet use (per week), n	35					
40+ hours		3	2	1	0	.66[f]
10-39 hours		19	8	9	2	
≤ 9 hours		13	8	5	0	
Not working, no (%)	34	16 (47)	11 (58)	5 (33)	0 (0)	.19[e]
Living alone, n (%)	17	5 (29)	2 (18)	3 (50)	0 (0)	.28[e]
Diagnosis, n (%)	35	32 (91)	19 (100)	12 (80)	1 (100)	.08[e]
Medication, n (%)	37	27 (73)	17 (89)	10 (63)	0 (0)	.11[e]
Profile,[g] n (%)	37	34 (92)	17 (89)	15 (94)	2 (100)	.99[e]
Accessing the SNS,[h] n (%)						
By personal computer or cell phone	36	24 (67)	16 (89)	8 (50)	0 (0)	.02[e]
By cell phone	36	1 (3)	1 (6)	0 (0)	0 (0)	.99[e]
Updating blogs,[h] n (%)	36	11 (31)	6 (33)	5 (31)	0 (0)	.99[e]
Communities, median (range)	37	2 (2-12)	2 (2-12)	2 (2-6)	3 (2-4)	.71[d]
Friends, median (range)	37	7 (0-27)	8 (1-21)	2 (0-27)	2 (1-3)	.01[d]
Relationship (social network analysis)						
Centrality, median (range)						
Degree	37	7 (0-27)	8 (1-21)	2 (0-27)	2 (1-3)	.01[d]
Closeness	37	1992 (1948-2177)	1977 (1951-2177)	2008 (1948-2034)	2056 (2034-2078)	.03[d]
Betweenness	37	27 (0-960)	74 (0-459)	3 (0-960)	0 (0-0)	.02[d]
Outcome						
Depressive state,[i] mean ± SD	31	50.9 ± 9.7	52.8 ± 8.7	50.4 ± 9.0	38.5 ± 19.1	.47[j]
Not depressed, n		3	0	2	1	.29[f]
Mildly depressed, n		6	5	1	0	
Moderately depressed, n		12	5	6	1	
Severely depressed, n		10	6	4	0	
Mood state,[k] mean ± SD						
(A) During normal time	31	5.6 ± 2.3	6.2 ± 2.1	5.5 ± 2.1	1.5 ± 2.1	.41[j]
(B) While using the Internet	31	4.9 ± 2.4	5.6 ± 2.2	4.7 ± 2.3	1.5 ± 2.1	.31[j]
(C) While using the SNS	31	5.1 ± 2.0	5.0 ± 2.0	5.6 ± 1.6	2.0 ± 2.8	.37[j]
Difference (B) − (A)	31	−0.7 ± 2.0	−0.6 ± 2.3	−0.8 ± 1.9	0.0 ± 0.0	.78[j]

	n[a]	Total (N = 37)	Assessment of the SNS[b]			*P* [c]
			Positive (N = 19)	Not Positive (N = 16)	No Assessment (N = 2)	
Difference (C) − (A)	31	−0.5 ± 1.9	−1.2 ± 2.3	0.1 ± 1.2	0.5 ± 0.7	.07[j]

[a] Several items included missing data.

[b] Assessment of the SNS was evaluated by the question "Do you feel your sense of illness management for depression improved compared to before participating in the SNS?" Response choices were "much more," "more," "no change," or "less." "Much more" and "more" were classified as "positive assessment," and the other responses were classified as "not positive assessment."

[c] Comparing positive and not positive assessments of the SNS.

[d] Mann-Whitney test.

[e] Fisher exact test.

[f] Pearson chi-square test.

[g] Number of people who had written their profile.

[h] Three times or more per week.

[i] Depressive states were measured by the Zung Self-Rating Depression Scale and were categorized as follows: not depressed (≤ 39), mildly depressed (40-47), moderately depressed (48-55), and severely depressed (≥ 56).

[j] Student t test.

[k] Mood states were measured by the numerical scale, which ranged from 0 (good mood) to 10 (bad mood).

Qualitative Content Analysis

Potential benefits and harms were examined by qualitative content analysis for 30 valid answers to the open-ended question; 19 concepts and 7 categories were developed (Table 3). A developed concept was described by < >, and a developed category was described by << >>.

Through the analysis, we generated the following storyline that described the reasons and process of SNS usage. The positive comments revealed that some channels (eg, message, blog, and community) or some functions (eg, invitation function, footprints function, and privacy controls function) ensured <<Advantage conditions>> like <Anonymity>, <Easiness>, and <Expectation>, creating a domain where participants could face each other honestly and obtain <<peer support>>. This indicated that the SNS helped network members (1) <Recognize the existence of peers>, who were others suffering from a similar disease; (2) <Acquire information> about their disease, treatment, and experience; (3) <Narrate their experience>; (4) <Support each other> through online interaction; and (5) <Encourage peer support> more quantitatively and/or qualitatively (they made more friends and/or they got closer friends). As an <<Advantage consequence>>, peer support enabled them to understand themselves and <Feel positive>. For some participants, peer support may even have led to <Changing behavior>, such as changing treatment as a consequence of the support.

On the other hand, few participants listed <Egocentric comments> and <Infrequent usage> as <<Disadvantage conditions>> of participation. <Solely cyber communication> with the SNS and intensified <Dependency> by depressed people were identified as negative aspects that encumbered some members with additional psychological burdens. Such increases in psychological burden could subsequently trigger a <Downward depressive spiral>, with the SNS exacerbating certain members' symptoms, like <Reading negative comments>, <Being depressed>, and <Writing negative comments>. As a <<Disadvantage consequence>>, they experienced <Disappointment> in SNS participation.

Table 3. Concepts and categories developed by qualitative content analysis (N = 30)

	Assessment of the SNS (N = 30)[a]		
	Positive (N = 19)	Not Positive (N = 10)	No Assessment (N = 1)
Advantage Aspect [b]			
<<Advantage conditions>>			
<Anonymousness>	54	30, 114	
<Easiness>	95	85	
<Expectation>	69	85	
<<Peer support>>			
<Recognizing the existence of peers>	9, 24, 39, 76, 84, 95, 96	19	93
<Acquiring information>	39, 76, 95	64	
<Narrating their experiences>	24, 57, 76		
<Supporting each other>	24, 49, 58, 70, 84, 92, 95, 113	19, 55, 64, 87, 114	
<Encouraging peer support>	3, 21, 105	64, 106	
<<Advantage consequence>>			
<Feeling positive>	9, 21, 39, 58, 76, 92	59	93
<Changing behavior>	49, 76, 113		
Disadvantage Aspect			
<<Disadvantage conditions>>			
<Egocentric comments>	3		
<Infrequent usage>	69, 96	106, 114	
<<Additional psychological burdens>>			
<Solely cyber communication>	57, 76	64	
<Dependency>	21, 24	59, 109	
<<Downward depressive spiral>>[c]			
<Downward depressive spiral >[c]		87	
<Reading negative comments>	58	30	
<Being depressive>	39, 70	109	
<Writing negative comments>	58		
<<Disadvantage consequence>>			
<Disappointment>		30	

[a] Numbers stand for anonymous registrants' IDs, corresponding to Figure 3.

[b] < > denotes a concept; << >> denotes a category.

[c] After <Downward depressive spiral> was developed as a concept, a category <<Downward depressive spiral>> was developed that included four concepts: <Downward depressive spiral>, <Reading negative comments>, <Being depressive>, and <Writing negative comments>.

Social Network Analysis

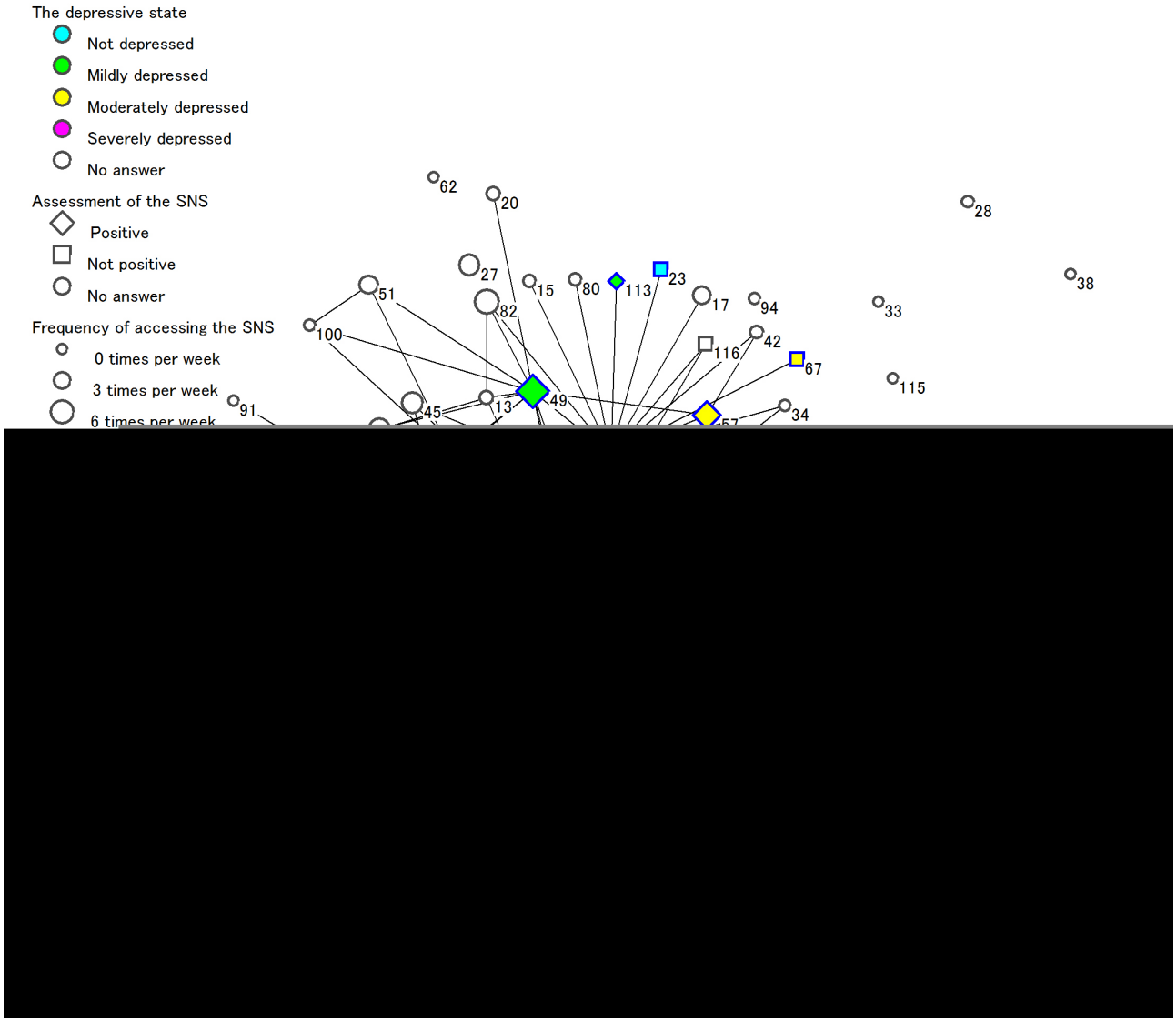
Figure 3 depicts the social network of the 105 participants.

The number of cliques was 115, and the two biggest cliques each included five people, (IDs 9, 61, 76, 99, 112 and IDs 58, 86, 99, 110, 111), which meant they were friends with each other. The top five registrants (in order) as measured by degree centrality were 99, 114, 9, 70, and 36; by closeness centrality, were 99, 114, 9, 65, and 21; and by betweenness centrality, were 99, 114, 70, 9, and 49. People using the SNS frequently, having more friends, and with a "positive" assessment seem to be clustered around the central region of Figure 3. Some gatekeepers, such as 99, 114, and 70, who had high betweenness centrality, are also connected to individuals in the periphery of the figure. ID 114 had some friends who were moderately or severely depressed, such as 57, 58, 60, and 87. ID 114 was unsatisfied with <Infrequent usage>, although pointing out <Anonymity> and <Supporting each other> as advantages of

the SNS. ID 87, who is a friend of 114, pointed out the potential harm, using the words “downward depressive spiral.”

In terms of the extent of interpersonal association, the odds ratio (95% confidence interval) for assessment of the SNS, depressive state, and frequency of accessing the SNS was 1.0 (0.5-2.2), 0.9 (0.3-2.6), and 0.9 (0.5-1.5), respectively.

Figure 3. Social network in the SNS. Each node with a number represents one person. Numbers refer to anonymous registrants’ IDs, corresponding to Table 3. Each line between nodes indicates a “friend” relationship. Depressive states, assessment of the SNS, and mood states are explained in Table 2. We compared the mood states in two situations: (A) during normal time (time not using the SNS or Internet) and (C) while using the SNS. “Better than during normal time” means that the mood state in (C) is better than the mood state in (A).



Discussion

Main Findings

Most participants in this study were depressed, since 32/35 people (91%) could be diagnosed with mood disorders and 28/31 people (90%) had depressive tendencies, being either mildly, moderately, or severely depressed as measured by the depressive scale. The median age of participants was in the late 30s, older than users of MySpace or mixi, who tend to be adolescents. Almost half of valid respondents (16/34) were not working, which might suggest that people over 30 had trouble at work due to a depressive state. The data showed that the SNS was used by most participants since 34/37 (92%) had written a profile and 24/36 (67%) accessed the SNS three times or more per week. However, only a few people updated their blogs and/or registered in a large number of communities (see Table 1).

Qualitative content analysis indicated that the SNS could provide various types of peer support that users could select based on anonymity and ease of use. User-selectable peer support was passive, active, and/or interactive. On the other hand, there was the potential harm of a downward depressive spiral triggered by aggravated psychological burden. Social network analysis showed that users communicated one-on-one with each other or in small groups since the median of friends was three and the largest group included only five people. It also implied that a downward depressive spiral was related to friends who were

moderately or severely depressed and friends with negative assessment of the SNS.

Potential Benefits of SNS Participation: Peer Support

The SNS could provide user-selectable peer support. One participant felt encouraged, recognizing that a peer who was more depressed than he/she was continued to persevere (ID 9's comment). At times, participants also felt empowered by giving support to others. As peer support was not a one-way interaction from the less depressed individual to the more depressed individual, it could be passive, active, and/or interactive. Accordingly, peer support might spread incrementally in the SNS. Moreover, participants could decide for themselves how many friends they connected with and how close they became to these friends. Therefore, a peer support SNS could meet participants' needs, such as avoiding face-to-face communication and maintaining privacy, while providing peer support. Given that SNSs are widely used among the public, this suggests that SNSs offer effective opportunities for people with depressive tendencies to obtain peer support.

Potential Harm of SNS Participation: Downward Depressive Spiral

As mentioned in the qualitative content analysis, a detrimental effect can result given that depressive tendencies can be exacerbated by cyber communication and dependency can be intensified for people with higher depressive tendencies. Dennis [6] also points out that an adverse outcome from peer support is the potential for emotional over-involvement that results in contagion stress. The downward depressive spiral emerged in part of the SNS, possibly because people with depressive tendencies tend to have greater dependency or because the mood state while using the SNS can worsen compared to normal time or while using the Internet if people had a "not positive" assessment of the SNS. In contrast, social network analysis showed that people with similar attributes, such as assessment of the SNS, depressive state, and frequency of accessing the SNS, did not tend to be connected with each other as a whole.

There is a potential discrepancy between the results of the qualitative content analysis and social network analysis. We interpret this discrepancy as follows. The SNS could, as a whole, prevent participants from experiencing the downward depressive spiral because it provides a domain where participants can face each other honestly. However, content analysis presented the possibility that the downward depressive spiral is an adverse event resulting from SNS participation. Future studies will be required to fully resolve this discrepancy.

Limitations and Strengths

Our study has several limitations and strengths. We acknowledge that group members were not selected through theoretical sampling and were by no means a representative sample of patients with clinical depression. However, they were people with complaints of depression who used the SNS. A study reported that the suicide rate from 2003 to 2004 in the United States increased, and the influence of Internet social networks was included as a potential factor to consider [38]. The need to examine relationships between Internet social networks and depression or suicide must be addressed.

We also acknowledge that the sample size was very small (105 total group members). A limitation of this study is determining if the results are representative because the study included just one SNS, which had a small number of members. Moreover, the survey was done through the Internet, and the ratio of respondents was relatively low (34%). To address selection bias, we compared valid respondents with invalid/nonrespondents using SNS log files, which showed that people using the SNS frequently were selected. To address information bias, we confirmed the answers' reliability and that the questionnaire responses corresponded with the SNS log files.

Finally, this study was designed as a cross-sectional survey, so we could not estimate causality between the SNS's effects and outcomes. As new information technology on the Internet is being developed, recently referred to as Web 3.0 [39], new technologies and services emerge and are put to use even before their benefits or harms are assessed. It might be not practical to examine the SNS by an experimental study because it will not be used until a few years later. However, it is useful to research existing technologies using observational studies since most new technologies stand on the shoulders of existing technologies.

Implications

We believe that this study has three public health implications. First, this study evaluated potential benefits and harm of SNSs, which are widely used among the public. Since SNSs can address individual needs of the public, it is important to analyze consumers' needs from the viewpoint of consumer health informatics [40,41]. Finding a balance between potential benefits and harms [15] contributes to both health and eHealth providers as well as patients and SNS users. Moreover, these benefits and harms could spread as "network phenomena" [18-20]. From the perspective of public health, it is important to prevent initial undesirable events. Desirable intervention would also provide desirable results.

Second, this study used mixed methods, combining content analysis of qualitative data with social network analysis of quantitative data. These data were derived from questionnaires and SNS log files. A qualitative approach is required to analyze relationships among people, and therefore this mixed methods strategy could be useful for exploring social networks.

Finally, this study suggested that the SNS was a kind of social network, as depicted in Figure 3. Features of Web 2.0 in health care, called Medicine 2.0, are social networking, collaboration, participation, apomediation, and openness [42]. Therefore, in the future, personally controlled online health data using Google Health or Microsoft HealthVault [43] could be linked by a common application programming interface for social applications across multiple websites (eg, OpenSocial) [44]. In future studies we believe it will be necessary to consider the influence of interpersonal associations or social networks.

Conclusions

A peer support SNS for people with depression might offer effective opportunities to obtain support for people with depressive tendencies because the SNS can provide

user-selectable passive, active, and/or interactive peer support based on anonymity or ease of use. It can meet participants' needs, such as avoiding face-to-face communication and maintaining privacy, while providing peer support. On the other hand, to avoid a downward depressive spiral, we recommend that participants refrain from using the SNS when they feel that the SNS is not user-selectable (eg, when they get egocentric comments, when friends have a negative assessment of the SNS, or when they have an additional psychological burden).

As communication on the Internet becomes more social, the mixed methods strategy used here, combining content analysis of qualitative data and social network analysis of quantitative data, is available to explore benefits and harms of this communication.

Acknowledgments

We are grateful to Takuya Okada, who voluntarily built and administrated the SNS, and all participants. We thank Professor Minoru Yabuuchi in the Graduate School of Psychology at Kyoto Notre Dame University for his valuable suggestions, Michal Nissim for her assistance with manuscript preparation, and Chiaki Takahashi and Shizuka Watanabe for analyzing quantitative data.

Y. Takahashi received a Pfizer Health Research Foundation Grant for Internal Joint Research by Young Researchers in 2005. The sponsor had no role in study design; in collection, analysis, or interpretation of data; in writing the paper; or in the decision to submit the paper for publication.

For author's contributions, Y Takahashi conceived the study, designed the protocol, participated in data collection, was responsible for the data analysis, and drafted the manuscript. C Uchida participated in the study design and gave a great deal of valuable advice regarding psychiatry. K Miyaki, M Sakai, and T Shimbo gave advice regarding data analysis. M Sakai was one of the leading members analyzing qualitative data. T Nakayama filled the role of supervisor and gave valuable advice regarding all aspects of the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Glossary of terms used in this article

[PDF file (Adobe PDF), 82 KB - jmir_v11i3e29_app1.pdf]

Multimedia Appendix 2

Framework of analytic methodology

[PDF file (Adobe PDF), 27 KB - jmir_v11i3e29_app2.pdf]

References

1. Powell J, McCarthy N, Eysenbach G. Cross-sectional survey of users of Internet depression communities. BMC Psychiatry 2003 Dec 10;3(1):19-26 [FREE Full text] [Medline: 14664725] [doi: 10.1186/1471-244X-3-19]
2. Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. Lancet 2007 Sep 8;370(9590):851-858. [Medline: 17826170] [doi: 10.1016/S0140-6736(07)61415-9]
3. ; American Psychiatric Association Work Group on Major Depressive Disorder. Practice Guideline for the Treatment of Patients With Major Depressive Disorder, Second Edition. 2000. URL: http://www.psychiatryonline.com/pracGuide/pracGuideTopic_7.aspx [WebCite Cache ID 5b5x3vKRD]
4. ; U.S. Preventive Services Task Force. Screening for depression: recommendations and rationale. Ann Intern Med 2002 May 21;136(10):760-764 [FREE Full text] [Medline: 12020145]
5. Sharp LK, Lipsky MS. Screening for depression across the lifespan: a review of measures for use in primary care settings. Am Fam Physician 2002 Sep 15;66(6):1001-1008 [FREE Full text] [Medline: 12358212]
6. Dennis CL. Peer support within a health care context: a concept analysis. Int J Nurs Stud 2003 Mar;40(3):321-332. [Medline: 12605954] [doi: 10.1016/S0020-7489(02)00092-5]
7. Doull M, O'Connor AM, Robinson V, Tugwell P, Wells GA. [Protocol] Peer support strategies for improving the health and well-being of individuals with chronic diseases. Cochrane Database of Systematic Reviews 2005;2007(4) CD005352.
8. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. BMJ 2004 May 15;328(7449):1166-1171 [FREE Full text] [Medline: 15142921] [doi: 10.1136/bmj.328.7449.1166]

9. Houston TK, Cooper LA, Ford DE. Internet support groups for depression: a 1-year prospective cohort study. Am J Psychiatry 2002 Dec;159(12):2062-2068 [FREE Full text] [Medline: 12450957] [doi: 10.1176/appi.ajp.159.12.2062]
10. Andersson G, Bergström J, Holländare F, Carlbring P, Kaldo V, Ekselius L. Internet-based self-help for depression: randomised controlled trial. Br J Psychiatry 2005 Nov;187(5):456-461 [FREE Full text] [Medline: 16260822] [doi: 10.1192/bjp.187.5.456]
11. O'Reilly T. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. 2005. URL: http://www.oreilly.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html [WebCite Cache ID 5b5woXPA8]
12. Moreno MA, Fost NC, Christakis DA. Research ethics in the MySpace era. Pediatrics 2008 Jan;121(1):157-161 [FREE Full text] [Medline: 18166570] [doi: 10.1542/peds.2007-3015]
13. Hinduja S, Patchin JW. Personal information of adolescents on the Internet: A quantitative content analysis of MySpace. J Adolesc 2008 Feb;31(1):125-146. [Medline: 17604833] [doi: 10.1016/j.adolescence.2007.05.004]
14. Ybarra ML, Mitchell KJ. How risky are social networking sites? A comparison of places online where youth sexual solicitation and harassment occurs. Pediatrics 2008 Feb;121(2):e350-e357 [FREE Full text] [Medline: 18227194] [doi: 10.1542/peds.2007-0693]
15. Mitchell KJ, Ybarra M. Social networking sites: finding a balance between their risks and benefits. Arch Pediatr Adolesc Med 2009 Jan;163(1):87-89. [Medline: 19124709] [doi: 10.1001/archpediatrics.2008.534]
16. Block JJ. Issues for DSM-V: internet addiction. Am J Psychiatry 2008 Mar;165(3):306-307 [FREE Full text] [Medline: 18316427] [doi: 10.1176/appi.ajp.2007.07101556]
17. Baker JR, Moore SM. Distress, coping, and blogging: comparing new myspace users by their intention to blog. Cyberpsychol Behav 2008 Feb;11(1):81-85. [Medline: 18275317] [doi: 10.1089/cpb.2007.9930]
18. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007 Jul 26;357(4):370-379 [FREE Full text] [Medline: 17652652] [doi: 10.1056/NEJMsa066082]
19. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med 2008 May 22;358(21):2249-2258 [FREE Full text] [Medline: 18499567] [doi: 10.1056/NEJMsa0706154]
20. Fowler JH, Christakis NA. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. BMJ 2008 Dec 04;337:a2338 [FREE Full text] [Medline: 19056788] [doi: 10.1136/bmj.a2338]
21. Pope P, Mays N. Qualitative Research in Health Care. 3rd Edition. Oxford, UK: Blackwell Publishing; 2006.
22. Lingard L, Albert M, Levinson W. Grounded theory, mixed methods, and action research. BMJ 2008 Aug 07;337:a567. [Medline: 18687728]
23. Tashakkori A, Teddlie C. Handbook of Mixed Methods in Social and Behavioral Research. Thousand Oaks, CA: SAGE Publications; 2003.
24. O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. J Health Serv Res Policy 2008 Apr;13(2):92-98. [Medline: 18416914] [doi: 10.1258/jhsrp.2007.007074]
25. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [Medline: 15471760] [doi: 10.2196/jmir.6.3.e34]
26. Zung WW. A Self-Rating Depression Scale. Arch Gen Psychiatry 1965;12:63-70. [Medline: 14221692]
27. Fukuda K, Kobayashi S. Japanese Version of Self-Rating Depression Scale. Kyoto, Japan: Sankyobo; 1983.
28. Williamson A, Hoggart B. Pain: a review of three commonly used pain rating scales. J Clin Nurs 2005 Aug;14(7):798-804. [Medline: 16000093] [doi: 10.1111/j.1365-2702.2005.01121.x]
29. 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. [Medline: 14769454] [doi: 10.1016/j.nedt.2003.10.001]
30. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [Medline: 18352969] [doi: 10.1111/j.1365-2648.2007.04569.x]
31. Scott J. Social Network Analysis: A Handbook. Thousand Oaks, CA: SAGE Publications; 2000.
32. Hawe P, Webster C, Shiell A. A glossary of terms for navigating the field of social network analysis. J Epidemiol Community Health 2004 Dec;58(12):971-975 [FREE Full text] [Medline: 15547054] [doi: 10.1136/jech.2003.014530]
33. Luke DA, Harris JK. Network analysis in public health: history, methods, and applications. Annu Rev Public Health 2007;28(1):69-93. [Medline: 17222078] [doi: 10.1146/annurev.publhealth.28.021406.144132]
34. Borgatti SP, Mehra A, Brass DJ, Labianca G. Network analysis in the social sciences. Science 2009 Feb 13;323(5916):892-895. [Medline: 19213908] [doi: 10.1126/science.1165821]
35. Steve S. An Overview of Content Analysis. Practical Assessment, Research & Evaluation 2001;7:17.
36. Glaser B, Strauss A. The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago, IL: Aldine; 1967.
37. Takahashi Y. A survey of a peer support social network service on the Internet for people with depressive tendencies [in Japanese]. 2007. URL: http://www.snsstudy.umin.jp/ [WebCite Cache ID 5bNlQ0Y37]
38. Bridge JA, Greenhouse JB, Weldon AH, Campo JV, Kelleher KJ. Suicide trends among youths aged 10 to 19 years in the United States, 1996-2005. JAMA 2008 Sep 3;300(9):1025-1026. [Medline: 18768413] [doi: 10.1001/jama.300.9.1025]
39. Giustini D. Web 3.0 and medicine. BMJ 2007 Dec 22;335(7633):1273-1274. [Medline: 18156223] [doi: 10.1136/bmj.39428.494236.BE]

40. Eysenbach G. Consumer health informatics. BMJ 2000 Jun 24;320(7251):1713-1716 [FREE Full text] [Medline: 10864552] [doi: 10.1136/bmj.320.7251.1713]
41. LaVenture M. Using the power of googling and health informatics to improve public health practice. Am J Prev Med 2007 Jul;33(1):75-76. [Medline: 17572315] [doi: 10.1016/j.amepre.2007.04.008]
42. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [Medline: 18725354] [doi: 10.2196/jmir.1030]
43. Steinbrook R. Personally controlled online health data--the next big thing in medical care? N Engl J Med 2008 Apr 17;358(16):1653-1656. [Medline: 18420496] [doi: 10.1056/NEJMp0801736]
44. Google. OpenSocial. 2007. URL: http://code.google.com/apis/opensocial [WebCite Cache ID 5b5ww3Wgq]

Abbreviations

SNS: social network service

Edited by L Van Gemert-Pijnen; submitted 07.10.08; peer-reviewed by M Moreno, N Nijland; revised version received 14.03.09; accepted 15.03.09; published 23.07.09.

Please cite as:
Takahashi Y, Uchida C, Miyaki K, Sakai M, Shimbo T, Nakayama T
Potential Benefits and Harms of a Peer Support Social Network Service on the Internet for People With Depressive Tendencies: Qualitative Content Analysis and Social Network Analysis
J Med Internet Res 2009;11(3):e29
URL: http://www.jmir.org/2009/3/e29/
doi:10.2196/jmir.1142
PMID:19632979

Original Paper

A Brief Web-Based Screening Questionnaire for Common Mental Disorders: Development and Validation

Tara Donker[1], MSc; Annemieke van Straten[1], PhD; Isaac Marks[1,2], MD; Pim Cuijpers[1], PhD

[1]Department of Clinical Psychology, VU University Amsterdam, Amsterdam, The Netherlands

[2]Institute of Psychiatry, Kings College London, London, UK

Corresponding Author:
Tara Donker, MSc
VU University Amsterdam
FPP, Department of Clinical Psychology
Van der Boechorststraat 1
1081 BT Amsterdam
The Netherlands
Phone: +31 20 5988265
Fax: +31 20 5988758
Email: t.donker@psy.vu.nl

Abstract

Background: The advent of Internet-based self-help systems for common mental disorders has generated a need for quick ways to triage would-be users to systems appropriate for their disorders. This need can be met by using brief online screening questionnaires, which can also be quickly used to screen patients prior to consultation with a GP.

Objective: To test and enhance the validity of the Web Screening Questionnaire (WSQ) to screen for: depressive disorder, alcohol abuse/dependence, GAD, PTSD, social phobia, panic disorder, agoraphobia, specific phobia, and OCD.

Methods: A total of 502 subjects (aged 18 - 80) answered the WSQ and 9 other questionnaires on the Internet. Of these 502, 157 were assessed for DSM-IV-disorders by phone in a WHO Composite International Diagnostic Interview with a CIDI-trained interviewer.

Results: Positive WSQ “diagnosis” had significantly ($P < .001$) higher means on the corresponding validating questionnaire than negative WSQ “diagnosis”. WSQ sensitivity was 0.72 - 1.00 and specificity was 0.44 - 0.77 after replacing three items (GAD, OCD, and panic) and adding one question for specific phobia. The Areas Under the Curve (AUCs) of the WSQ’s items with scaled responses were comparable to AUCs of longer questionnaires.

Conclusions: The WSQ screens appropriately for common mental disorders. While the WSQ screens out negatives well, it also yields a high number of false positives.

(J Med Internet Res 2009;11(3):e19) doi:10.2196/jmir.1134

KEYWORDS

Depression; anxiety; mental disorders, common; needs assessment; Internet; online systems

Introduction

The thriving development of Internet-based self-help aids [1] for particular mental disorders [2,3] has generated a need for quick ways to triage would-be users to systems appropriate for their disorders. Many sufferers do not easily recognize their particular mental problem [4] and could be guided by a Web-screening questionnaire to a self-help system appropriate for their problem. This could reduce the likelihood of their becoming disenchanted with using a self-help system not intended for their disorder. Such a questionnaire would preferably be conducted via the Internet, as it offers quick and easy access to large numbers of users at a low cost [5,6]. This kind of questionnaire could also assist professionals such as general practitioners (GPs) in screening their patients prior to consultation.

The screening must be brief, as subjects will undergo screening more readily if it is short, quick [7], and easy to read. A few brief online screening questionnaires [8-10] appear to be reliable and valid. The Internet-based Self-assessment Program for Depression (ISP-D [10]), for example, reported sensitivity, specificity, positive predictive values (PPV), and negative

predictive values (NPV) for major depressive disorder of 0.82, 0.73, 0.67, and 0.86, respectively [10]. Sensitivity of another online test, the Web-Based Depression and Anxiety Test (WB-DAT [8]), ranged from 0.71 to 0.95, while specificity ranged from 0.87 to 0.97 for major depressive disorder (MDD), obsessive compulsive disorder (OCD), post-traumatic stress disorder (PTSD), panic disorder with and without agoraphobia, and social phobia. Sensitivity for generalized anxiety disorder (GAD) was somewhat lower (0.63). However, existing online screening questionnaires do not assess all mental disorders for which self-help systems are now being created. To reduce this paucity, we developed a brief online screening questionnaire which screens for different mental disorders: the Web Screening Questionnaire for common mental disorders (WSQ), based on the Screening Questionnaire (SQ) of Marks and colleagues [9]. The WSQ contains only 15 items and screens for depression, GAD, panic disorder with and without agoraphobia, social phobia, specific phobia, OCD, PTSD, and alcohol abuse/dependence. This paper reports optimization and validation of the WSQ.

Methods

Participants and Procedure

Participants were recruited (between May and December 2007) from the general Dutch population by using Internet banners (eg, Google and Dutch Internet sites on mental health issues). The advertisements linked to a Web page containing information about common mental disorders, Internet treatment and this study, an application form, and a link to the questionnaires. Subjects were asked to input their name and email address, so they could be identified and added to the data pool only once.

We specifically targeted adults (18 years of age or older) with Internet access and who felt anxious, depressed, or thought of themselves as drinking too much alcohol. We targeted a population with a high rate of common mental disorders as the kind likely to use the WSQ in the future. Since this population can only illuminate false negative and true positive rates, we needed controls to test those rates. Therefore, we also recruited 20 undergraduate psychology students who were not required to have symptoms, using banners at the VU University's students' Web page seeking participants for VU studies.

We excluded people reporting a high suicide risk (ie, a score of 3 on Q15 of the WSQ); they were advised to contact their GP. To raise the response rate, participants were told in advance that completers of the screening questionnaires would be offered a self-help book for common mental problems. Students received academic credit for participating. The study protocol was approved by the Medical Ethics Committee at the VU Medical Centre in Amsterdam, Netherlands.

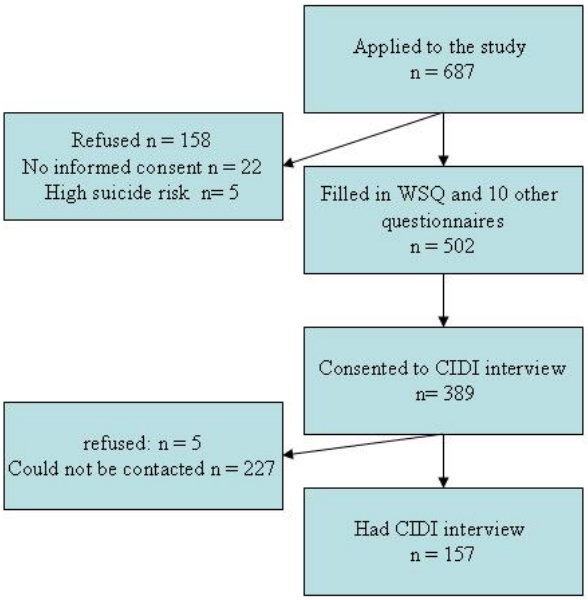
Our study tested the WSQ's validity and consisted of two parts (Figure 1):

1. Completion of 10 sets of questions: Internet demographic questions, the WSQ, and other questionnaires for common mental disorders: Center for Epidemiological Studies Depression scale (CES-D [11]), Generalized Anxiety Disorder scale (GAD-7 [12]), Fear Questionnaire (FQ [13]) plus a further question about eight kinds of specific phobia, Panic Disorder Severity Scale - Self Report (PDSS-SR [14,15]), Yale-Brown Obsessive Compulsive Scale (YBOCS [16,17]), Impact of Events Scale (IES [18]), Alcohol Use Disorders Identification Test (AUDIT [19]; details below),
2. A DSM-IV-diagnostic phone interview with a Composite International Diagnostic Interview (CIDI)-trained interviewer (CIDI lifetime, World Health Organization (WHO) version 2.1 [20]) to assess the presence of a current (ie, within the last 6 months) DSM-IV diagnosis [21] of MDD, dysthymia (Dyst), minor depression (MinD), social phobia, GAD, panic disorder, agoraphobia, specific phobia, OCD, PTSD, and alcohol abuse/dependence. CIDI-interviewers were blind to the subjects' self-reports and the inclusion of control subjects (undergraduate psychology students).

In all, 687 people applied for the study, of whom 185 (27%) were excluded because they represented a high suicide risk (n = 5); there was no written informed consent (n = 22); or they refused to participate (n = 158). This left 502 participants, of whom 389 consented to a diagnostic phone interview, but 232 (60%) of those 389 either could not be contacted (n = 227) or refused (n = 5), leaving 157 participants who were phoned by a CIDI-trained interviewer within a mean of 13 days.

If participants had never experienced a traumatic event, they skipped the IES; if they had never drunk alcohol, they skipped the AUDIT; and if they had never suffered a panic attack, they skipped the PDSS-SR. Those who completed the screening questionnaires and gave informed consent entered the study.

Figure 1. Flowchart of participants (WSQ)



Measures

Development of the Web Screening Questionnaire for Common Mental Disorders (WSQ)

The WSQ for common mental disorders [22] has 15 self-rated questions based on the screening questionnaire (SQ) of Marks and colleagues [9] which screens for most common mental disorders. Of the SQ's original questions we used 6 unchanged (WSQ Q1, 3, 5, 6, 11, and 15) and added 8 questions from further reliable and valid instruments. These are:

- WSQ Q2 for depression, from CIDI [20],
- WSQ Q4, 8, 9, 10, and 12 (for panic, social phobia, PTSD, and OCD from Mini-International Neuropsychiatric Interview (M.I.N.I. [23]), and
- WSQ Q13 and 14 (for alcohol, from AUDIT [19]).

Three questions of the original WSQ reached either low specificity or low sensitivity. To enhance validity, we used logistic regression analysis to determine whether other items from appropriate questionnaires could replace these WSQ-items. We amended three questions using items for GAD (WSQ Q3, from GAD-7 [12]) for panic (WSQ Q4 from PDSS-SR [15]), and for OCD (WSQ 12, from YBOCS [16,17]). We also added one question for the WSQ subscale specific phobia (WSQ Q7) which concerned further types of specific phobia. Each WSQ subscale has 1 - 2 items (for GAD, panic disorder, OCD, alcohol addiction, depression, agoraphobia, specific phobia, social phobia, and PTSD). Of the 15 WSQ questions, 8 had "yes" or "no" answers while the other 7 were Likert-type scales.

Further Screening Questionnaires

Depressive Symptoms

The Dutch version of the CES-D [11] has 20 self-rated items with each scored on a range of 0 - 3 and a total score of 0 - 60. The paper-pencil CES-D has good psychometric properties with a cut-off score of 16 [24]. The Internet CES-D is also reliable and valid with a cut-off score of 22 (Cronbach alpha: .93; sensitivity: 0.90; specificity: 0.74 [25]).

GAD

We translated the GAD-7 [12] into Dutch for self-rating of generalized anxiety symptoms. Each of its 7 questions is rated 0 - 3 ("not at all" to "nearly every day"), and the total score range is 0 - 21. Reliability is excellent (Cronbach alpha = .92). With a cut-off point of ≥ 10, sensitivity is 0.89 and specificity is 0.82 among primary care participants [12]. The GAD-7 was translated into Dutch by forward-translation (translated and discussed by two independent health professionals) and blind backward-translation (by an independent translator whose mother tongue is English). Since psychometric properties may differ among other populations, the Dutch version of the GAD-7 is validated in another study (TD, AVS, IMM, and PC, unpublished data, 2009).

Panic Disorder

The Dutch version of the PDSS [14] self-report SR form [15] asks 7 questions about 7 dimensions of panic disorder, each self-rated 0 - 4, with a total score range of 0 - 28. With a cut-off score of 8, sensitivity is 0.83 and specificity is 0.64 [14].

Phobias (Agoraphobia, Social Phobia, Specific Phobia)

The Dutch version of the FQ [13] detects agoraphobia, social phobia, and blood-injury phobia. The FQ's total phobia scale contains 15 items; each self-rated 0 - 8, with a total score range of 0 - 120. Several studies support the validity of the FQ's social and agoraphobia subscales [26-29]. To the FQ's 5 blood/injury questions we added a single self-rated question ("are you scared of …?") concerning further types of specific phobias, to be ticked as present or absent: animals (eg, dogs, cats), natural events (eg, earthquakes, storms, flooding), body fluids (eg, faeces, vomit, semen), materials (eg, cleaning products, medicine, poison), medical appointments (eg, dentist, hospital), items at home (eg, telephone, toilet, soap), specific situations (eg, driving, riding elevators, crossing bridges), and other (eg, vomiting, children). We omitted the FQ's 6 anxiety-depression items.

PTSD

The IES [18] assesses signs and symptoms of avoidance and intrusion after a serious or traumatic life event. It has 15 items, each self-rated 0 - 5, with a total score range of 0 - 75. People who score ≥ 26 are likely to have PTSD. The Dutch version is reliable and sensitive [30].

OCD

We used the Dutch 10-item severity subscale of the YBOCS [16,17]. Each self-rated item is rated 0 - 4, with a total score range of 0 - 40. Tests of internal consistency of the total scale (Dutch version) are .69 to .91 (Cronbach alpha) and compare well with several but not all measures often used to assess OCD [31]. A total score of 13 or more denotes clinically significant obsessive-compulsive symptoms [32].

Alcohol Abuse/Dependence

The Dutch version of WHO's self-rated AUDIT [19] identifies people with hazardous alcohol consumption and dependence in primary care. Each of its 10 items is rated 0 - 4, with a total score range of 0 - 40. Cronbach alpha is .65 to .93: overall sensitivity is 0.92, and specificity is 0.94 [19]. A cut-off score of 8 is recommended for various endpoints (eg, alcohol-related social problems or medical problems) [33].

Diagnoses

We used the Lifetime version 2.1 of the CIDI [19] in its Dutch version [34,35] as a "gold standard' to assess the presence of DSM-IV disorders in the last 6 months (GAD, panic disorder, OCD, alcohol abuse/dependence, MDD, Dyst, MinD, agoraphobia, specific phobia, social phobia, and PTSD). The CIDI is reliable and valid [36,37]. The CIDI was administered by phone by trained CIDI interviewers who were psychologists or master's-level psychology students. The CIDI interviews used in this trial lasted 69 minutes on average.

Analyses

To establish whether WSQ scores differed significantly between subjects with positive and with negative screen results, we conducted t-tests on the mean and standard deviation of each screening instrument separately. In the sub-sample that had a diagnostic interview, we performed chi-square tests to ascertain whether WSQ scores differed between subjects with and without DSM-IV disorders.

We calculated sensitivity and specificity, and positive and negative predictive values, for each WSQ subscale regarding its corresponding DSM-IV disorder (predictive validity). Sensitivity is the probability that a person who has a disorder is screen positive. Specificity is the probability that a person not suffering from a disorder is screen negative. There is no consensus of what levels of sensitivity and specificity are acceptable, as they depend on the test's aim, costs, and benefits [38]. The WSQ aims to detect clinically-relevant mood, anxiety, and alcohol-related problems. Therefore, to minimize missed cases we set threshold levels of sensitivity at 0.70 or more, and of specificity at 0.40 or more. PPV is the probability of a positive diagnosis after a positive screening, and NPV is the probability of a negative diagnosis after a negative screening. PPV and NPV depend on prevalence (PPV increases when prevalence increases), so we did not set acceptable levels of these.

For WSQ questions which turned out to have unacceptable sensitivity or specificity, we replaced them with relevant items from the appropriate screening questionnaire. To find which items best predicted the chance of detecting a diagnosis, we used logistic regression analyses (Forward Likelihood Ratio method). We replaced items only if they improved validity. We calculated the Area Under the Curve (AUC) for the WSQ's scaled and dichotomous response options and its appropriate screening questionnaires. The AUC (the sum of sensitivity versus [1 –] specificity) measures a scale's accuracy; it equals the probability that a randomly chosen case will score higher than a randomly chosen non-case [39]. AUCs of 0.5 - 0.7 are said to reflect low accuracy, 0.7 - 0.9 moderate accuracy, and 0.9 - 1.0 high accuracy [40]. Furthermore, we performed t-tests and χ^2 tests to examine differences in demographic and questionnaire results between subjects who had a CIDI diagnostic interview and those who did not, and tests to examine whether a student sub-sample's WSQ scores differed from the whole sample.

Our analyses used diagnoses reached within the last 6 months. MDD, Dyst, and MinD were combined into the category depressive disorder. For all analyses we used SPSS version 15.0 for Windows.

Results

Sample Characteristics

The total sample (N = 502) had a mean age of 43 years (SD 13, range 18 - 80), and 285 (57%) of the subjects were female. Of the 157 subjects who had a CIDI interview, the mean age was 43 (SD 15, range 18 - 80). Of these, 89 (57%) were female, and 107 (68%) subjects met DSM-IV criteria for any current (ie,

within the past 6 months) depressive disorder, anxiety disorder, and/or alcohol abuse/dependence. A total of 67 (43%) subjects had more than one diagnosis (Table 1).

Table 1. Demographic characteristics and prevalence of diagnosis

	N (%)	
	Complete sample	CIDI sub-sample
Completed all questionnaires on Internet	502 (100)	157
Gender, N (%)		
Male	217 (43)	68 (43)
Female	285 (57)	89 (57)
Age, Mean (SD)	43 (13)	43 (15)
(Range)	(18 - 80)	(18 - 80)
Education		
Low[a]	99 (20)	27 (17)
Medium[b]	217 (43)	73 (47)
High[c]	186 (37)	57 (36)
Country		
Netherlands	474 (94)	146 (94)
Other	28 (6)	11 (6)
Marital status		
Single	180 (36)	65 (41)
Married or cohabiting	241 (48)	67 (43)
Divorced/widowed	81 (16)	25 (16)
DSM-IV diagnosis within last 6 months, on CIDI phone interview	157	
Any depressive disorder	52 (33)	
Major depressive disorder	46 (29)	
Dysthymia	9 (6)	
Minor depression	8 (5)	
Any anxiety disorder	94 (60)	
Social phobia	32 (20)	
GAD	30 (19)	
Panic disorder	10 (6)	
Panic with agoraphobia	22 (14)	
Agoraphobia	10 (6)	
Specific phobia	40 (26)	
Obsessive-compulsive disorder	10 (6)	
PTSD	12 (8)	
Alcohol abuse/dependence	23 (15)	
Any disorder	107 (68)	
> one diagnosis	67 (43)	

[a]Low education: primary and lower general secondary education.

[b]Medium education: Intermediate Vocational Training,school of higher general secondary education or pre-university education.

[c]High education: higher vocational education or university.

Comparisons of WSQ With Other Questionnaires

Table 2 shows that subjects who scored "Yes" for any particular WSQ "diagnosis" had significantly higher means ($P < .001$) on the corresponding validating questionnaire than those who scored "No" for that WSQ "diagnosis".

Table 2. WSQ and screening questionnaires: means, standard deviations (SDs) and prevalence (N = 502)

Other screening questionnaires:	"Diagnosis" on WSQ (Web Screening Questionnaire)				
	Yes		No		t (d.f. = 500)
	N(%)	M (SD)	N (%)	M (SD)	
Any depressive disorder					
CES-D (score range 0 - 60)	296 (59.0)	32.2 (7.1)	206 (41.0)	18.1 (10.3)	15.2[a]
Generalized anxiety disorder					
GAD-7 (score range 0 - 21)	320 (63.8)	13.6 (3.9)	182 (36.3)	5.5 (3.1)	24.3 [a]
Panic disorder (without agoraphobia)					
PDSS-SR (score range 0 - 28)	278 (55.4)	9.3 (5.1)	224 (44.6)	0.6 (1.7)	24.2 [a]
Panic with agoraphobia					
PDSS-SR	153 (30.5)	11.2 (5.1)	349 (69.5)	2.9 (4.1)	19.3 [a]
Agoraphobia (without panic disorder)					
FQ-agoraphobia (score range 0 - 40)	205 (40.8)	12.7 (10.9)	297 (59.2)	2.9 (4.5)	14.0 [a]
Social phobia					
FQ-social phobia (score range 0 - 40)	226 (45.0)	16.6 (8.7)	276 (55.0)	7.0 (6.0)	14.6[a]
Specific phobia					
FQ-specific phobia[b] (score range 0 - 40)	290 (57.8)	7.6 (7.7)	212 (42.2)	2.3 (3.5)	9.2 [a]
Obsessive-compulsive disorder					
YBOCS (score range 0 - 40)	182 (36.3)	11.0 (6.3)	320 (63.8)	0.8 (2.3)	26.2 [a]
Post-traumatic stress disorder					
IES (score range 0 - 75)	273 (54.4)	33.5 (20.1)	229 (45.6)	0.0[c]	25.3 [a]
Alcohol abuse/dependence					
AUDIT (score range 0 - 40)	198 (39.4)	19.6 (6.2)	260 (60.6)	6.3 (5.5)	24.4 [a]

[a]Significant at $P < .001$.

[b]Additional specific phobia questions were dichotomous, so their means and standard deviations could not be calculated.

[c]If participants had never experienced a traumatic event then they skipped the IES.

Predictive Validity and Refinement of the WSQ

For the three WSQ subscales, GAD, OCD, and panic, validity was below threshold levels of 0.70 for sensitivity and 0.40 for specificity, so we replaced those (based on logistic regression analysis) with relevant items from the appropriate screening questionnaires (GAD-7, YBOCS, and PDSS-SR, respectively). This improved sensitivity or specificity. The WSQ subscale-specific phobia had an unacceptably low sensitivity (0.60), but we did not replace it with an item from the appropriate screening questionnaire as that did not improve sensitivity or specificity.

Based on the log-likelihood ratio statistic, using logistic regression analyses, we added three categories of the specific phobia question, "Are you scared of …?". These categories were (1) animals, (2) specific situations, and (3) medical issues, which improved the sensitivity of the WSQ subscale for specific

phobia but not for specificity (sensitivity: from 0.60 to 0.80; specificity: from 0.77 to 0.47).

Table 3 shows that for all 10 CIDI DSM-IV diagnoses more subjects with a CIDI diagnosis scored positive on the corresponding WSQ questions than did subjects without that CIDI diagnosis. The differences were all significant at the $P <$ .001 level except for specific phobia ($P = .003$). Table 3 also shows that the WSQ's sensitivity ranged from 0.72 (social phobia) to 1.00 (agoraphobia). The WSQ's specificity ranged from 0.44 (panic disorder) to 0.77 (panic disorder with agoraphobia). PPV varied from 0.11 (PTSD) to 0.51 (any depressive disorder), and NPV varied from 0.87 (specific phobia) to 1.00 (agoraphobia).

Table 3. WSQ vs CIDI diagnoses: prevalence, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) (n = 157)

WSQ "Diagnosis"	CIDI DSM-IV Diagnosis							
		No	Yes	χ^2 (d.f. = 1)	Sensitivity	Specificity	PPV	NPV
		n	n					
Any depressive disorder								
WSQ-depression	No	62	8	26.8[a]	0.85	0.59	0.51	0.89
	Yes	43	44					
Generalized anxiety disorder								
WSQ-GAD	No	57	2	[b]	0.93	0.45	0.29	0.97
	Yes	70	28					
Social phobia								
WSQ-social phobia	No	91	9	22.0[a]	0.72	0.73	0.40	0.91
	Yes	34	23					
Panic disorder								
WSQ-panic	No	65	1	[b]	0.90	0.44	0.10	0.98
	Yes	82	9					
Panic with agoraphobia								
WSQ-panic+agoraphobia	No	104	3	[b]	0.86	0.77	0.38	0.97
	Yes	31	19					
Agoraphobia								
WSQ-agoraphobia	No	92	0	[b]	1.00	0.63	0.15	1.00
	Yes	55	10					
Specific phobia								
WSQ-specific phobia	No	55	8	9.1[c]	0.80	0.47	0.34	0.87
	Yes	62	32					
Obsessive-compulsive disorder								
WSQ-OCD	No	102	2	[b]	0.80	0.69	0.15	0.98
	Yes	45	8					
Post-traumatic stress disorder								
WSQ-PTSD	No	68	2	[b]	0.83	0.47	0.11	0.99
	Yes	77	10					
Alcohol abuse/dependence								
WSQ-alcohol	No	97	4	[b]	0.83	0.72	0.34	0.96
	Yes	37	19					

[a]Significant at $P < .001$.

[b]Not able to calculate χ^2 due to small numbers (< 5) in cells.

[c]Significant at $P = .003$.

Compared to the corresponding CIDI DSM-IV diagnoses, the AUC for the WSQ subscales with scaled responses (WSQ subscales GAD, OCD, alcohol, and panic) were similar to the AUC of the longer questionnaires, ranging from an AUC of

0.76 for the WSQ subscale panic versus an AUC of 0.70 of the PDSS-SR, to an AUC of 0.81 for the WSQ subscale OCD versus an AUC of 0.85 for the YBOCS. The AUC for the dichotomous WSQ's subscales of panic with agoraphobia and agoraphobia were similar to the AUC of the longer, scaled questionnaires (PDSS: AUC of 0.79 versus WSQ panic with agoraphobia: AUC of 0.82; both WSQ and FQ subscale agoraphobia: AUC of 0.81), but not for the WSQ dichotomous subscales of depression, social phobia, and PTSD (ranging from WSQ subscale depression: AUC of 0.72 versus CES-D: AUC of 0.84 to WSQ subscale PTSD: AUC of 0.65 versus IES: AUC of 0.82) (Table 4).

Table 4. WSQ and screening questionnaires versus CIDI diagnoses: Area Under the Curve (AUC) and 95% CI for scaled and dichotomous response options

WSQ "Diagnosis"	CIDI DSM-IV Diagnosis	
	AUC	95% C.I.
Any depressive disorder		
WSQ-depression	0.72	0.64 - 0.80
CES-D	0.84	0.77 - 0.90
Generalized anxiety disorder		
WSQ-GAD	0.78	0.69 - 0.86
GAD-7	0.77	0.68 - 0.85
Social phobia		
WSQ-social phobia	0.72	0.62 - 0.82
FQ-social phobia	0.82	0.74 - 0.89
Panic disorder		
WSQ-panic	0.76	0.59 - 0.93
PDSS-SR	0.70	0.57 - 0.88
Panic with agoraphobia		
WSQ-panic+agoraphobia	0.82	0.72 - 0.91
PDSS-SR	0.79	0.69 - 0.89
Agoraphobia		
WSQ-agoraphobia	0.81	0.73 - 0.90
FQ-agoraphobia	0.81	0.70 - 0.91
Obsessive-compulsive disorder		
WSQ-OCD	0.81	0.65 - 0.97
YBOCS	0.86	0.72 - 0.99
Post-traumatic stress disorder		
WSQ-PTSD	0.65	0.51 - 0.80
IES	0.82	0.67 - 0.97
Alcohol abuse/dependence		
WSQ-alcohol	0.77	0.68 - 0.86
AUDIT	0.75	0.66 - 0.84

Differences Between Students and Non-students

As expected, students compared to non-students had significantly lower scores on the WSQ subscales for depression (P = .004), alcohol (P < .001), GAD (P < .001), OCD (P < .001), panic (P < .001), and panic with agoraphobia (P = .004).

Differences Between CIDI Interviewed and Non-interviewed Sub-samples

Demographic variables did not differ significantly between subjects who had a CIDI diagnostic interview and those who did not. However, those who had a CIDI interview scored significantly lower on one WSQ subscale (social phobia; P = .009), on the CES-D (P = .05), and on the FQ social-phobia subscale (P = .03).

Discussion

Principal Results

It takes about two minutes to complete the WSQ to detect common mental disorders. The WSQ quickly detects clinically-relevant mood, anxiety, and alcohol-related problems and so can guide Internet users to Internet-self-help modules appropriate for their problem, or quickly screen patients prior to consultation with a GP. This measure can also be used in more homogeneous samples to screen out people with co-morbid disorders. The WSQ turned out to be a valid screener for social phobia, panic disorder with agoraphobia, agoraphobia, OCD, and alcohol abuse/dependence (sensitivity: 0.72 - 1.00; specificity: 0.63 - 0.80), and appropriate for depressive disorder, GAD, PTSD, specific phobia, and panic disorder (without agoraphobia) (sensitivity: 0.80 - 0.93; specificity: 0.44 - 0.51) in our study population. Interestingly, the AUC's of the WSQ's scaled single items, and some of the dichotomous items, were comparable to the AUC's of the longer questionnaires, supporting our conclusion that short questionnaires, sometimes with just one item, can be as valid as longer ones. This is in line with previous studies [7,41-43].

Compared to psychometric properties of other online screening questionnaires [8,10] (sensitivity: 0.63 - 0.95; specificity: 0.73 - 0.97), WSQ's sensitivity was similar (sensitivity: 0.72 - 1.00), but specificity was, for some disorders, considerably lower (specificity: 0.44 - 0.80). One explanation for this lower specificity might be that we have used 6-month prevalence rates rather than point prevalence rates, whereas the WSQ assesses current symptoms rather than symptoms during the previous 6 months. Therefore, specificity might be higher when the WSQ is validated against concurrent DSM-IV diagnoses. Although only one of the two symptoms is required for a diagnosis of MDD, the "WSQ depression diagnosis" is based on elevated mood and anhedonia. However, when only one of the two symptoms would give a positive "WSQ depression diagnosis", specificity was below the threshold level of 0.40. Therefore, both core depression symptoms are needed to fulfill the criteria of a positive "WSQ depression diagnosis". Although sensitivity, specificity, and NPV's were acceptable for most WSQ "diagnosis", PPV's were low (0.10 - 0.51), indicating that the WSQ misidentified many participants as (falsely) positive. NPV and PPV depend on prevalence. When prevalence is high, which might be the case in self-selected samples such as those in this study, "true" negatives will have a greater impact, and when prevalence is low, "true positives" have a higher impact on the NPV and PPV. When prevalence is low, a positive diagnosis from the WSQ should be regarded with caution. Subjects with a positive WSQ score can then undergo more in-depth screening with a longer questionnaire or CIDI with a higher specificity. However, the test successfully identified "true" negatives (high NPV), which is to say that subjects with no WSQ positive score ("diagnosis") of any kind are likely to have no relevant DSM-IV diagnosis when interviewed by CIDI. In brief, the WSQ screens out negatives well but yields many false positives.

Although WSQ's false positives do not have a diagnosis, they might have symptoms of depression, anxiety, or alcohol problems, since they have elevated scores on the relevant screening questionnaires.

Limitations

One limitation of our study is that the CIDI-diagnosis live phone interviews were not taped, so inter-rater reliability could not be calculated. Second, subjects always completed the WSQ on the Internet before the other screening questionnaires, so order effects could not be ruled out. Third, though sensitivity and specificity do not depend on prevalence of the disorders in the population, the PPV and NPV do; consequently, the values we found might not generalize to situations where prevalence is different. Fourth, it is not known how representative our self-recruited participants are of Internet self-help applicants. Fifth, subjects who had a CIDI interview had significantly less social phobia on that WSQ-subscale than those who did not, so the WSQ-social-phobia results might be less generalizable to other populations. Sixth, as described earlier, 6-month prevalence rates of DSM-IV diagnoses were used, whereas the WSQ assesses current symptoms. Ideally, the WSQ should be validated against concurrent DSM-IV diagnoses. Seventh, norms are unavailable for acceptable levels of sensitivity and specificity which depend on the test's aim, costs, and benefits [38]. As the WSQ aims to detect clinically-relevant mood, anxiety, and alcohol-related problems in order to minimize missed cases, we chose thresholds of sensitivity at 0.70 or more and of specificity at 0.40 or more. Finally, the WSQ for common mental disorders could be further simplified [44]. However, before using this simplified WSQ, psychometric properties have to be evaluated.

Despite its limitations, the WSQ is a useful and quick Internet screening tool to detect people likely to have common mental disorders.

Future Research

Many false positives were found for WSQ subscales GAD, panic, specific phobia, and PTSD, while far fewer false positives were found for alcohol abuse/dependence, social phobia, panic disorder with agoraphobia, and OCD. The high rate of false positives may, for some questions, be due to a lack of clarity or classification criteria. Future research which enhances clarity of questions and classification criteria is needed to improve the predictive power of the WSQ.

Acknowledgments

This study is funded by the Faculty of Psychology and Education of the VU University, Amsterdam.

Conflicts of Interest

None declared.

Multimedia Appendix 1

WSQ

[PDF file (Adobe PDF), 117 KB - jmir_v11i3e19_app1.pdf]

References

1. Marks IM, Cavanagh K, Gega L. Computer-aided psychotherapy: revolution or bubble? Br J Psychiatry 2007 Dec;191(6):471-473 [FREE Full text] [Medline: 18055948] [doi: 10.1192/bjp.bp.107.041152]
2. Andersson G, Cuijpers P, Carlbring P, Lindefors N. Effects of Internet-delivered cognitive behaviour therapy for anxiety and mood disorders. Psychiatry 2007;1(2):9-14.
3. Carlbring P, Nilsson-Ihrfelt E, Waara J, Kollenstam C, Buhrman M, Kaldo V, et al. Treatment of panic disorder: live therapy vs. self-help via the Internet. Behav Res Ther 2005 Oct;43(10):1321-1333. [Medline: 16086983] [doi: 10.1016/j.brat.2004.10.002]
4. Angermeyer MC, Dietrich S. Public beliefs about and attitudes towards people with mental illness: a review of population studies. Acta Psychiatr Scand 2006 Mar;113(3):163-179. [Medline: 16466402] [doi: 10.1111/j.1600-0447.2005.00699.x]
5. Austin DW, Carlbring P, Richards JC, Andersson G. Internet administration of three commonly used questionnaires in panic research: equivalence to paper administration in Australian and Swedish samples of people with panic disorder. International Journal of Testing 2006;6(1):25-39. [doi: 10.1207/s15327574ijt0601_2]
6. Buchanan T. Internet-based questionnaire assessment: appropriate use in clinical contexts. Cogn Behav Ther 2003;32(3):100-109. [Medline: 16291542] [doi: 10.1080/16506070310000957]
7. Cuijpers P, Smits N, Donker T, ten Have M, de Graaf R. Screening for mood and anxiety disordes with the five-item, the three-item, and the two-item Mental Health Inventory. Psychiatry Res 2009 Jan 28 (forthcoming) . [doi: 10.1016/j.psychres.2008.05.012]
8. Farvolden P, McBride C, Bagby RM, Ravitz P. A Web-based screening instrument for depression and anxiety disorders in primary care. J Med Internet Res 2003 Sep 29;5(3):e23 [FREE Full text] [Medline: 14517114] [doi: 10.2196/jmir.5.3.e23]
9. Gega L, Kenwright M, Mataix-Cols D, Cameron R, Marks IM. Screening people with anxiety/depression for suitability for guided self-help. Cogn Behav Ther 2005;34(1):16-21. [Medline: 15844684] [doi: 10.1080/16506070410015031]
10. Lin CC, Bai YM, Liu CY, Hsiao MC, Chen JY, Tsai SJ, et al. Web-based tools can be used reliably to detect patients with major depressive disorder and subsyndromal depressive symptoms. BMC Psychiatry 2007;7(1):12 [FREE Full text] [Medline: 17425774] [doi: 10.1186/1471-244X-7-12]
11. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Applied Psychological Measurement 1977;1(3):385-401 . [doi: 10.1177/014662167700100306]
12. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097 [FREE Full text] [Medline: 16717171] [doi: 10.1001/archinte.166.10.1092]
13. Marks IM, Mathews AM. Brief standard self-rating for phobic patients. Behav Res Ther 1979;17(3):263-267. [Medline: 526242] [doi: 10.1016/0005-7967(79)90041-X]
14. Shear MK, Rucci P, Williams J, Frank E, Grochocinski V, Vander Bilt J, et al. Reliability and validity of the Panic Disorder Severity Scale: replication and extension. J Psychiatr Res 2001;35(5):293-296. [Medline: 11591432] [doi: 10.1016/S0022-3956(01)00028-0]
15. Houck PR, Spiegel DA, Shear MK, Rucci P. Reliability of the self-report version of the panic disorder severity scale. Depress Anxiety 2002;15(4):183-185. [Medline: 12112724] [doi: 10.1002/da.10049]
16. Goodman WK, Price LH, Rasmussen SA, Mazure C, Delgado P, Heninger GR, et al. The Yale-Brown Obsessive Compulsive Scale. II. Validity. Arch Gen Psychiatry 1989 Nov;46(11):1012-1016. [Medline: 2510699]
17. Goodman WK, Price LH, Rasmussen SA, Mazure C, Fleischmann RL, Hill CL, et al. The Yale-Brown Obsessive Compulsive Scale. I. Development, use, and reliability. Arch Gen Psychiatry 1989 Nov;46(11):1006-1011. [Medline: 2684084]
18. Horowitz M, Wilner N, Alvarez W. Impact of Event Scale: a measure of subjective stress. Psychosom Med 1979 May;41(3):209-218 [FREE Full text] [Medline: 472086]
19. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. Addiction 1993 Jun;88(6):791-804. [Medline: 8329970] [doi: 10.1111/j.1360-0443.1993.tb02093.x]
20. ; World Health Organization. Composite International Diagnostic Interview (CIDI): Version 2.1. Geneva: World Health Organization; 1997.
21. ; American Psychiatric Association. Diagnosis and Statistical Manual of Mental Disorders. 4th edition. Washington, DC: American Psychiatric Association; 1994.
22. ; Faculty of Psychology and Education. Web Screening Questionnaire for Common Mental Disorders (WSQ). VU University Amsterdam. URL: http://webscreeningquestionnaire.org/files/WSQ.pdf [accessed 2009 Apr 6] [WebCite Cache ID 5fqVoAKCG]

23. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20(Suppl 20):22-33;quiz 34-57. [Medline: 9881538]
24. Beekman AT, Deeg DJ, Van Limbeek J, Braam AW, De Vries MZ, Van Tilburg W. Criterion validity of the Center for Epidemiologic Studies Depression scale (CES-D): results from a community-based sample of older subjects in The Netherlands. Psychol Med 1997 Jan;27(1):231-235. [Medline: 9122304] [doi: 10.1017/S0033291796003510]
25. Cuijpers P, Boluijt P, van Straten A. Screening of depression in adolescents through the Internet : sensitivity and specificity of two screening questionnaires. Eur Child Adolesc Psychiatry 2008 Feb;17(1):32-38. [Medline: 17876508] [doi: 10.1007/s00787-007-0631-2]
26. Van Zuuren FJ. The fear questionnaire. Some data on validity, reliability and layout. Br J Psychiatry 1988 Nov;153(5):659-662. [Medline: 3255455] [doi: 10.1192/bjp.153.5.659]
27. Cox BJ, Parker JD, Swinson RP. Confirmatory factor analysis of the Fear Questionnaire with social phobia patients. Br J Psychiatry 1996 Apr;168(4):497-499. [Medline: 8730947] [doi: 10.1192/bjp.168.4.497]
28. Cox BJ, Swinson RP, Shaw BF. Value of the Fear Questionnaire in differentiating agoraphobia and social phobia. Br J Psychiatry 1991 Dec;159(6):842-845. [Medline: 1790455] [doi: 10.1192/bjp.159.6.842]
29. Hoyer J, Becker ES, Neumer S, Soeder U, Margraf J. Screening for anxiety in an epidemiological sample: predictive accuracy of questionnaires. J Anxiety Disord 2002;16(2):113-134. [Medline: 12194539] [doi: 10.1016/S0887-6185(01)00094-9]
30. van der Ploeg E, Mooren TTM, Kleber RJ, van der Velden PG, Brom D. Construct validation of the Dutch version of the impact of event scale. Psychol Assess 2004 Mar;16(1):16-26. [Medline: 15023089] [doi: 10.1037/1040-3590.16.1.16]
31. de Haan L, Hoogeboom B, Beuk N, Wouters L, Dingemans PMAJ, Linszen DH. Reliability and validity of the Yale-Brown Obsessive-Compulsive Scale in schizophrenia patients. Psychopharmacol Bull 2006;39(1):25-30 [FREE Full text] [Medline: 17065972]
32. van Oppen P, van Balkom AJLM, de Haan E, van Dyck R. Cognitive therapy and exposure in vivo alone and in combination with fluvoxamine in obsessive-compulsive disorder: a 5-year follow-up. J Clin Psychiatry 2005 Nov;66(11):1415-1422. [Medline: 16420079]
33. Conigrave KM, Saunders JB, Reznik RB. Predictive capacity of the AUDIT questionnaire for alcohol-related harm. Addiction 1995 Nov;90(11):1479-1485. [Medline: 8528033] [doi: 10.1111/j.1360-0443.1995.tb02810.x]
34. Janca A, Ustün TB, Sartorius N. New versions of World Health Organization instruments for the assessment of mental disorders. Acta Psychiatr Scand 1994 Aug;90(2):73-83. [Medline: 7976462] [doi: 10.1111/j.1600-0447.1994.tb01559.x]
35. Smeets RMR; Dingemans PMAJ. Composite International Diagnostic Interview (CIDI), Version 1.1, Interviewers Version, Dutch Translation. World Health Organization edition. Amsterdam/Geneva: World Health Organization; 1993.
36. Andrews G, Peters L. The psychometric properties of the Composite International Diagnostic Interview. Soc Psychiatry Psychiatr Epidemiol 1998 Feb;33(2):80-88. [Medline: 9503991] [doi: 10.1007/s001270050026]
37. Wittchen HU. Reliability and validity studies of the WHO--Composite International Diagnostic Interview (CIDI): a critical review. J Psychiatr Res 1994;28(1):57-84. [Medline: 8064641] [doi: 10.1016/0022-3956(94)90036-1]
38. Smits N, Smit F, Cuijpers P, De Graaf R. Using decision theory to derive optimal cut-off scores of screening instruments: an illustration explicating costs and benefits of mental health screening. Int J Methods Psychiatr Res 2007;16(4):219-229 [FREE Full text] [Medline: 18188835] [doi: 10.1002/mpr.230]
39. Cairney J, Veldhuizen S, Wade TJ, Kurdyak P, Streiner DL. Evaluation of 2 measures of psychological distress as screeners for depression in the general population. Can J Psychiatry 2007 Feb;52(2):111-120. [Medline: 17375867]
40. Fischer JE, Bachmann LM, Jaeschke R. A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. Intensive Care Med 2003 Jul;29(7):1043-1051. [Medline: 12734652] [doi: 10.1007/s00134-003-1761-8]
41. Lim PP, Ng LL, Chiam PC, Ong PS, Ngui FT, Sahadevan S. Validation and comparison of three brief depression scales in an elderly Chinese population. Int J Geriatr Psychiatry 2000 Sep;15(9):824-830. [Medline: 10984729] [doi: 10.1002/1099-1166(200009)15:9<824::AID-GPS207>3.0.CO;2-C]
42. McKenzie N, Marks I. Quick rating of depressed mood in patients with anxiety disorders. Br J Psychiatry 1999 Mar;174(3):266-269. [Medline: 10448454] [doi: 10.1192/bjp.174.3.266]
43. Mitchell AJ, Coyne JC. Do ultra-short screening instruments accurately detect depression in primary care? A pooled analysis and meta-analysis of 22 studies. Br J Gen Pract 2007 Feb;57(535):144-151 [FREE Full text] [Medline: 17263931]
44. ; Faculty of Psychology and Education. Web Screening Questionnaire for Common Mental Disorders (WSQ) (simple version). VU University Amsterdam. URL: http://webscreeningquestionnaire.org/files/WSQ_simple_version.pdf [accessed 2009 Apr 6] [WebCite Cache ID 5fqVaHWE1]

Abbreviations

AUC: area under the curve
AUDIT: alcohol use disorders identification test
CES-D: Center for Epidemiological Studies Depression scaleCIDI:composite international diagnostic interview

Dyst: dysthymia
DSM-IV: Diagnostic Statistical Manual, 4thedition
FQ: fear questionnaire
GAD: generalized anxiety disorder
GAD-7: generalized anxiety disorder - 7
GPs: general practitioners
IES: impact of events scale
ISP-D: Internet-based self-assessment program for depression
M: mean
MinD: minor depression
MDD: major depressive disorder
MINI: mini-international neuropsychiatric interview
NPV: negative predictive value
OCD: obsessive compulsive disorder
PDSS-SR: panic disorder severity scale self-report
PPV: positive predictive value
PTSD: Post-Traumatic Stress Disorder
SD: standard deviation
SQ: screening questionnaire
VU: Vrije Universiteit
WB-DAT: Web-based depression and anxiety test
WHO: World Health Organization
WSQ: Web screening questionnaire
YBOCS: Yale-Brown Obsessive Compulsive Scale

Edited by L Ritterband; submitted 21.09.08; peer-reviewed by B Klein, J Seeley; comments to author 18.11.08; revised version received 06.03.09; accepted 06.03.09; published 24.07.09.

Please cite as:
Donker T, Straten AV, Marks I, Cuijpers P
A Brief Web-Based Screening Questionnaire for Common Mental Disorders: Development and Validation
J Med Internet Res 2009;11(3):e19
URL: http://www.jmir.org/2009/3/e19/
doi:10.2196/jmir.1134
PMID:19632977

Original Paper

An Internet Tool for Creation of Cancer Survivorship Care Plans for Survivors and Health Care Providers: Design, Implementation, Use and User Satisfaction

Christine E Hill-Kayser[1], MD; Carolyn Vachani[1], MSN, RN, AOCN; Margaret K Hampshire[1], RN, BSN, OCN; Linda A Jacobs[2], PhD, RN; James M Metz[1], MD

[1]Department of Radiation Oncology, University of Pennsylvania, Philadelphia, PA, USA

[2]Department of Medical Oncology, University of Pennsylvania, Philadelphia, PA, USA

Corresponding Author:
Christine E Hill-Kayser, MD
Department of Radiation Oncology
Hospital of the University of Pennsylvania
3400 Spruce Street, 2 Donner
Philadelphia, PA 19104
USA
Phone: +1 215 662 2428
Fax: +1 215 349 5445
Email: hill@xrt.upenn.edu

Abstract

Background: Survivorship care plans have been recommended by the Institute of Medicine for all cancer survivors. We implemented an Internet-based tool for creation of individualized survivorship care plans. To our knowledge, this is the first tool of this type to be designed and made publicly accessible.

Objective: To investigate patterns of use and satisfaction with an Internet-based tool for creation of survivorship care plans.

Methods: OncoLife, an Internet-based program for creation of survivorship care plans, was designed by a team of dedicated oncology nurses and physicians at the University of Pennsylvania. The program was designed to provide individualized, comprehensive health care recommendations to users responding to queries regarding demographics, diagnosis, and cancer treatments. After being piloted to test populations, OncoLife was made publicly accessible via Oncolink, a cancer information website based at the University of Pennsylvania which averages 3.9 million page views and over 385,000 unique visits per month. Data entered by anonymous public users was maintained and analyzed.

Results: From May 2007 to November 2008, 3343 individuals utilized this tool. Most (63%) identified themselves as survivors, but also health care providers (25%) and friends/family of survivors (12%). Median age at diagnosis was 48 years (18 - 100+), and median current age 51 (19 - 100+). Most users were Caucasian (87%), female (71%), and college-educated (82%). Breast cancer was the most common diagnosis (46%), followed by hematologic (12%), gastrointestinal (11%), gynecologic (9%), and genitourinary (8%). Of all users, 84% had undergone surgery, 80% chemotherapy, and 60% radiotherapy. Half of users (53%) reported receiving follow-up care from only an oncologist, 13% only a primary care provider (PCP), and 32% both; 12% reported having received survivorship information previously. Over 90% of users, both survivors and health care providers, reported satisfaction levels of “good” to “excellent” using this tool.

Conclusions: Based on our experience with implementation of what is, to our knowledge, the first Web-based program for creation of survivorship care plans, survivors and health care providers appear both willing to use this type of tool and satisfied with the information provided. Most users have never before received survivorship information. Future iterations will focus on expanding accessibility and improving understanding of the needs of cancer survivors in the era of the Internet.

(J Med Internet Res 2009;11(3):e39) doi:10.2196/jmir.1223

KEYWORDS

Survivors; cancer survivor; patient care planning; survivorship care plan; late effects from cancer treatment; survivor issues; Internet

Introduction

Advances in cancer screening, detection, and treatment have increased the numbers of persons considered cured of cancer and those living with cancer as a chronic illness; as a result, the number of cancer survivors living in the United States (US) tripled from 3.0 million in 1971 to 9.8 million in 2001 [1]. A significant portion of the adult population is thus faced not only with the medical needs of normal aging, but with the unique health care concerns associated with cancer diagnosis and treatment, including recurrent and/or residual disease, treatment-related late effects, and threats to psychosocial and economic well-being [2-4].

Despite the unique needs of cancer survivors, this population may be at risk for receiving inadequate health care [5], and several groups have demonstrated that cancer survivors, a growing subset of the population, are at risk not only for cancer recurrence, but for receiving inadequate risk-based and routine preventive health care [5-7]. In response to this, national and international organizations have prioritized issues of survivorship over the past decade. In 2005, The Institute of Medicine (IOM) produced its report *From Cancer Patient to Cancer Survivor: Lost in Transition*. In this publication, the IOM outlined 10 recommendations intended to improve care of, and fiscal support for, cancer survivors. The second of these recommendations called on health care providers to provide patients with a "Survivorship Care Plan," or "a comprehensive care summary and follow-up plan" [8]. This recommendation is based on recognition that many cancer survivors do not receive comprehensive care after active treatment, and that inadequate communication likely contributes to this. Indeed, the majority of primary care providers (PCPs) surveyed rate the current transition process from oncologic care to the PCP as fair or poor [9], and up to one-third of cancer survivors report being unsure of which of their physicians is in charge of their follow-up care [10]. Survivorship care plans are a conduit not only between active cancer care and survivorship care, but between physicians and survivors.

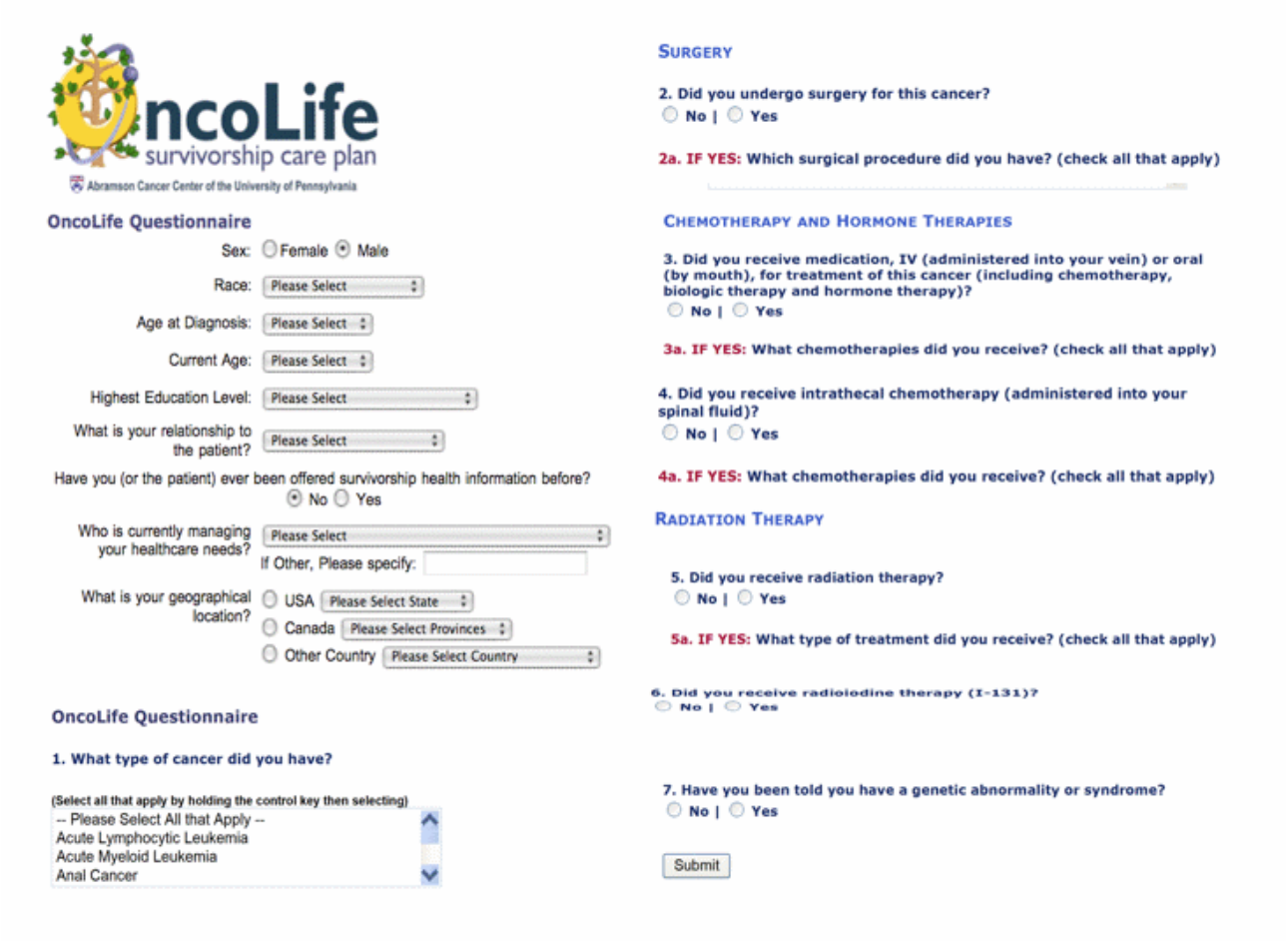
The Internet is an increasing source of health information worldwide. The Pew Internet & American Life Project reported in 2006 that 113 million Americans had used the Internet for health-related purposes; of these, over 50% reported that Internet use impacted their health care [11]. As the complexity of Internet-based systems has increased, several groups have demonstrated improvement in quality of life [12-13] and overall care [14-16] with use of Internet-based coaching for management of pain, diabetes, and heart/lung disease. Cancer is one of the top three diseases about which Internet users seek information [17], and recent studies suggest that the Internet may offer opportunities to actively improve health care for cancer patients and survivors [18]. Use of the Internet to actively manage symptoms related to cancer treatments is currently being examined in a European clinical trial [19].

In May 2007, we launched the world's first Internet-based tool for creation of survivorship care plans, *OncoLife* [20]. *OncoLife* is a publicly accessible tool that is available through *OncoLink* [21], a cancer information website based at the University of Pennsylvania's Abramson Cancer Center. *OncoLife* was designed to supply dynamic, personalized information to cancer survivors, and to prompt interventions with regard to both surveillance and management of late effects when indicated. The launch of *OncoLife* was anticipated to fill an unmet need for survivorship information and care; however, the willingness of survivors and their health care providers to use this type of tool, the satisfaction it would provide them, and the demographic, diagnosis, and treatment characteristics of users could not be predicted. The study described here was undertaken in order to investigate these questions. Our findings, as well as the *OncoLife* design and implementation process, are described here.

Methods

OncoLink is a general cancer information website maintained by physicians and nurses at the Abramson Cancer Center of the University of Pennsylvania, serving 3.9 million pages to over 385,000 unique Internet Protocol (IP) addresses monthly. *OncoLife*, a section of *OncoLink*, was developed by a dedicated team of oncology nurses and physicians. The *OncoLife* format includes a publicly accessible, five-screen series of 17 queries regarding demographics, cancer diagnosis, and cancer treatments received, and provides users with lists from which to select surgeries, sites of radiotherapy, and chemotherapy/biologic agents by both generic and trademark names (Figure 1).

Figure 1. *OncoLife* user interface. The *OncoLife* tool for creation of survivorship care plans is available via *OncoLink* and *OncoLife* websites



The *OncoLife* survey is of open design, accessible to any visitor to the *OncoLink* site, with a target population of cancer survivors, health care providers, and friends/family members of survivors and a convenience sample frame. *OncoLife* is advertised via *OncoLink* (See Multimedia Appendix 1); flyers and bookmarks with the *OncoLife* trademark and website address are also available in our clinic and have been made available to health care providers in other institutions for distribution.

Completion of the *OncoLife* survey results in generation of individualized, detailed, comprehensive survivorship care plans providing surveillance recommendations for tumor recurrence, in addition to guidelines for overall health care in the setting of increased risk for certain morbidities secondary to cancer treatment (See Multimedia Appendix 2). These guidelines have been designed to be specific to types of treatments that patients have received, as well as their primary cancer diagnoses. Guidelines are based on type and site of surgical procedures, radiotherapy sites, and specific drugs received. All survivors are provided with information regarding second malignancy and other global issues pertaining to cancer survivorship. Guidelines are evidence- or consensus-based whenever possible and are in accordance with guidelines provided by the IOM, Children's Oncology Group (COG), National Cancer Institute (NCI), and American Society of Clinical Oncology (ASCO). In areas in which evidence- or consensus-based guidelines are not available, guidelines provided are based on practice at our own institution. All guidelines provided as part of *OncoLife* survivorship care plans have been constructed with both nursing and physician input and are described in plain language.

The *OncoLife* survey and the information provided to survivors using *OncoLife* were piloted with groups of survivors who tested usability and technical functionality, prior to the public launch. The pilot version of *OncoLife* included queries regarding cancer diagnosis and treatment only (Table 1). Following the pilot process, queries regarding demographics were added, and version 1 was made publicly accessible. Over the course of the 18 months following the implementation of *OncoLife*, three further iterations were developed with the intent of increasing the comprehensive nature of the survivorship care plans produced by *OncoLife*, as well as increasing accessibility through improved understanding of the user population. Changes incorporated into iterations were based on user feedback, as well as observations regarding use patterns. Version 2 included additional queries regarding follow-up care and the availability of survivorship care and also requested that individuals completing the survey describe themselves as survivors, friends/family members of a survivor, or health care providers. Version 3 made use of the same series of queries used in version 2, but provided more individualized and extensive information to survivors, including adaptive questioning regarding surgeries, radiotherapy, and chemotherapies implemented for survivors specifically of breast cancer. Additionally, with the launch of version 3, a five-question, one-page user satisfaction survey

was added through an optional link accessible upon receipt of survivorship care plans. Version 4, launched in January 2009, includes queries regarding menopausal status to allow further individualization of guidelines provided.

Table 1. *OncoLife* queries and response options according to version (vers.)

		Pilot Version	Vers. 1	Vers. 2[a]	Vers. 3[a]	Vers. 4
Number of Users (Non-duplicate)		40	1374	1124	805	Recent Launch
OncoLife Query	Response Options					
Sex	Male		•	•	•	•
	Female					
Race	Caucasian		•	•	•	•
	African American					
	Asian/Pacific Islander					
	Hispanic/Latino/a					
	Mixed Race					
	Other					
Age at Diagnosis	Please select		•	•	•	•
Current Age	Please select		•	•	•	•
Highest Education Level	Grade School		•	•	•	•
	High School					
	Some College					
	College Degree					
	Graduate School					
What is your relationship to the patient?	Self			•	•	•
	Family member/friend					
	Health care provider					
Have you ever been offered survivorship health information before?	Yes			•	•	•
	No					
Who is currently managing your health care needs?	Oncologist			•	•	•
	PCP/internist					
	Oncologist and PCP					
	Other (specify)					
What is your geographical location?	USA (select state)			•	•	•
	Canada (select province)					
	Other country (select)					
What type of cancer did you have?	Please select	•	•	•	•	•
Did you undergo surgery for this cancer?	Yes (select procedure)	•	•	•	•	•
	No					
Did you receive medication, intravenous or oral, for treatment of this cancer?	Yes (select medication[s])	•	•	•	•	•
	No					
Did you receive intrathecal chemotherapy?	Yes (select medication)	•	•	•	•	•
	No					
Did you receive radiation therapy?	Yes (select site)	•	•	•	•	•
	No					
Did you receive radioiodine therapy (I-131)?	Yes	•	•	•	•	•
	No					

		Pilot Version	Vers. 1	Vers. 2[a]	Vers. 3[a]	Vers. 4
Number of Users (Non-duplicate)		40	1374	1124	805	Recent Launch
OncoLife Query	Response Options					
Have you ever been told you have a genetic abnormality or syndrome?	Yes	•	•	•	•	•
	No					
What is your menopausal status? (females only)	Menopause before cancer therapy					•
	Postmenopausal (due to surgery or chemo/radiotherapy)					
	Premenopausal					
	Perimenopausal					
	Not sure					

[a] Versions 2 and 3 differ only by survivorship care plans (SCP) generated.

In addition to its evolution through these versions, *OncoLife* has been completely translated into Spanish. Spanish translation was, and continues to be, performed by a bilingual (English- and Spanish-speaking) health care provider practicing in the field of oncology, with culturally relevant revisions occasionally made to the wording used on the website.

OncoLife remains an anonymous tool, and users are not asked for identifying information. Prior to submission of the *OncoLife* survey, users are able to review and change answers; however, in order to protect and ensure anonymity, users are not asked to "log in," and entries are not maintained or saved for reuse or review at a later date. Use of *OncoLife* surveys is completely voluntary, with production of the survivorship care plan being the only incentive for use. Survivorship care plans produced using *OncoLife* were designed to address issues faced by adult cancer survivors. Pediatric cancer survivors are referred on the *OncoLife* introductory page to the COG website guidelines for survivors of childhood cancer [22]. *OncoLife* survivorship care plans are intended to provide guidance for survivors and physicians providing follow-up care to survivors, and they are not intended to replace interactions or recommendations provided by health care providers of individual survivors. Instead, plans may serve as aids for communication between survivors and their caregivers.

Data from each use of *OncoLife* have been maintained anonymously on a secure server, with automatic database entry. Data collection and maintenance procedures were approved by the Institutional Review Board (IRB) prior to the launch of *OncoLife*. Only data from completed questionnaires are recorded and/or analyzed—JavaScript encryption ensures that surveys cannot be submitted without completion of all queries. Where appropriate, queries provide non-response options (such as "I don't know," or "not applicable"). Data are password protected and are available only to the small team of physicians and nurses (five in total) involved in the creation of *OncoLife*. Entries are screened by IP address to avoid analysis of duplicate entries. A Chi-squared contingency table with one degree of freedom was used to compare user survey data regarding availability of information reported by survivors versus health care providers; an exact contingency test was used to compare satisfaction data after binning of Likert-type responses between the two groups to account for sparse cell population [23].

Results

Between May 2007 and November 2008, 3647 *OncoLife* surveys were completed, 40 using a pilot version, 1562 using version 1, 1211 using version 2, and 834 using version 3. Based on duplicate IP address and data entry, 304 of these were identified as duplicates, leaving 3343 unique *OncoLife* users. Of these, 79 reported more than one cancer diagnosis. Of the 3343 responders, the median age at the time of cancer diagnosis was 48 years (mean 48, range 18 - 100+). Median current age was 51 years (mean 51, range 18 - 100+). The majority of users were women (71.3%, n = 2385) and described themselves as Caucasian (85.6%, n = 2861) and college-educated (78.2%, n = 2617) (Table 2). Of 1880 users who completed *OncoLife* surveys after the implementation of its second version, most described themselves as survivors (64.2%, n = 1198), although significant proportions were health care providers (24.8%, n = 461) and friends/family members of survivors (12.4%, n = 221) (Table 2). Health care providers were predominantly nurses (61.8%, n = 285) and nurse practitioners (23.2%, n = 107). Of 1872 users for whom data on location were available, the majority (91.0%, n = 1704) were US residents, representing 48 different states, and 5.9% (n = 110) were Canadian. The remaining 3% (n = 58) of users were residents of 24 other countries.

Table 2. Demographic information reported by users of *OncoLife*

Demographic	Total n = 3343	%
Sex		
Male	957	28.6
Female	2385	71.3
Race		
Caucasian	2861	85.6
African American	179	5.4
Asian/Pacific Islander	91	3
Hispanic/Latino/a	85	3
Mixed Race	36	1
Other	47	1
Unknown	40	1
Education		
Grade School	73	2
High School	612	18.3
Some College	699	20.9
College Degree	1107	33.1
Graduate School	811	24.3
Unknown	40	1
Relationship to Patient	n = 1880[a]	
Self	1198	64.2
Family member/friend	221	12.4
Health care provider	461	24.8
Nurse	285	61.8[b]
Nurse practitioner	107	23.2[b]
Physician	48	10[b]
Other health care	26	6[b]

[a] Query added with implementation of version 2, so data available for n = 1880 users.

[b] Refers to percent of health care providers (n = 461)

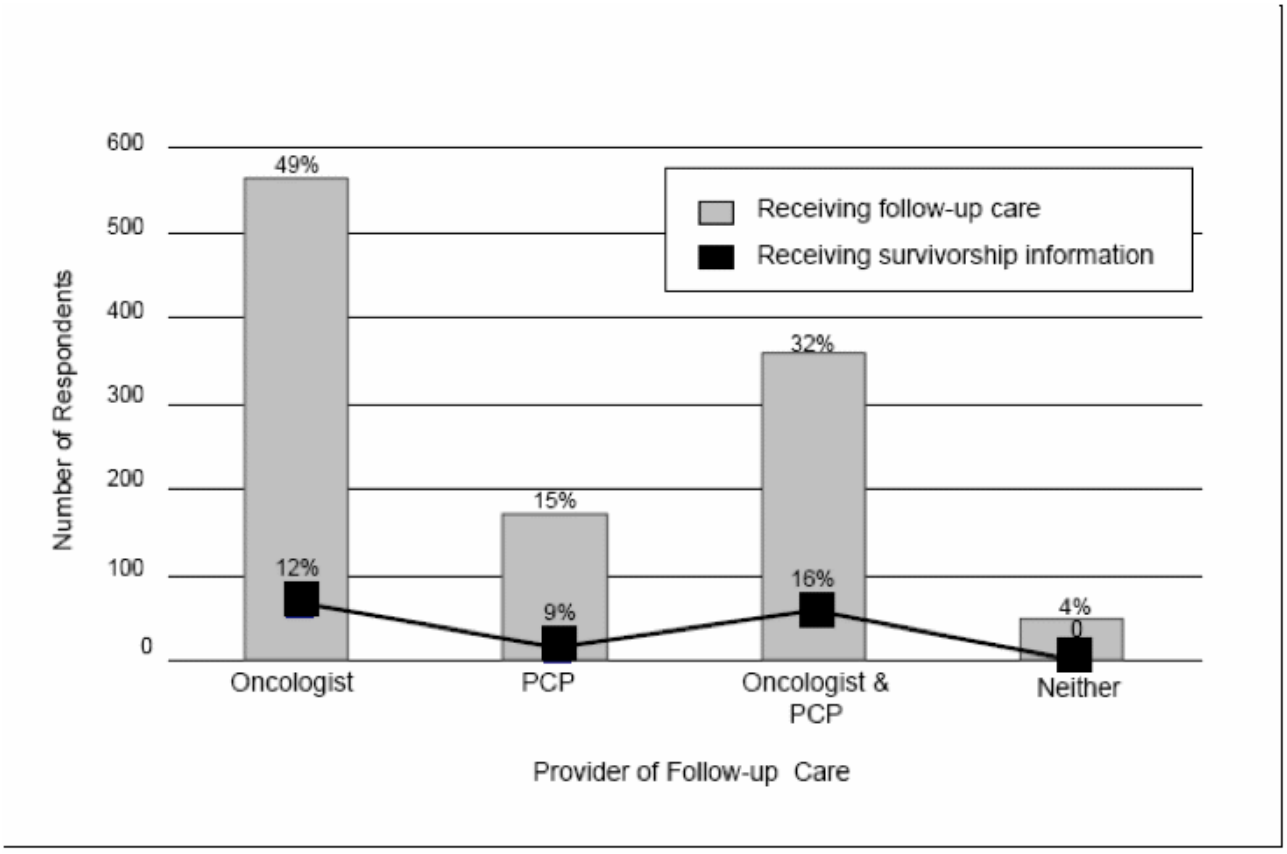
Breast cancer represented the most commonly reported primary cancer diagnosis (45.9%, n = 1537) among the 3343 *OncoLife* users, followed by hematologic (12.0%, n = 401), gastrointestinal (11.7%, n = 391), gynecologic (8.6%, n = 287), and genitourinary malignancies (8.3%, n = 278) (Table 3). Overall, 79.8% of these 3343 users (n = 2670) reported being treated with chemotherapy, 59.0% (n = 1973) with radiotherapy, and 83.5% (n = 2793) with surgery.

Table 3. Primary cancer diagnoses among users of *OncoLife*

Primary Cancer Diagnosis/Site	Number of *OncoLife* Users	%
Breast	1537	45.9
Hematologic	401	12.0
Gastrointestinal	391	11.7
Gynecologic	287	8.6
Genitourinary	278	8.3
Thoracic	149	4.5
Head & Neck	90	3
Melanoma	69	2
Central Nervous System	59	2
Thyroid	46	1
Sarcoma	38	1
Non-Melanoma Skin Cancer	17	< 1
Other	81	2

Of 1869 users who provided information regarding follow-up care, half (52.5%, n = 982) reported receiving follow-up care only from an oncologist, and only 12.6% (n = 235) reported having previously received information on cancer survivorship (Figure 2). The majority of patients having received survivorship information prior to *OncoLife* were followed by an oncologist: Of these 235 patients, 89.7% (n = 211) reported receiving follow-up care from an oncologist, and 10% (n = 14) only from a PCP.

Figure 2. Follow-up care and survivorship information offered to users of *OncoLife*



The user satisfaction survey was launched in July 2008, and 150 satisfaction surveys were completed. Of these, 57% (n = 86) were completed by survivors or family members/friends of survivors. The remaining 43% (n = 64) were completed by health care providers. According to overall user response, *OncoLife* survivorship care plan questionnaires took an average of 6.7 minutes to complete (range 1 - 30 minutes). Health care providers reported average time of 4.4 minutes, compared to

7.2 minutes reported by survivors and friends/family members. Overall, over 90% of users rated their experience and level of satisfaction using *OncoLife as* "good," "very good," or "excellent." Specifically, 98% (n = 64) of the 65 health care providers rated their experiences as "good" to "excellent." This was similar to 95% (n = 81) of the 86 survivors/friends/family members rating their experience "good" to "excellent" (probability = 0.2, *P* = .39). Most users (92%, n = 138) felt that they had the information needed to complete the *OncoLife* questionnaire, and this did not differ significantly between health care providers and survivors/friends/family members (89% [n = 57] versus 94% [n = 81] respectively [$\chi^2 = 0.16 \times 10^{-3}$, *P* = .99]). Most survivors (83%, n = 71) answered that they would plan to share the information provided with their health care team. Health care providers reported "good" to "excellent" levels of satisfaction with the information provided to the patient via *OncoLife* in 95% of cases (Table 4).

Table 4. *OncoLife* user satisfaction survey queries and responses

OncoLife User Satisfaction Survey Query	Response Options	Total Responses n = 150 (%)	Number Health Care Provider n = 64 (%)	Number Survivors/ Family Members/ Friends n = 86 (%)	*P*
How would you rate your experience completing this survey?	1 = Poor	0	0	0	.39[a]
	2 = Fair	5 (3)	1(2)	4 (5)	
	3 = Good	18 (12)	11 (17)	7 (8)	
	4 = Very good	43 (27)	16 (25)	27 (31)	
	5 = Excellent	84 (56)	36 (56)	48 (56)	
How long did it take to complete (in minutes) (Free entry)	Mean	6.7	4.4	7.2	n/a
	Median	5	5	5	
Did you have all of the information needed to complete the questionnaire?	Yes	138 (92)	57 (89)	81 (94)	.99
	No	12 (8)	7 (11)	5 (6)	
Was the information helpful?[b]	Yes		n/a	83 (97)	n/a
	No		n/a	3 (3)	
Will you share your plan with your health care team?[b]	Yes		n/a	71 (83)	n/a
	No		n/a	15 (17)	
How satisfied are you with the information provided to the patient?[c]	1 = Poor		0	n/a	n/a
	2 = Fair		3 (5)	n/a	
	3 = Good		17 (27)	n/a	
	4 = Very good		15 (23)	n/a	
	5 = Excellent		29 (45)	n/a	

[a] For comparison of users responding "good" - "excellent" vs "fair" - "poor," based on exact contingency test [23].

[b] Query posed to survivors/family members/friends only

[c] Query posed to health care providers only

Discussion

Here we describe the design and implementation of, as well as use patterns and user satisfaction with, Internet-based survivorship care plans for cancer survivors. To our knowledge, *OncoLife* is the first such tool to be made publicly available. The intent of survivorship care plans is multifold: Plans are developed to assist with communication between physicians of various specialties, to increase physician-patient communication, and to increase awareness in both the physician and survivor populations of known and suspected late-effects associated with cancer and its treatments. The use of the Internet to allow creation of survivorship care plans allows information to be widely accessible and instantly available.

Based on the NCI definition of a cancer survivor, which includes all people diagnosed with cancer, as well as their caregivers, as survivors, several phases of survivorship certainly exist, and the needs of any one survivor may change dramatically over time. *OncoLife* survivorship care plans may be of use to survivors in any phase—from the moment of diagnosis until the end of life; however, their design may be most appropriate for those survivors who have completed cancer treatment or who continue to receive long-term cancer treatment. For this group of survivors, for whom the acute phase of cancer treatment may have ended, the designation of which health care provider(s) will provide various types of health care may be particularly ambiguous. This transition period may be associated with both survivor and physician uncertainty and dissatisfaction [9,10], potentially leading to important disparities in health care. Although all survivors may potentially be at risk for receiving

inadequate health care after cancer treatment, prior data suggests those followed by an oncologic specialist may be more likely to receive adequate screening for late effects and disease recurrence [6], while those followed by a PCP may be more likely to receive adequate preventive care [5,7]. In reality, however, most survivors do not appear to be followed by both types of providers [7], a finding that is confirmed by users of *OncoLife:* Approximately one third of survivors using *OncoLife* reported routinely receiving follow-up care from both a PCP and an oncologist. Of the remaining two-thirds, the majority reported seeing only an oncologist. These findings emphasize the need for comprehensive communication among physicians and between physicians and survivors. The vast majority of survivors utilizing *OncoLife* reported never having received survivorship information, and this suggests a broad communication deficit. Both PCP feedback and the improvement in comprehensive care when survivors are followed by multiple physicians indicate that gaps in communication are a significant barrier to care of cancer survivors. Survivorship care plans are a communication bridge between physicians and survivors, allowing all of the individuals involved in a survivor's care (including the survivor) to be aware of survivorship health issues and to be assured that they are addressed.

According to the IOM, survivorship care plans should address issues of health maintenance, cancer screening, healthy behaviors, late effects of treatment, possible signs of recurrence, second malignancy risk, and financial consequences of cancer, and they should offer referrals to follow-up providers and lists of cancer-related resources [8]. Not surprisingly, in our current milieu of shrinking resources, the oncology community has expressed concern regarding time and monetary constraints limiting the feasibility of offering survivorship care plans, specifically voicing concerns that a survivorship care plan tool requiring more than 20 minutes per patient would be unrealistic [25]. The IOM recommended in its report that the service of provision of survivorship care plans "be reimbursed by third-party payers of health care" [8]. Hopefully, this concept will become reality in the future—the Comprehensive Cancer Care Improvement Act, currently under consideration in the US (HR. 1078/S. 2790), would allow Medicare reimbursement for oncologists to create survivorship care plans. In the meantime, *OncoLife* has been designed as a free service that does not rely on insurance re-imbursement, and *OncoLife* surveys take on average less than 7 minutes to complete. Both survivors and health care providers report high levels of satisfaction utilizing *OncoLife*, a tool that provides survivors with timely, comprehensive information that addresses the goals delineated by the IOM without insurance or payment delays.

The data presented here demonstrate that survivors, as well as their family members, friends, and health care providers, appear to be willing to use this type of tool. From our data, certain subsets of survivors appear more likely to use *OncoLife* than others—breast cancer survivors represent approximately one-quarter of adult cancer survivors living in the US today (22%) [1] and 45% of *OncoLife* users. This stands in contrast to prostate cancer survivors, who represent the second most prominent survivor population in the US (17%) [1] but only 6% of *OncoLife* users. The disproportionately low use of *OncoLife* by prostate cancer survivors is in all likelihood multifactorial and may have to do with decreased awareness of survivorship issues in this population when compared to the breast cancer survivor population. Another contributing factor may be the overall increased frequency of Internet use by women as opposed to men for health care needs [26-27]. Additionally, *OncoLife* users were predominantly Caucasian, well-educated, and young when compared to the overall survivor population. In 2001, persons over 65 years represented 61% of all cancer survivors [1], while the median age of *OncoLife* users was 51 years. These findings may reflect increased Internet access and level of comfort with Internet use among younger survivors, and are consistent with findings from other groups demonstrating increased Internet use in young, highly educated cancer survivors [28] and under representation of African Americans in online cancer support groups [29]. Since the initial development of *OncoLife*, efforts have been made to increase accessibility to underserved populations, including translation of *OncoLife* into Spanish, distribution of *OncoLife* materials at national meetings to health care providers for distribution to patients, and use of *OncoLife* by nurses at the University of Pennsylvania who complete surveys for patients when they complete cancer treatment. The vast majority of health care providers utilizing *OncoLife* are nurses, and oncology nurses represent a tremendous resource for provision of survivorship care plans to survivors with limited access to the Internet. Efforts are underway to raise awareness among nurses nationwide of the *OncoLife* tool. Efforts to further increase accessibility will continue with future iterations. Additionally, as more centers make use of computer-based data gathering by and for patients, we expect that availability of *OncoLife* to patients completing cancer treatment will continue to increase.

The anonymous nature of *OncoLife* has been maintained in order to protect user privacy and alleviate survivor fear of discrimination following cancer diagnosis; data obtained via *OncoLife* use is, however, limited by its anonymous nature. Data are strictly based on user responses and cannot be verified or validated. National efforts are ongoing to provide cancer survivors with comprehensive summaries of all cancer treatments received, which can then be entered directly into a tool such as *OncoLife*. Future versions of this program may be interfaced directly with electronic medical records to ensure accuracy of all data. Although nearly all users reported having access to the information needed to create a care plan using *OncoLife*, it is conceivable that other survivors might not utilize the tool because of limited access to information needed to complete the survey. Additionally, because users are not required to "log in," plans are not currently saved on our system, although they may be printed and/or converted to electronic files for users themselves to save (both options are available at the time of survivorship care plan production). Future *OncoLife* iterations may be developed with a log in option, so that users may return to their own plans and update their information in order to received updated guidelines. Other limitations of *OncoLife* are associated with the current lack of evidence allowing construction of guidelines for follow-up care of patients after cancer. Our data suggest that most cancer survivors utilizing *OncoLife* have undergone multimodality treatment and are at risk for late effects; however, recognition of this risk may

not translate into clear screening recommendations: Cardiac toxicity is recognized as a concern for survivors of breast cancer [30], but ASCO guidelines for screening for cardiac late effects do not exist due to "the lack of direct, high-quality evidence on the benefits and harms of [this] screening" [31]. In the development of *OncoLife*, we described published, evidence-based guidelines whenever possible, and lacking those, consensus-based guidelines. In situations in which these types of published guidelines are not available, *OncoLife* information is provided to increase survivor and physician awareness of late effects and their possible treatments. Only a small fraction of *OncoLife* users (12%) reported ever having received survivorship information in the past. Certainly, our hope is that the information provided by *OncoLife*, whether evidence-, consensus-, or practice-based, will be useful to survivors, especially in a setting in which most report having had very little information offered to them. Future efforts will focus on increasing the individualization of *OncoLife* survivorship care plans, as well as understanding of the survivorship population in efforts to expand use and accessibility.

Acknowledgments

This project is supported in part by the Lance Armstrong Foundation. The authors would like to thank Lara N Sloboda, MS, for her consultation and advice.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The *OncoLink* homepage, with display of the *OncoLife* announcement.

[PDF file (Adobe PDF), 516 KB - jmir_v11i3e39_app1.pdf]

Multimedia Appendix 2

Sample *OncoLife* survivorship care plan, developed for a 69-year-old woman with history of breast cancer at age 61, treated with lumpectomy/sentinel lymph node biopsy, adriamycin and cytoxan chemotherapies, breast conserving radiotherapy, and tamoxifen followed by an aromatase inhibitor.

[PDF file (Adobe PDF), 114 KB - jmir_v11i3e39_app2.pdf]

References

1. ; Centers for Disease Control and Prevention (CDC). Cancer survivorship--United States, 1971-2001. MMWR Morb Mortal Wkly Rep 2004 Jun 25;53(24):526-529 [FREE Full text] [Medline: 15215740]
2. de Boer AGEM, Taskila T, Ojajärvi A, van Dijk FJH, Verbeek JHAM. Cancer survivors and unemployment: a meta-analysis and meta-regression. JAMA 2009 Feb 18;301(7):753-762 [FREE Full text] [Medline: 19224752] [doi: 10.1001/jama.2009.187]
3. Stanton AL. Psychosocial concerns and interventions for cancer survivors. J Clin Oncol 2006 Nov 10;24(32):5132-5137. [Medline: 17093275] [doi: 10.1200/JCO.2006.06.8775]
4. Short PF, Vargo MM. Responding to employment concerns of cancer survivors. J Clin Oncol 2006 Nov 10;24(32):5138-5141. [Medline: 17093276] [doi: 10.1200/JCO.2006.06.6316]
5. Earle CC, Neville BA. Under use of necessary care among cancer survivors. Cancer 2004 Oct 15;101(8):1712-1719 [FREE Full text] [Medline: 15386307] [doi: 10.1002/cncr.20560]
6. Keating NL, Landrum MB, Guadagnoli E, Winer EP, Ayanian JZ. Surveillance testing among survivors of early-stage breast cancer. J Clin Oncol 2007 Mar 20;25(9):1074-1081. [Medline: 17369571] [doi: 10.1200/JCO.2006.08.6876]
7. Snyder CF, Earle CC, Herbert RJ, Neville BA, Blackford AL, Frick KD. Preventive care for colorectal cancer survivors: a 5-year longitudinal study. J Clin Oncol 2008 Mar 1;26(7):1073-1079. [Medline: 18309941] [doi: 10.1200/JCO.2007.11.9859]
8. Hewitt M, Greenfield S, Stoval E. From cancer patient to cancer survivor—Lost in transition. Washington, DC: National Academies Press; 2005.
9. Nissen MJ, Beran MS, Lee MW, Mehta SR, Pine DA, Swenson KK. Views of primary care providers on follow-up care of cancer patients. Fam Med 2007;39(7):477-482 [FREE Full text] [Medline: 17602321]
10. Miedema B, MacDonald I, Tatemichi S. Cancer follow-up care. Patients' perspectives. Can Fam Physician 2003 Jul;49:890-895 [FREE Full text] [Medline: 12901486]
11. Fox S. Health information online. Washington, DC: Pew Internet & American Life Project; 2005. May 17 URL: http://www.pewinternet.org/~/media//Files/Reports/2005/PIP_Healthtopics_May05.pdf.pdf
12. Lorig KR, Ritter PL, Laurent DD, Plant K. Internet-based chronic disease self-management: a randomized trial. Med Care 2006 Nov;44(11):964-971. [Medline: 17063127] [doi: 10.1097/01.mlr.0000233678.80203.c1]

13. Lorig KR, Ritter PL, Laurent DD, Plant K. The internet-based arthritis self-management program: a one-year randomized trial for patients with arthritis or fibromyalgia. Arthritis Rheum 2008 Jul 15;59(7):1009-1017 [FREE Full text] [Medline: 18576310] [doi: 10.1002/art.23817]
14. Allen M, Iezzoni LI, Huang A, Huang L, Leveille SG. Internet-based coaching to improve patient-clinician communication in primary care. AMIA Annu Symp Proc 2007:861. [Medline: 18693963]
15. Allen M, Iezzoni LI, Huang A, Huang L, Leveille SG. Improving patient-clinician communication about chronic conditions: description of an internet-based nurse E-coach intervention. Nurs Res 2008;57(2):107-112. [Medline: 18347482] [doi: 10.1097/01.NNR.0000313478.47379.98]
16. Leveille SG, Huang A, Tsai SB, Weingart SN, Iezzoni LI. Screening for chronic conditions using a patient internet portal: recruitment for an internet-based primary care intervention. J Gen Intern Med 2008 Apr;23(4):472-475 [FREE Full text] [Medline: 18373147] [doi: 10.1007/s11606-007-0443-6]
17. Larkin M. Online support groups gaining credibility. Lancet 2000 May 20;355(9217):1834. [Medline: 10832868] [doi: 10.1016/S0140-6736(05)73097-X]
18. Cantrell MA, Conte T. Enhancing hope among early female survivors of childhood cancer via the internet: a feasibility study. Cancer Nurs 2008;31(5):370-379. [Medline: 18772662] [doi: 10.1097/01.NCC.0000305766.42475.33]
19. Ruland CM, Jeneson A, Andersen T, Andersen R, Slaughter L, Bente-Schjødt-Osmo, et al. Designing tailored Internet support to assist cancer patients in illness management. AMIA Annu Symp Proc 2007:635-639 [FREE Full text] [Medline: 18693913]
20. Metz JM. Abramson Cancer Center of the University of Pennsylvania. 2009. URL: http://www.oncolink.org/oncolife [accessed 2009 Aug 6] [WebCite Cache ID 5frBokU6H]
21. Metz JM. Abramson Cancer Center of the University of Pennsylvania. 2009. URL: http://www.oncolink.org/ [accessed 2009 Aug 6] [WebCite Cache ID 5frBGU1SE]
22. Reaman GH, Woelkers JF, Hallers SP. National Cancer Foundation, Children's Oncology Group. 2009. URL: http://www.curesearch.org/customize.aspx?stageid=19&lc=http://www.curesearch.org/for_patients/aftertreatment [accessed 2009 Aug 6] [WebCite Cache ID 5fvrJ12c7]
23. Kirkman TW. Tools for Science. 1996. URL: http://www.physics.csbsju.edu/stats [accessed 2007 Mar 26] [WebCite Cache ID 5gLIvjjL0]
24. National Cancer Institute. Office of Cancer Survivorship. About cancer survivorship research: survivorship definitions. URL: http://cancercontrol.cancer.gov/ocs/definitions.html [accessed 2006] [WebCite Cache ID 5fvo4ULEL]
25. Hewitt ME, Bamundo A, Day R, Harvey C. Perspectives on post-treatment cancer care: qualitative research with survivors, nurses, and physicians. J Clin Oncol 2007 Jun 1;25(16):2270-2273. [Medline: 17538172] [doi: 10.1200/JCO.2006.10.0826]
26. Metz JM, Devine P, DeNittis A, Jones H, Hampshire M, Goldwein J, et al. A multi-institutional study of Internet utilization by radiation oncology patients. Int J Radiat Oncol Biol Phys 2003 Jul 15;56(4):1201-1205. [Medline: 12829160] [doi: 10.1016/S0360-3016(03)00407-3]
27. Wilson JJ, Mick R, Wei SJ, Rustgi AK, Markowitz SD, Hampshire M, et al. Clinical trial resources on the internet must be designed to reach underrepresented minorities. Cancer J 2006;12(6):475-481. [Medline: 17207317] [doi: 10.1097/00130404-200611000-00007]
28. van de Poll-Franse LV, van Eenbergen MCHJ. Internet use by cancer survivors: current use and future wishes. Support Care Cancer 2008 Oct;16(10):1189-1195. [Medline: 18293014] [doi: 10.1007/s00520-008-0419-z]
29. Fogel J, Ribisl KM, Morgan PD, Humphreys K, Lyons EJ. Underrepresentation of African Americans in online cancer support groups. J Natl Med Assoc 2008 Jun;100(6):705-712. [Medline: 18595573]
30. Harris EER, Correa C, Hwang WT, Liao J, Litt HI, Ferrari VA, et al. Late cardiac mortality and morbidity in early-stage breast cancer patients after breast-conservation treatment. J Clin Oncol 2006 Sep 1;24(25):4100-4106. [Medline: 16908933] [doi: 10.1200/JCO.2005.05.1037]
31. Carver JR, Shapiro CL, Ng A. American Society of Clinical Oncology clinical evidence review on the ongoing care of adult cancer survivors: cardiac and pulmonary late effects. J Clin Oncol 2007;25(25):3991-4008. [Medline: 17577017] [doi: 10.1200/JCO.2007.10.9777]

Abbreviations

ASCO: American Society of Clinical Oncology
COG: Children's Oncology Group
IOM: Institute of Medicine
IP: Internet protocol
IRB: Institutional Review Board
NCI: National Cancer Institute
PCP: primary care provider
US: United States

Edited by G Eysenbach; submitted 02.03.09; peer-reviewed by H Jones, R Smith, S Urowitz; comments to author 23.03.09; revised version received 07.05.09; accepted 17.06.09; published 04.09.09.

Please cite as:
Hill-Kayser CE, Vachani C, Hampshire MK, Jacobs LA, Metz JM
An Internet Tool for Creation of Cancer Survivorship Care Plans for Survivors and Health Care Providers: Design, Implementation, Use and User Satisfaction
J Med Internet Res 2009;11(3):e39
URL: http://www.jmir.org/2009/3/e39/
doi:10.2196/jmir.1223
PMID:19734128

Original Paper

Effect of Home-Based Telemonitoring Using Mobile Phone Technology on the Outcome of Heart Failure Patients After an Episode of Acute Decompensation: Randomized Controlled Trial

Daniel Scherr[1], MD; Peter Kastner[2], MSc, MBA; Alexander Kollmann[2], PhD; Andreas Hallas[3], MD; Johann Auer[4], MD; Heinz Krappinger[5], MD; Herwig Schuchlenz[6], MD; Gerhard Stark[7], MD; Wilhelm Grander[8], MD; Gabriele Jakl[9], MD; Guenter Schreier[2], PhD; Friedrich M Fruhwald[1], MD, FESC; and the MOBITEL investigators[10]

[1]Department of Cardiology, Medical University Graz, Graz, Austria

[2]Information Management & eHealth, AIT Austrian Institute of Technology GmbH, Graz, Austria

[3]Department of Internal Medicine, Hospital Tulln, Tulln, Austria

[4]Department of Cardiology, Hospital Wels, Wels, Austria

[5]Department of Cardiology, Hospital Villach, Villach, Austria

[6]Department of Cardiology, Hospital Graz West, Graz, Austria

[7]Department of Internal Medicine, Hospital Deutschlandsberg, Deutschlandsberg, Austria

[8]Department of Internal Medicine, Hospital Hall/Tirol, Hall/Tirol, Austria

[9]Department of Cardiology, Wilhelminenspital Vienna, Vienna, Austria

[10]For a list of additional MOBITEL investigators see Acknowledgments

Corresponding Author:
Friedrich M Fruhwald, MD, FESC
Department of Cardiology
University Clinic of Internal Medicine
Auenbruggerplatz 15
A-8036 Graz
Austria
Phone: +43 316 385 2544
Fax: +43 316 385 3733
Email: friedrich.fruhwald@medunigraz.at

Abstract

Background: Telemonitoring of patients with chronic heart failure (CHF) is an emerging concept to detect early warning signs of impending acute decompensation in order to prevent hospitalization.

Objective: The goal of the MOBIle TELemonitoring in Heart Failure Patients Study (MOBITEL) was to evaluate the impact of home-based telemonitoring using Internet and mobile phone technology on the outcome of heart failure patients after an episode of acute decompensation.

Methods: Patients were randomly allocated to pharmacological treatment (control group) or to pharmacological treatment with telemedical surveillance for 6 months (tele group). Patients randomized into the tele group were equipped with mobile phone–based patient terminals for data acquisition and data transmission to the monitoring center. Study physicians had continuous access to the data via a secure Web portal. If transmitted values went outside individually adjustable borders, study physicians were sent an email alert. Primary endpoint was hospitalization for worsening CHF or death from cardiovascular cause.

Results: The study was stopped after randomization of 120 patients (85 male, 35 female); median age was 66 years (IQR 62-72). The control group comprised 54 patients (39 male, 15 female) with a median age of 67 years (IQR 61-72), and the tele group included 54 patients (40 male, 14 female) with a median age of 65 years (IQR 62-72). There was no significant difference between groups with regard to baseline characteristics. Twelve tele group patients were unable to begin data transmission due to the inability of these patients to properly operate the mobile phone (“never beginners”). Four patients did not finish the study due to personal reasons. Intention-to-treat analysis at study end indicated that 18 control group patients (33%) reached the primary endpoint (1 death, 17 hospitalizations), compared with 11 tele group patients (17%, 0 deaths, 11 hospitalizations; relative risk reduction 50%, 95% CI 3-74%, $P = .06$). Per-protocol analysis revealed that 15% of tele group patients (0 deaths, 8 hospitalizations)

reached the primary endpoint (relative risk reduction 54%, 95% CI 7-79%, *P*= .04). NYHA class improved by one class in tele group patients only (*P*< .001). Tele group patients who were hospitalized for worsening heart failure during the study had a significantly shorter length of stay (median 6.5 days, IQR 5.5-8.3) compared with control group patients (median 10.0 days, IQR 7.0-13.0; *P*= .04). The event rate of never beginners was not higher than the event rate of control group patients.

Conclusions: Telemonitoring using mobile phones as patient terminals has the potential to reduce frequency and duration of heart failure hospitalizations. Providing elderly patients with an adequate user interface for daily data acquisition remains a challenging component of such a concept.

(J Med Internet Res 2009;11(3):e34) doi:10.2196/jmir.1252

KEYWORDS

Heart failure; telemedicine; mobile phone; eHealth

Introduction

Chronic heart failure (CHF) is a major cause of hospitalization in Western countries, with a high rate of re-admission [1]. Despite effective therapies, patients with CHF suffer from a high rate of hospital re-admission and bear a high risk of death within the first months after an episode of acute heart failure [1,2].

Reduction of re-admissions is a major goal in the management of CHF patients. Therefore, both the European and the American guidelines on CHF recommend disease management programs that include telephone follow-up or home nursing [3,4].

Monitoring body weight, home visits by nurses, telephone conferences with nurses, and complete telemedical supervision have been tested with inconsistent results [5-20]. Simple interventions such as measuring body weight with telemedical surveillance produced conflicting results for re-hospitalization [5,8]. Nurse visits at home or telephone contact with nurses [6-7,10-12,14] inconsistently showed a benefit for the prognosis of heart failure patients.

Telemedicine using transmission of vital parameters can be expensive and technically demanding [12,14,21]. The majority of reports show a positive influence on the outcome of heart failure patients [9,12,15], although it is not clear which system is best for this challenge.

Rapid advances and the ubiquitous availability of mobile phones have created new perspectives toward telemedical interaction between patients and health care professionals [22]. Recently, we were able to show that mobile phone–based surveillance is safe and promising in cardiac patients [23].

Based on these promising early results using mobile phone technology, we designed the MOBITEL (MOBIle TELemonitoring in heart failure patients) study in order to test the hypothesis that telemedical surveillance using widespread mobile phone technology as patient terminals improves the outcome of CHF patients after an episode of acute cardiac decompensation.

Methods

Patients

MOBITEL was a prospective, randomized, open-label study. Recruitment started on October 1, 2003, and the study was closed with the sign-off of the last patient on April 29, 2008. Registration in a public trials registry was not performed for MOBITEL since it was optional at the time of study start. Local regulatory authorities approved the study, with the University Clinic Graz being the lead ethics committee.

Patients were eligible for the study if they met all of the following inclusion criteria: acute worsening of heart failure (acute cardiac decompensation) with hospital admission lasting > 24 hours within the last 4 weeks, treatment according to the guidelines of the European Society of Cardiology (ESC) with an angiotensin converting enzyme (ACE) inhibitor or an angiotensin receptor blocker (ARB), diuretic, and beta-blocker (except in cases with documented intolerance to beta-blockers). Initially, patients older than 18 years and younger than 75 years were eligible; the latter was amended to 80 years after 4 months of recruitment. For the definition of CHF, we adopted the ESC guidelines [4].

Patients with one of the following conditions were not eligible for MOBITEL: unstable coronary artery disease with revascularization within the last 6 months, planned revascularization (percutaneous or surgical) for coronary artery disease, planned heart valve surgery, planned or completed heart transplantation, uncontrolled arterial hypertension, acute myocarditis, inability to read the display of a handheld phone, or malignancy.

After receiving verbal and written study information, patients gave written consent to participate in the study. Patients were allocated randomly to pharmacological treatment (control group) or pharmacological treatment plus telemedical surveillance (tele group). The adaptive randomization procedure was stratified by patient age, New York Heart Association (NYHA) class, gender, and study center.

Baseline demographics and medication were recorded for all patients, and an appointment for the 6-month follow-up was made. There was no planned interaction between study site and patients in the control group within the follow-up period of 6 months. Patients in the tele group were given the telemonitoring equipment and an appointment for telephone or face-to-face technical training.

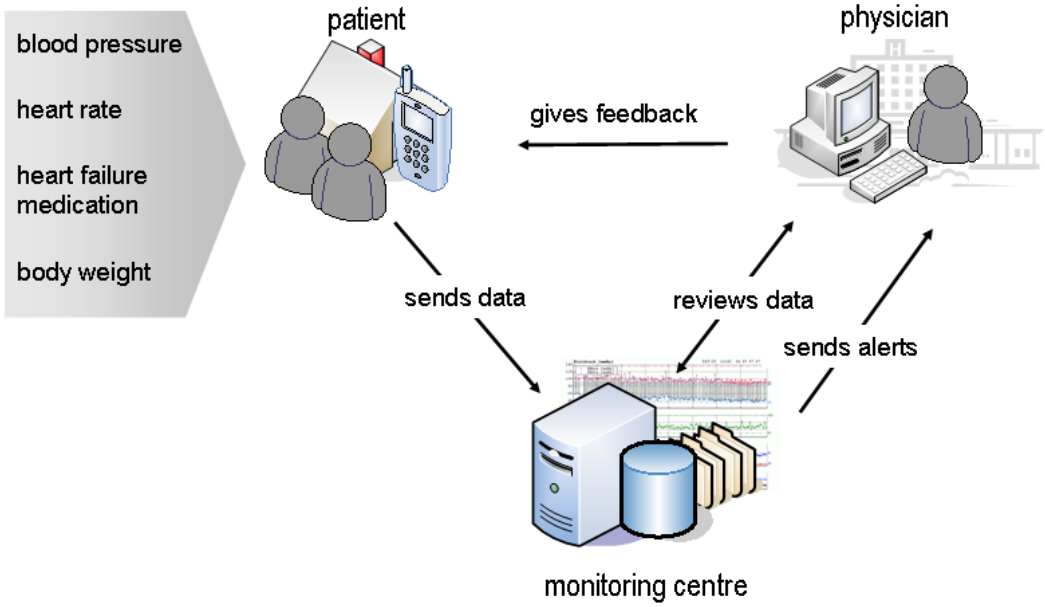
Equipment and Data Processing

The telemonitoring equipment consisted of three commercially available components: (1) a mobile phone (Nokia 3510, Finland), (2) a weight scale with 0.1 kg accuracy and electronic

display (Soehnle creta, Germany), and (3) a sphygmomanometer for fully automated measurement of blood pressure and heart rate (BosoMedicus, Bosch&Sohn, Germany). Tele group patients were trained in measurement of blood pressure and weight using the equipment prior to discharge home. Furthermore, tele group patients were instructed by a study technician in the use of the mobile phone.

Tele group patients were asked to measure vital parameters (blood pressure, heart rate, body weight) on a daily basis at the same time, preferably in the morning after emptying the bladder and before dressing and taking medication. Thereafter, patients were advised to enter these values as well as their dosage of heart failure medication into the mobile phone's Internet browser and send them to the monitoring center provided by the Austrian Institute of Technology (AIT) – Information Management & eHealth, Graz. Study physicians had access to a secure website providing both numerical and graphical depiction of data for each patient. Whenever necessary, study physicians could contact patients using the mobile phone. Figure 1 outlines this process.

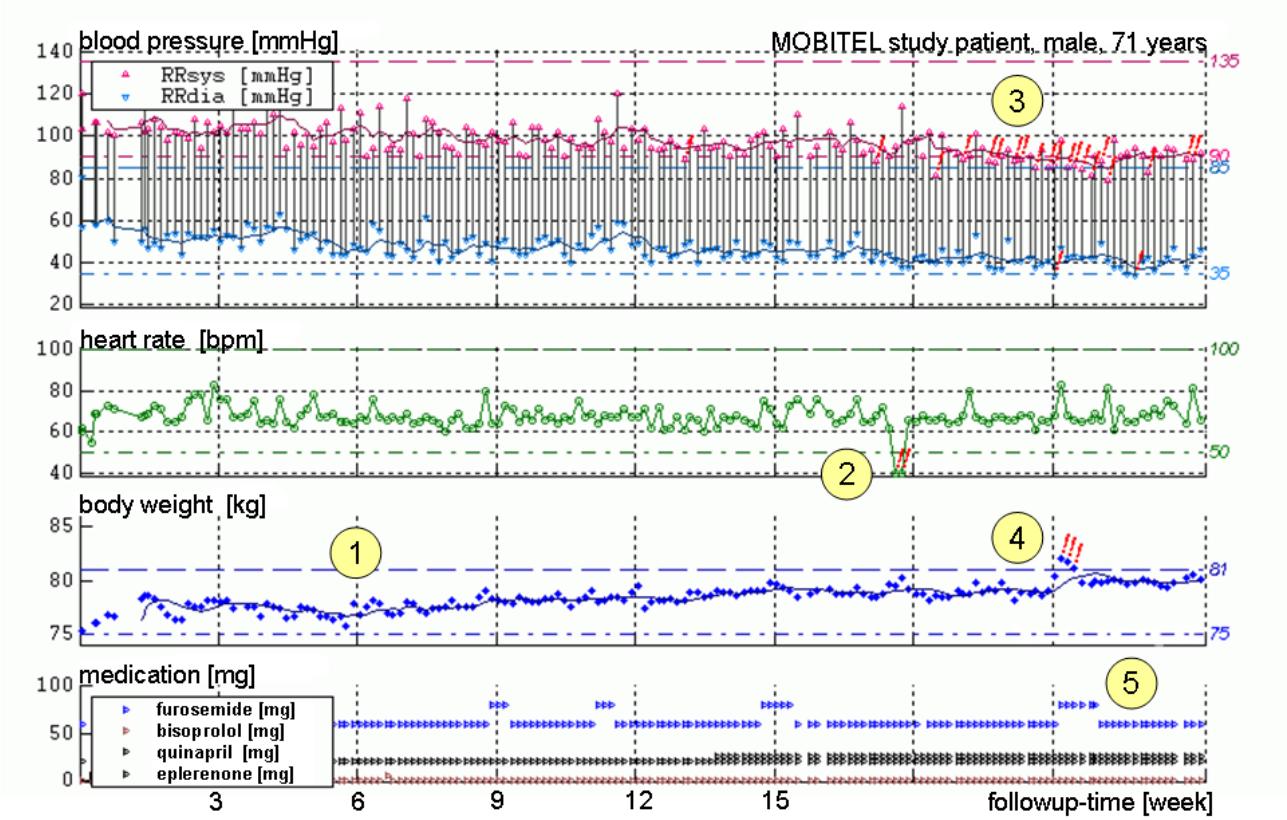
Figure 1. Schematic depiction of the equipment and data collection process used in MOBITEL



At the monitoring center, data were depicted both numerically and graphically in an electronic case record form (CRF; Figure 2). Study physicians had continuous access to the CRFs of their patients via a secure website. Physicians were advised to use the automated warning system for the monitoring of vital parameters of their patients. If transmitted values went outside individually adjustable borders, study physicians were sent an email alert. Additionally, an email alert was generated if a patient's body weight increased or decreased more than 2 kg in 2 days. After receiving an alert, study physicians could contact the patient directly via the mobile phone to confirm the parameters and, if appropriate, could ask the patient to adjust his or her medication.

For technical questions, patients had access to a 24-hour hotline at the service center.

Figure 2. Trend chart of vital parameters of a typical patient (male, 71 years) over the 6-month study period: Each dot represents a transmitted value of (from top to bottom) systolic (red) and diastolic blood pressure (light blue), heart rate (green), body weight (blue), and heart failure medication (bottom panel, furosemide in blue); (1) After discharge from hospital, all transmitted values are stable, (2) Red exclamation marks indicate low heart rate, (3) Alarms due to decreased blood pressure herald a phase of instability, (4) Increase in body weight of more than 2 kg in 2 days, causing an alarm (red exclamation marks), (5) After telephone contact with the patient and confirmation of values, the daily dose of furosemide is temporarily increased and body weight returns to stable values within 3 days



MOBITEL Telemedicine Platform

The MOBITEL telemedicine platform was developed as a three-tier, client-server architecture (data, logic, and representation layers) using state-of-the-art Internet technology. The Zope Web/application server (Zope 2.6.1, Zope Corporation, Fredericksburg, VA, USA) and the relational database system (Interbase 6.0, Borland Software Corporation, Cupertino, CA, USA) were chosen for the basic system.

While the application server provided core logic, particular services were developed as independent modules in the sense of service-oriented architecture (SOA). Particular functionalities were clustered into services that were able to communicate and share data with each other.

- data processing and graphic service: For sophisticated data processing and visualization of time-series data (e.g. blood pressure measurements), the MatLab 6.5 environment (The MathWorks, Natick, MA) was used.
- notification service: The database was checked at regular intervals for arrival of new alerts, notifications, or reminders generated by the data processing service. Subsequently, a personalized message was composed and sent to the responsible physician by text messaging, email, or both.

The components of the MOBITEL telemedicine platform were designed with respect to a high level of security and confidentiality to comply with regulatory requirements. Data transfer was encrypted, and access to the data was restricted to authorized users.

Outcome Parameters and Endpoint Definitions

The combined primary endpoint of this study was cardiovascular mortality or re-hospitalization for worsening heart failure. Besides evaluation of patients' functional status according to the NYHA classification and length of stay during re-hospitalizations, further secondary endpoints focused on technical parameters: system availability, cumulative transmissions, and transmissions per patient.

Study patients were classified as reaching the primary endpoint in case of hospital admission for worsening heart failure or cardiovascular death within the study period.

Tele group patients who dropped out immediately after randomization due to inability to handle the telemonitoring equipment despite intensified training were classified as "never

beginners." These patients were contacted 6 months after randomization to obtain data for the primary endpoint.

Study patients who stopped prematurely for personal reasons were classified as "early termination on patient's request." These patients were invited to an earlier follow-up (i.e. end-of-study visit) to obtain as much information as possible and were included in the final analysis according to randomization.

Statistics

For statistical planning, we assumed that patients in the control arm would show an event rate of 30% over 6 months [1]. For the telemedicine arm, we expected a 50% reduction of the event rate. To show a statistically significant difference at an error of .05 with a power of 80%, a sample size of 240 subjects was calculated.

Randomization was stopped after 120 patients due to an increasing number of never beginners who were unable to operate the mobile phone, indicating the urgent need for a new technology. However, as we tried to avoid a mix of technologies within one study, we decided to stop randomization in coordination with the ethics committee of the Medical University Graz. Therefore, the results must be interpreted cautiously.

Final data analysis was performed according to the per-protocol principle, including all patients except never beginners. Additionally, intention-to-treat analysis was performed, including all randomized patients. The log-rank test, the Kaplan-Meier estimation method with 95% confidence interval (CI), and calculation of the relative risk reduction were used to analyze the primary endpoint.

With respect to the secondary endpoints, difference in functional NYHA class between baseline and end of study was compared between patients in the control and tele group. Dosage of heart failure medication was calculated as percentage of the ESC-recommended dosage and compared between both groups.

Normally distributed values were compared using the t-test, while the chi-square test was used to compare nominal distributed values. We used the Wilcoxon rank sum test to compare independent samples and the Wilcoxon signed rank test to compare dependent samples when values were not normally distributed.

Patients' adherence with the telemonitoring system was expressed as a percentage of effectively to expected received datasets. Thus, the cumulative monitoring period and the total number of received values were calculated for all participants.

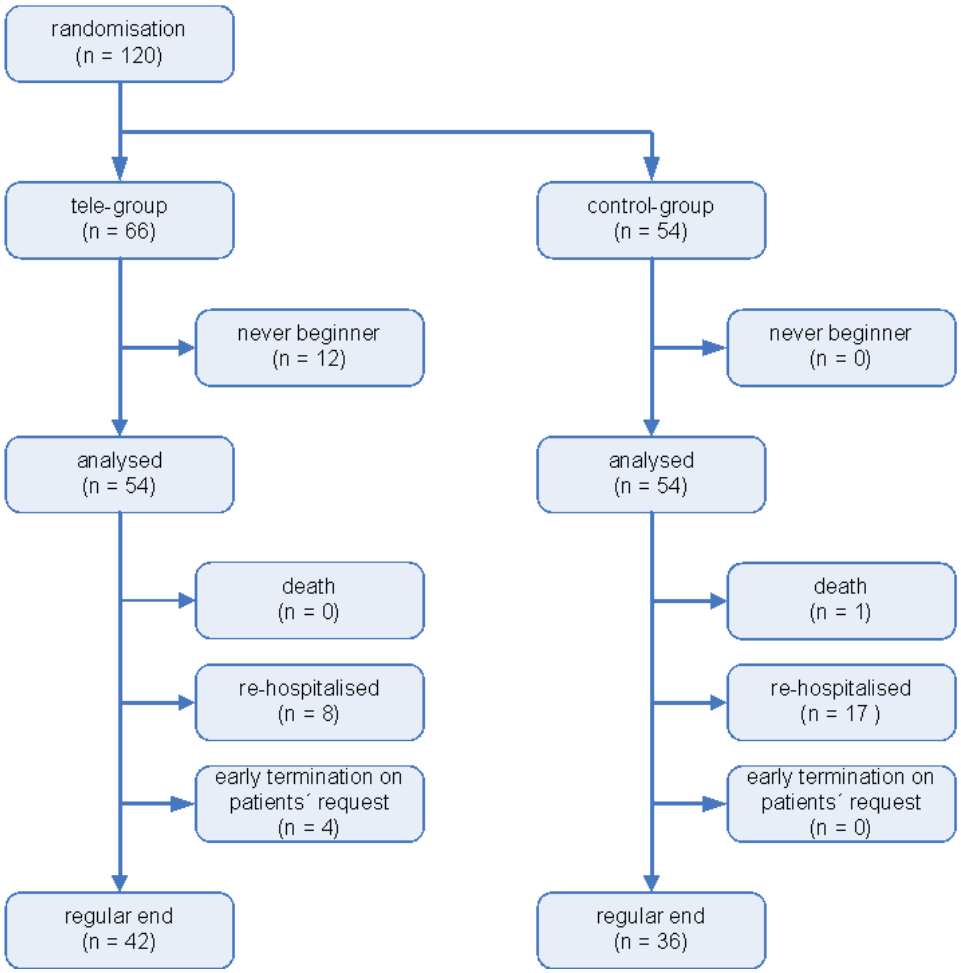
Statistics were calculated using R statistical software, version 2.4.1 (R Foundation for Statistical Computing, Vienna, Austria). We considered a P-value < .05 to show a statistically significant difference in all comparisons. All data are given as median and interquartile range (IQR).

Results

Patients

In the tele group, 12 patients (6 male, 6 female; median age 68 years [IQR 64-74]) emerged unable to begin transmission of data and were classified as never beginners. Never beginners were included in the intention-to-treat analysis but not the per-protocol analysis. Furthermore, there were four patients who requested early termination of the study (4 male, median age 63 years [IQR 60-65]) and were included in both the intention-to-treat analysis and the per-protocol analysis. Figure 3 shows the study flowchart.

Figure 3. Study flowchart of the MOBITEL population



Overall, 120 patients (85 male, 35 female) with a median age of 66 years (IQR 62-72) were randomized in eight centers: 54 patients were randomized to the control group (39 male, 15 female; median age 67 years [IQR 61-72]), and 66 patients were randomized to the tele group (46 male, 20 female; median age 66 years [IQR 62-73]).

For the per-protocol analysis (Table 1), the control group comprised 54 patients (39 male, 15 female; median age 67 years [IQR 61-72]) and was compared with the tele group, including 54 patients (40 male, 14 female; median age 65 years [IQR 62-72]). At baseline there was no statistically significant difference between the two groups regarding age, gender, cause of heart failure, NYHA functional class, left ventricular (LV) ejection fraction, heart failure medication, as well as frequency and duration of heart failure hospitalizations in the year prior to randomization (Table 1).

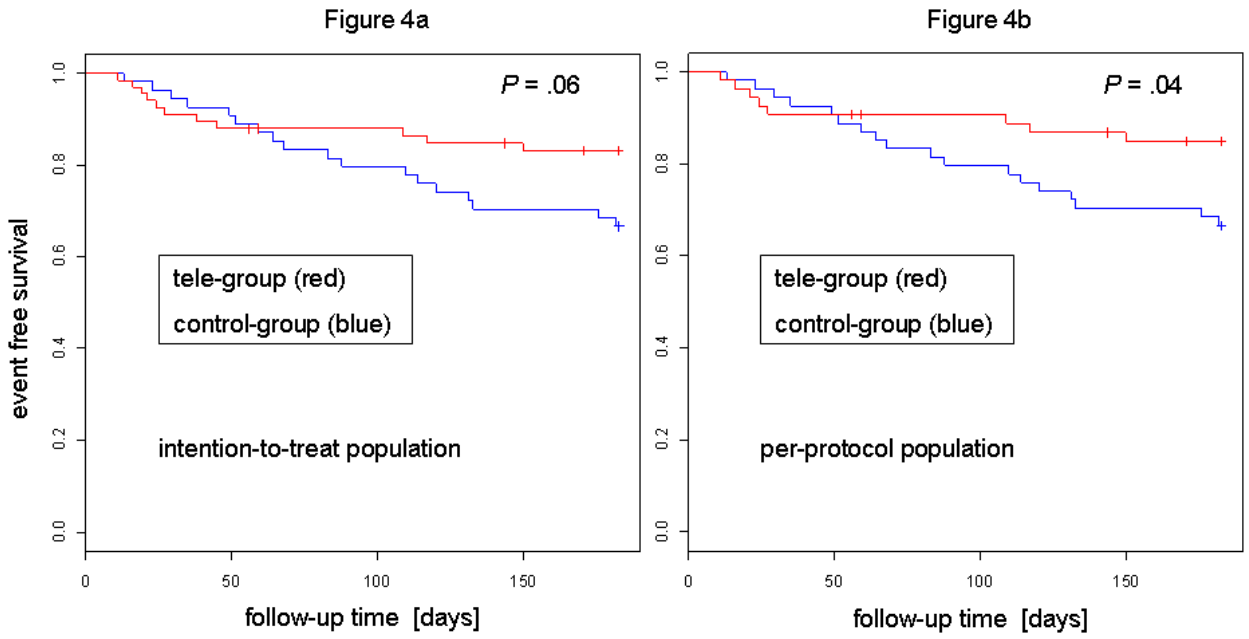
Table 1. Baseline characteristics of the MOBITEL population (per-protocol population)

Characteristic	Control Group (n = 54)	Tele Group (n = 54)	*P* (for difference)
Median age, years (IQR)	67 (61-72)	65 (62-72)	.79
Male, no. (%)	39 (72)	40 (74)	.92
Median LV ejection fraction (IQR)	29 (21-36)	25 (20-38)	.70
NYHA class II, no. (%)	7 (13)	7 (13)	.40
NYHA class III, no. (%)	37 (68.5)	33 (61)	.40
NYHA class IV, no. (%)	10 (18.5)	14 (26)	.40
Median number of HF[a] hospitalizations in past 12 months, no. (IQR)	1 (1-2)	1 (1-2)	.36
Median length of stay for HF hospitalizations, days (IQR)	11 (7-17)	12 (9-15)	.67
Cardiovascular Risk Profile			
Ischemic heart disease, no. (%)	23 (43)	20 (37)	.69
Hypertension, no. (%)	24 (44)	29 (54)	.44
Valvular disease, no. (%)	1 (2)	1 (2)	.48
Diabetes mellitus, no. (%)	16 (30)	12 (22)	.44
HF Treatment at Study Entry			
ACE inhibitor, no. (%)	41 (76)	45 (83)	.28
ARB, no. (%)	13 (24)	9 (17)	.47
Diuretic, no. (%)	44 (81)	49 (91)	.27
Beta-blocker, no. (%)	42 (78)	47 (87)	.31
Spironolactone, no. (%)	23 (43)	21 (39)	.85

[a]HF = heart failure.

Intention-to-treat analysis indicated that 18 control group patients (33%) reached the primary endpoint (1 death, 17 hospitalizations) compared with 11 tele group patients (17%, 0 deaths, 11 hospitalizations; relative risk reduction 50%, 95% CI 3-74%, *P*= .06; Figure 4a). The number of never beginners reaching the primary endpoint was not higher than for control group patients.

Figure 4. Kaplan-Meier curve for primary endpoint in MOBITEL: intention-to-treat analysis (4a) and per-protocol analysis (4b)



Per-Protocol Analyses

Per-protocol analysis at study end revealed that 15% of tele group patients (0 deaths, 8 hospitalizations) reached the primary endpoint (relative risk reduction 54%, 95% CI 7-79%, *P*= .04; Figure 4b).

Median NYHA class improved from 3 to 2 in tele group patients only (*P*< .001) compared with control group patients at study end and compared with tele group baseline values. Ejection fraction showed a nonsignificant improvement in both the control group, from 29% (IQR 21-36) to 35% (IQR 24-40), and the tele group, from 25% (IQR 20-38) to 35% (IQR 25-45).

Tele group patients who were hospitalized for worsening heart failure during the study had a significantly shorter length of hospital stay (median 6.5 days, IQR 5.5-8.3) compared with control group patients (median 10.0 days, IQR 7.0-13.0; *P*= .04).

Dosage of heart failure medication was calculated as percent of the ESC-recommended daily dose. At baseline, tele group patients had a higher percentage of the ESC-recommended daily dose of ACE inhibitors compared to control group patients. However, this difference was no longer significant at study end. All other types of drugs were equally balanced.

Tele group patients started transmission of data within 6.5 days (IQR 4-11) after randomization. During the entire study, a total of 7554 value sets were received from tele group patients, corresponding to a median of 162 transmissions per patient (IQR 136-173). On 7554 out of 7962 cumulative monitoring days, at least one set of values was sent, which indicates a patient adherence rate of 95%.

There were 375 alerts sent to study physicians in cases of exceeding predefined thresholds for body weight or exceeding the dynamic threshold of ± 2 kg in 2 days. Consequently, tele group patients were contacted 170 times. In 55 of those times, an adjustment of heart failure medication was made (Table 2). Adverse events or side effects were not observed.

Table 2. Technical data and interventions in the telemedicine arm

Variable	Tele Group (n = 54)
System availability, %	98
Patient adherence, %	95
Cumulative monitoring, days	7962
Cumulative transmissions, no.	7554
Median transmissions/patient, no. (IQR)	162 (136-173)
Patient contacts following alarms, no.	170
Adjustments of individual limits, no.	118
Adjustments of medication, no.	55

Discussion

The MOBITEL study addressed important issues in the management of heart failure patients after discharge from hospital, namely the following:

- how to reduce the high risk of re-hospitalization for worsening heart failure and/or cardiovascular death
- how to detect early warning signs of impending decompensation
- how to inform the treating physician

For this purpose, mobile phone–based patients terminals were used in a telemonitoring scenario. The system was developed and tested in 120 heart failure patients. After 6 months, there was a reduction of relative risk in tele group patients (re-admission or death) by 54%. However, the small sample size and the premature termination of randomization because of relevant technological issues with the patient terminal have to be considered as the main limitations when interpreting the results.

Telemonitoring and Telemedical Intervention for Patients With Heart Failure

Patients suffering an episode of acute heart failure are at a high risk of adverse events within the first months after such an episode [1,2]. Therefore, efforts are needed to reduce this risk and to look out for early warning signs of acute heart failure.

Telemonitoring in general may provide a powerful tool to look for early warning signs and to reduce the high risk of adverse events in patients with CHF [19]. However, there are some uncertainties about the parameters to monitor as well as the proper monitoring tool.

The simple concept of monitoring body weight with telemedical surveillance did not show consistent results [5,8] but led to more sophisticated monitoring tools. However, recently it has been shown that the risk for re-hospitalization for worsening heart failure is highest in patients with weight gain of more than 4.5 kg within the week prior to hospitalization [24]. In our study, we also used increasing body weight as an early warning sign of fluid retention, although our limits were much closer (2 kg weight gain within 2 days). This stringent limit led to a significant number of alerts prompting communication between study physicians and patients, and, in some but not all cases, led to an adjustment of medication.

Increase in body weight often signals worsening heart failure. In addition, changes in blood pressure and heart rate may also herald worsening heart failure. To detect such episodes, we used a fully automated sphygmomanometer to measure both blood pressure and heart rate to gain information about potential refinement of heart failure therapy.

Our study is also the first to investigate the ability of patients to transmit information on heart failure medication dose on a daily basis. The vast majority of patients were able to enter and transmit their daily doses together with vital parameters by using the mobile Internet browser.

Furthermore, this study offered the unique ability to see that patients with CHF underwent an educational process: In many tele group patients who entered the study with some instability and who were (in response to alarms) advised by the study physician to adapt diuretic therapy, we saw, over time, some form of “self-monitoring” and, consequently, self-treatment.

As indicated by the late separation of Kaplan-Meier curves in Figure 4, telemedical surveillance does not appear to be superior to conventional treatment in the first month of follow-up. In our study, the majority of re-hospitalizations of tele group patients was seen in the first month of patient follow-up.

Alarms and Alarm Management

In this study, sophisticated pre-analysis of data was not implemented. False-positive alarm messages sent to the physicians were predominately influenced by two factors:

- Outlier and type errors were not identified.
- Individual thresholds were not quickly adjusted in response to the slight variation of the body weight over time.

However, alarm management in general (i.e. timely confirmation of an alarm) and alarm generation in particular turned out to be crucial factors in a home-monitoring scenario. Hence, future research will focus on the development of sophisticated algorithms for automatic data analysis of home-monitoring data. The algorithm should provide high sensitivity and specificity in order to make telemedical services effective and efficient.

Mobile Phone–Based Patient Terminal

MOBITEL was the first study showing a positive influence on outcome in CHF patients using mobile phones as patient terminals for daily data transmission to the monitoring center.

Mobile phone technology (either GSM or 3G) is widespread in Europe and is a familiar tool, even in elderly patients. Hence, after 4 months of recruitment, we amended the upper age limit for inclusion up to 80 years.

The approach of using the mobile phone as telemonitoring equipment is different from all other studies reported so far: previous trials used more sophisticated equipment to obtain and transmit vital parameters from heart failure patients [9,12,15,21].

As the 95% patient adherence to the telemonitoring system indicates, the system was well accepted by those who were able to handle the mobile phone. Contrary to our expectations, data entry errors were rare, and the quality of self-reported data was appropriate for further clinical evaluation.

Recently, our group reported the first use of mobile phone–based telemedical surveillance in cardiac patients [23]. With this trial we were able to show that this technology can be used to monitor heart failure patients and to improve their outcome.

Limitations

Not surprisingly, our study has a number of limitations. Although we were careful not to include patients with visual impairment (this was an explicit exclusion criterion), there were some patients who were not able to properly handle the mobile phone's Internet browser (never beginners). However, there was no difference at baseline between these never beginners and the remaining study patients, and the never beginners subsequently had an event rate that was not different from control patients.

As a consequence, we looked for new technologies to improve the usability for elderly, technically unskilled patients. A promising solution could be the use of near field communication (NFC) in telemonitoring technology. Unlike Bluetooth, NFC supports a touch-based method for data acquisition using upcoming NFC-enabled mobile phones [24]. In addition to data acquisition from medical devices, it provides access to data stored on radio frequency identification (RFID) tags (eg, electronic barcodes), which could be embedded in future telemonitoring technology. However, as we did not want to mix two completely different technologies in one study, we decided to stop randomization after 120 patients. Therefore, the intention-to-treat analysis only revealed a trend in favor of telemedical surveillance, while the per-protocol analysis showed a significant difference between groups. This is the second major limitation of MOBITEL.

Lessons Learned

Based on the experiences from the MOBITEL study, the following key requirements for the utilization of a telemonitoring system in CHF management have been identified:

- an easy-to-use patient terminal to allow safe and secure data acquisition, especially for unskilled, elderly patients
- sophisticated alarm management with high sensitivity and specificity in order to focus physicians' attention on those patients whose data indicate signs of worsening health status

The development of an integrated care concept might be helpful to allow optimal integration of telemonitoring tasks into the existing workflows and processes of clinicians, general practitioners, and home care nurses.

Conclusion

The results of the MOBITEL study indicate that home-based telemonitoring using mobile phones improves outcome in CHF patients and reduces both frequency and duration of heart failure hospitalizations. Providing elderly patients with an adequate user interface for data acquisition and transmission remains a challenging part of this concept.

Acknowledgments

This study was partly funded by restricted research grants from Novartis Pharma Austria, Roche Pharma Austria, and Mobilkom Austria.

Additional MOBITEL Coinvestigators are as follows:

Department of Cardiology, Medical University Graz, Austria: Christoph Castellani, Stefan Prisker, Melanie Reicher, Michael Scheruebl, Hannah Walk

Safety & Security Department, AIT Austrian Institute of Technology GmbH, Graz, Austria: Juergen Morak

Department of Internal Medicine, Hospital Tulln, Austria: Bruno Brunner

Department of Cardiology, Wels, Austria: Elisabeth Lassnig

Department of Cardiology, General Hospital Graz West, Austria: Georg Saurer

Department of Internal Medicine, General Hospital Deutschlandsberg, Austria: Sonja Schroettner

Department of Cardiology, Wilhelminenspital Vienna, Austria: Elisabeth Weiss

Conflicts of Interest

None declared.

References

1. Zannad F, Briancon S, Juilliere Y, Mertes PM, Villemot JP, Alla F, et al. Incidence, Clinical and Etiologic Features, and Outcomes of Advanced Chronic Heart Failure: The EPICAL Study. J Am Coll Cardiol 1999 Mar;33(3):734-742. [Medline: 10080475] [doi: 10.1016/S0735-1097(98)00634-2]
2. Solomon SD, Dobson J, Pocock S, Skali H, McMurray JJV, Granger CB, et al. Influence of nonfatal hospitalization for heart failure on subsequent mortality in patients with chronic heart failure. Circulation 2007 Sep 25;116(13):1482-1487 [FREE Full text] [Medline: 17724259] [doi: 10.1161/CIRCULATIONAHA.107.696906]
3. ; Writing Committee to Update the 2001 Guidelines for the Evaluation and Management of Chronic Heart Failure. ACC/AHA 2005 guidelines for the diagnosis and management of chronic heart failure in the adult. J Am Coll Cardiol 2005;46(6):1144-1178. [Medline: 16168305] [doi: 10.1016/j.jacc.2005.07.012]
4. ; The Task Force for the Diagnosis and Treatment of Chronic Heart Failure of the ESC. Guidelines for the diagnosis and treatment of chronic heart failure: executive summary (2005). Eur Heart J 2005;26(11):1115-1140. [Medline: 15901669] [doi: 10.1093/eurheartj/ehi204]
5. Goldberg LR, Piette JD, Walsh MN, Frank TA, Jaski BE, Smith AL, et al. Randomized trial of a daily electronic home monitoring system in patients with advanced heart failure: The Weight Monitoring in Heart Failure (WHARF) trial. Am Heart J 2003 Oct;146(4):705-712. [Medline: 14564327] [doi: 10.1016/S0002-8703(03)00393-4]
6. Inglis SC, Pearson S, Treen S, Gallasch T, Horowitz JD, Stewart S. Extending the horizon in chronic heart failure: Effects of multidisciplinary, home-based intervention relative to usual care. Circulation 2006 Dec 5;114(23):2466-2473 [FREE Full text] [Medline: 17116767] [doi: 10.1161/CIRCULATIONAHA.106.638122]
7. Jaarsma T, van der Wal MHL, Lesman-Leegte I, Luttik ML, Hogenhuis J, Veeger NJ, et al. Effect of moderate or intensive disease management program on outcome in patients with heart failure: Coordinating Study Evaluating Outcomes of Advising and Counseling in Heart Failure (COACH). Arch Intern Med 2008 Feb 11;168(3):316-324 [FREE Full text] [Medline: 18268174] [doi: 10.1001/archinternmed.2007.83]
8. Cordisco ME, Benjaminovitz A, Hammond K, Mancini D. Use of telemonitoring to decrease the rate of hospitalization in patients with severe congestive heart failure. Am J Cardiol 1999 Oct 1;84(7):860-862. [Medline: 10513789] [doi: 10.1016/S0002-9149(99)00452-X]
9. de Lusignan S, Wells S, Johnson P, Meredith K, Leatham E. Compliance and effectiveness of 1 year's home telemonitoring. The report of a pilot study of patients with chronic heart failure. Eur J Heart Fail 2001 Dec;3(6):723-730. [Medline: 11738225] [doi: 10.1016/S1388-9842(01)00190-8]
10. Scalvini S, Zanelli E, Volterrani M, Martinelli G, Baratti D, Buscaya O, et al. A pilot study of nurse-led, home-based telecardiology for patients with chronic heart failure. J Telemed Telecare 2004;10(2):113-117. [Medline: 15068649] [doi: 10.1258/135763304773391576]
11. ; GESICA Investigators. Randomised trial of telephone intervention in chronic heart failure: DIAL trial. BMJ 2005 Aug 20;331(7514):425 [FREE Full text] [Medline: 16061499] [doi: 10.1136/bmj.38516.398067.E0]
12. Cleland JGF, Louis AA, Rigby AS, Janssens U, Balk AHMM. Noninvasive home telemonitoring for patients with heart failure at high risk of recurrent admission and death: the Trans-European Network-Home-Care Management System (TEN-HMS) study. J Am Coll Cardiol 2005 May 17;45(10):1654-1664. [Medline: 15893183] [doi: 10.1016/j.jacc.2005.01.050]
13. Clark RA, Yallop JJ, Piterman L, Croucher J, Tonkin A, Stewart S, et al. Adherence, adaptation and acceptance of elderly chronic heart failure patients to receiving healthcare via telephone-monitoring. Eur J Heart Fail 2007 Nov;9(11):1104-1111. [Medline: 17942364] [doi: 10.1016/j.ejheart.2007.07.018]
14. Mortara A, Pinna GD, Johnson P, Maestri R, Capomolla S, La Rovere MT, et al. Home telemonitoring in heart failure patients: the HHH study (Home or Hospital in Heart Failure). Eur J Heart Fail 2009 Mar;11(3):312-318. [Medline: 19228800] [doi: 10.1093/eurjhf/hfp022]
15. Giordano A, Scalvini S, Zanelli E, Corrà U, G L L, Ricci VA, et al. Multicenter randomised trial on home-based telemanagement to prevent hospital readmission of patients with chronic heart failure. Int J Cardiol 2009 Jan 9;131(2):192-199. [Medline: 18222552] [doi: 10.1016/j.ijcard.2007.10.027]
16. Wu RC, Delgado D, Costigan J, Maciver J, Ross H. Pilot study of an Internet patient-physician communication tool for heart failure disease management. J Med Internet Res 2005;7(1):e8 [FREE Full text] [Medline: 15829480] [doi: 10.2196/jmir.7.1.e8]
17. Ross SE, Moore LA, Earnest MA, Wittevrongel L, Lin CT. Providing a web-based online medical record with electronic communication capabilities to patients with congestive heart failure: randomized trial. J Med Internet Res 2004 May 14;6(2):e12 [FREE Full text] [Medline: 15249261] [doi: 10.2196/jmir.6.2.e12]
18. Dar O, Riley J, Chapman C, Dubrey SW, Morris S, Rosen SD, et al. A randomized trial of home telemonitoring in a typical elderly heart failure population in North West London: results of the Home-HF study. Eur J Heart Fail 2009 Mar;11(3):319-325. [Medline: 19174529] [doi: 10.1093/eurjhf/hfn050]
19. Clark RA, Inglis SC, McAlister FA, Cleland JGF, Stewart S. Telemonitoring or structured telephone support programmes for patients with chronic heart failure: systematic review and meta-analysis. BMJ 2007 May 5;334(7600):942. [Medline: 17426062] [doi: 10.1136/bmj.39156.536968.55]

20. Morguet AJ, Kühnelt P, Kallel A, Jaster M, Schultheiss HP. Impact of telemedical care and monitoring on morbidity in mild to moderate chronic heart failure. Cardiology 2008;111(2):134-139. [Medline: 18376125] [doi: 10.1159/000119701]
21. Pinna GD, Maestri R, Andrews D, Witkowski T, Capomolla S, Scanferlato JL, et al. Home telemonitoring of vital signs and cardiorespiratory signals in heart failure patients: System architecture and feasibility of the HHH model. Int J Cardiol 2007 Sep 3;120(3):371-379. [Medline: 17189654] [doi: 10.1016/j.ijcard.2006.10.029]
22. Kollmann A, Kastner P, Schreier G. Utilizing mobile phones as patient terminal in managing chronic diseases. In: Al-Hakim L, editor. Web Mobile-Based Application for Healthcare Management. Hershey, PA: IRM Press; 2007:227-257.
23. Scherr D, Zweiker R, Kollmann A, Kastner P, Schreier G, Fruhwald FM. Mobile phone-based surveillance of cardiac patients at home. J Telemed Telecare 2006;12(5):255-261. [Medline: 16848939] [doi: 10.1258/135763306777889046]
24. Chaudhry SI, Wang Y, Concato J, Gill TM, Krumholz HM. Patterns of weight change preceding hospitalization for heart failure. Circulation 2007 Oct 2;116(14):1549-1554 [FREE Full text] [Medline: 17846286] [doi: 10.1161/CIRCULATIONAHA.107.690768]
25. Morak J, Kollmann A, Schreier G. Feasibility and usability of a home monitoring concept based on mobile phones and near field communication (NFC) technology. Stud Health Technol Inform 2007;129(Pt 1):112-116. [Medline: 17911689]

Abbreviations

ACE: angiotensin converting enzyme
ARB: angiotensin receptor blocker
CHF: chronic heart failure
CRF: case report form
ESC: European Society of Cardiology
IQR: interquartile range
LV: left ventricular
MOBITEL: MOBIle TELemonitoring in heart failure patients
NFC: near field communication
NYHA: New York Heart Association

Edited by G Eysenbach; submitted 28.04.09; peer-reviewed by J Garcia, J Cafazzo; comments to author 29.05.09; revised version received 18.06.09; accepted 21.07.09; published 17.08.09.

Please cite as:
Scherr D, Kastner P, Kollmann A, Hallas A, Auer J, Krappinger H, Schuchlenz H, Stark G, Grander W, Jakl G, Schreier G, Fruhwald FM, and the MOBITEL investigators
Effect of Home-Based Telemonitoring Using Mobile Phone Technology on the Outcome of Heart Failure Patients After an Episode of Acute Decompensation: Randomized Controlled Trial
J Med Internet Res 2009;11(3):e34
URL: http://www.jmir.org/2009/3/e34/
doi:10.2196/jmir.1252
PMID:19687005

Original Paper

Test-Retest Reliability of Web-Based Retrospective Self-Report of Tobacco Exposure and Risk

Janet Brigham[1], PhD; Christina N Lessov-Schlaggar[2], PhD; Harold S Javitz[1], PhD; Ruth E Krasnow[1], MA; Mary McElroy[1], MPH; Gary E Swan[1], PhD

[1]Center for Health Sciences, SRI International, Menlo Park, CA, USA

[2]Washington University School of Medicine, St. Louis, MO, USA

Abstract

Background: Retrospectively collected data about the development and maintenance of behaviors that impact health are a valuable source of information. Establishing the reliability of retrospective measures is a necessary step in determining the utility of that methodology and in studying behaviors in the context of risk and protective factors.

Objective: The goal of this study was to examine the reliability of self-report of a specific health-affecting behavior, tobacco use, and its associated risk and protective factors as examined with a Web-based questionnaire.

Methods: Core tobacco use and risk behavior questions in the Lifetime Tobacco Use Questionnaire—a closed, invitation-only, password-controlled, Web-based instrument—were administered at a 2-month test-retest interval to a convenience sample of 1229 respondents aged 18 to 78 years. Tobacco use items, which covered cigarettes, cigars, smokeless tobacco, and pipe tobacco, included frequency of use, amount used, first use, and a pack-years calculation. Risk-related questions included family history of tobacco use, secondhand smoke exposure, alcohol use, and religiosity.

Results: Analyses of test-retest reliability indicated modest (.30 to .49), moderate (.50 to .69), or high (.70 to 1.00) reliability across nearly all questions, with minimal reliability differences in analyses by sex, age, and income grouping. Most measures of tobacco use history showed moderate to high reliability, particularly for age of first use, age of first weekly and first daily smoking, and age at first or only quit attempt. Some measures of family tobacco use history, secondhand smoke exposure, alcohol use, and religiosity also had high test-retest reliability. Reliability was modest for subjective response to first use.

Conclusions: The findings reflect the stability of retrospective recall of tobacco use and risk factor self-report responses in a Web-questionnaire context. Questions that are designed and tested with psychometric scrutiny can yield reliable results in a Web setting.

(J Med Internet Res 2009;11(3):e35) doi:10.2196/jmir.1248

KEYWORDS

Tobacco smokers; retrospective studies; psychometrics

Introduction

Studying Behavior in Context

Behaviors that can impact health are not isolated phenomena, separate from other behaviors and independent of forces influencing decisions and outcomes. Consequently, the use of tobacco and other potentially harmful substances is often studied in relation to the lifetime context of use. For example, the Rasch model analysis of smoking and alcohol use [1] reflects the intertwined relationships among different substances, as well as the advantage of studying multiple substances and risk factors in concert. Recent research focusing on aspects of social network that affect lifetime tobacco cessation outcomes [2] underscores the desirability of examining substance use in larger social and cultural contexts.

The contextual setting of tobacco use involves risk or protective factors that can affect tobacco use. These factors, as summarized by Sussman [3], can be examined readily through self-report and retrospective questioning. Factors include education, income, race and ethnicity, family use and peer use of tobacco, perceived consequences, access to tobacco, opportunities for use, cognition, habits, and addictions.

Retrospective Research

Relevant information about tobacco use is often not collected at the time events occur. This necessitates retrospective research, which has been scrutinized as a means of collecting information relating to lifetime patterns of tobacco use. Retrospective data collection allows exploration of events that may not have been perceived as important at the time they occurred. For example, a contemporary researcher desiring to study changes in tobacco

initiation ages across several decades probably would need to use retrospective techniques.

However, not all questions are amenable to retrospective inquiry, as Kenkel and colleagues [4] indicated. In examining the usefulness of retrospective measures of smoking from national survey samples, they reported that specificity in questions and statistical methods was critical for obtaining accurate retrospective measurements; they concluded that some aspects of tobacco use, such as frequent, temporary quit attempts, were not amenable to retrospective study, although the data could be helpful for studying more prolonged or permanent events. Johnson and Mott [5], studying age of onset of tobacco, alcohol, and other substance use, concluded that retrospective information typically obtained through questionnaires was adequate for most epidemiological applications. Reliability of reported age at onset [6] can benefit from aided recall and contextual information.

Reliability

Retrospective and contemporaneous examinations of risk and protective factors have demonstrated mixed psychometric adequacy. Post and colleagues [7] reported that maternal retrospective recall of smoking during pregnancy was "fairly stable over time," concurring with an earlier study by Matt and colleagues [8]. Grant and coauthors [9] studied reliability of numerous substance use disorders and associated behaviors, finding moderate to high reliability for tobacco use measures in a test-retest interval of 2 to 10 weeks. Reliability of reporting of family history of depression was high regarding parents' and siblings' conditions. Ruan and colleagues [10] found moderate to high reliability for addiction risk factor measures in 2- to 10-week test-retest. Reliability was highest for recent stressful events and stigma of alcoholism.

Self-described religiosity has been identified as a protective factor capable of attenuating an additive genetic risk for smoking initiation and thus moderating genetic influences on the liability for smoking [11]. A study of measures of spirituality, mindfulness, and substance use [12] reported moderate to high reliability for assessment of religiosity and spirituality.

These varied reports provide support for the feasibility of examining the psychometric properties of risk and protective factors related to substance use. These findings also reflect the potential utility of assessing tobacco use in the context of life events. The present study examined the reliability of retrospective questions about tobacco exposure and factors that could influence tobacco use. The instrument was a Web-based questionnaire designed to minimize error and maximize respondent involvement.

Research Goals

A primary goal of the present study was to examine 2-month test-retest reliability of retrospective self-report of tobacco use and risk-related behaviors. A previous study by the present authors [13] identified moderate to high 2-year test-retest reliability of items about lifetime tobacco use. In view of standards for reliability testing [14-16] and the positive findings from that longer-interval study, the present study addressed the following research questions: (1) How reliable is recall of tobacco use and elements of risk and protection? and (2) What factors moderate the reliability of recall?

A related goal was to continue to explore questionnaire reliability of a Web self-administration instrument. As Internet access expands [17], questions arise about the representativeness and generalizability of Web samples. We explored the reliability of Web administration within the framework of a closed, invitation-only, passcode-controlled Web-based questionnaire.

Methods

Recruitment

The Institutional Review Board of SRI International of Menlo Park, California, approved the study and determined that it was exempt from requirements for informed consent because respondents' identity was anonymous to the researchers and the responses presented no risk of jeopardy. Signed informed consent was not required. As described in the following paragraphs, participants were invited to participate in a study about tobacco use and were provided contact information for the investigators and for technical support. The study was identified on the introductory screen and succeeding screens as being sponsored by SRI International.

The sample size goal of 1200 respondents was established based on consideration of the half-width of confidence intervals for relevant statistics. This sample size was sufficient so that (1) the 95% confidence interval for a percent responding in a category at a given time would have a half-width no greater than 0.03, (2) the confidence interval for a Pearson correlation statistic of .50 for normally distributed variables would have a half-width no greater than .04, and (3) for a dichotomous variable with 75% agreement (evenly divided between agreement on each value of the variable) and a kappa statistic of 0.50, the 95% confidence interval for the true kappa would be no greater than 0.05.

The Web-based questionnaire, the Lifetime Tobacco Use Questionnaire (LTUQ), was self-administered two times, 2 months apart, by a randomly selected, invitation-only convenience sample of adults aged 18 and older drawn from a US consumer panel (e-Rewards Inc., Dallas, TX, USA). The Web panel comprised millions of persons invited from consumer databases, such as public utility customers; airline, dining, and hotel program members; and other customer groups. Opt-in membership was not allowed either to the panel or to the study.

The data were collected in two waves because of budget allocations. Nearly identical waves of 2-month test-retest administration occurred in January/March 2006, and August/October 2006. The January and August 2006 administrations were referred to as Time 1 and were grouped for statistical analysis. The March and October 2006 administrations were grouped and referred to as Time 2.

Time 1 invitations were emailed to randomly selected members of the consumer panel. Reminder invitations were sent 1 week later to those invitees who had not yet completed the LTUQ. For retesting at Time 2, all Time 1 respondents were invited to re-take the LTUQ. Time 1 respondents not responding to the

first Time 2 invitation within 1 week were sent a second invitation. Test-retest duration approximated 2 months, with variation ± 2 weeks.

Administration

Respondents' identity remained anonymous to the investigators. Respondents self-administered the questionnaire through a passcode-controlled website and could suspend the questionnaire and resume at their convenience using a passcode. The incentive was US$10 in e-Rewards scrip, the standard mechanism of payment to e-Rewards panel members.

Each respondent received a unique passcode to self-administer the questionnaire through a secure website. Cookies were not used, and IP addresses were not available to the investigators. SRI International researchers received all data without personal identifiers. Only the Web sample provider knew the respondents' identities. The sample provider did not have access to the LTUQ data and could not connect the respondents' identities with their responses. Data were encoded and collected on secure central servers and later decoded by the software provider (WebSurvent, CfMC, San Francisco, CA, USA) before the data were provided to the investigators.

Measures

The LTUQ [13] retrospectively assessed the use of any form of tobacco or nicotine across the lifespan. Developed initially in 1998, the LTUQ was tested in three earlier versions on more than 4000 respondents through computer-assisted self-interviewing (CASI), computer-assisted telephone interviewing (CATI), and computer-assisted personal interviewing (CAPI), and usability testing was conducted prior to the present CASI study. The programming utilized computerized features including skip logic, branching, and loops to shorten testing time and minimize attrition. Response options were randomized and rotated to reduce sequence effects and carryover/practice effects, with some response options anchored for consistency. The questionnaire included internal validity checks, accuracy checks, and response limitations that either prevented respondents from entering certain types of inaccurate data or flagged those responses for later examination. Because of these features, the LTUQ cannot be administered in noncomputer mode (see Multimedia Appendix 1).

A progress bar indicated the approximate percent completion of the survey as respondents proceeded through the questions. Respondents could review all prior questions and could change their responses prior to completion. Only completed questionnaires were used in the data analyses.

The LTUQ was structured around a core questionnaire that assessed the extent and nature of tobacco use from earliest exposure to the point of testing. Questions covered four major types of tobacco—cigarettes, cigars, smokeless tobacco, and pipe tobacco—and included an open-ended response option for other tobacco-delivery methods such as waterpipe or bidi. In addition to the core questions, module questions examined risk and protective factors related to tobacco use.

The core tobacco-use questions assessed initial use, transition to regular daily or weekly use, regular use, dependence, quit attempts, and abstinence. Modules of additional questions addressed (1) subjective reactions to initial use, (2) secondhand smoke exposure, (3) familial use of tobacco, (4) alcohol use, and (5) religiosity. Pack-years of smoking cigarettes was calculated from questions about the extent and duration of cigarette use and periods of abstinence.

Several minor typographical and programming errors were corrected for the August/October testing, and several additional risk-related questions were appended near the end of the questionnaire for the August/October testing.

Respondent Characteristics

Demographic characteristics of age, sex, and an estimate of median household income were evaluated as independent variables potentially affecting reliability. Since age and sex were screening variables for identifying invalid test-retest responses, reliability estimates were not calculated for those variables. Race/ethnicity and education data also were obtained but were not assessed for effects on reliability.

Respondents were grouped into terciles for examination of age effects on the reliability of recall. Age groupings were determined through the SAS procedure PROC RANK, to establish three groups of approximately the same size. The three similarly sized age groups were as follows: younger (18 to 37 years old, n = 422, mean = 30.8 years, SD = 4.2), middle (38 to 50 years old, n = 400, mean = 44.2 years, SD = 3.8), and older (51 to 78 years old, n = 402, mean = 57.6 years, SD = 5.4). Item reliabilities were calculated within each age group and compared using the chi-square test.

Median household income (in US dollars) was estimated from 2000 US Census ZIP codes [18], separated by terciles: lower ($16,383 to $41,430, n = 398, mean = $34,530, SD = $5160), middle ($41,554 to $56,585, n = 398, mean = $48,678, SD = $4466), and higher ($56,589 to $140,357, n = 396, mean = $70,895, SD = $13,200).

Tobacco Use

Questions about overall tobacco use included smoking 100 cigarettes in lifetime, frequency of use of all tobacco types, and current use.

Measures related to the first use of tobacco included (1) age at first tobacco use, (2) type of tobacco first used, (3) amount used at first exposure, and (4) subjective reactions to first tobacco use (dizzy, lightheaded, nauseated, enjoyed it, coughing/choking, liked taste, felt bad, relaxed/calm, irritated throat or lungs, head rush or buzz, felt good, difficulty inhaling, and liked smell) rated on a scale from 1 ("not at all") to 5 ("very much") or "unsure."

Lifetime frequency of tobacco use was assessed for each tobacco type (cigarettes, cigars, smokeless tobacco, pipe tobacco, other) on a 5-point scale ranging from "never used" to "used at least daily for at least 1 month." Additionally, when respondents indicated at least weekly or daily use, the frequency and amount of daily and weekly tobacco use were assessed with questions regarding (1) age at onset of weekly/daily tobacco use and (2) amount of tobacco used weekly/daily after the onset of weekly/daily use.

Current tobacco use was assessed for four primary types of tobacco (cigarettes, cigars, smokeless tobacco, pipe tobacco) plus other types. Dependence was assessed for onset of daily use of cigarettes. Quitting history included first and most recent quit attempt of at least 3 months' duration, allowing for brief lapses.

Pack-years typically is calculated by multiplying the number of packs of cigarettes smoked per day by the number of years an individual has smoked [19]. We did not calculate a comparable measure for tobacco types other than cigarettes [20] because of low use (see Table 1). We calculated pack-years in detail by averaging amount smoked across periods of known use, excluding periods of abstinence of at least 3 months' duration. Pack-years calculation was possible for only part of the subject sample because questions facilitating its calculation were added for the August/October respondents.

Risk and Protective Factors for Tobacco Use

Questions on family history of tobacco use were based on a family smoking index [21] that asked about paternal, maternal, sibling, and offspring use of tobacco.

Regarding secondhand smoke exposure, respondents were questioned about current home and vehicle rules and about children's exposure to secondhand smoke.

Alcohol use was probed with questions about ever use, age at first use, use of alcohol and tobacco together, and extent of alcohol use.

Religiosity questions were based on the Intrinsic Religious Motivation Scale [22-24], with additional questions regarding attendance at religious meetings, prayer/meditation, and participation in groups discouraging tobacco use.

Data Analyses

Analyses were calculated using SAS (SAS Institute Inc., Cary, NC, USA). Frequencies, means, percentages, standard deviations, and correlations were conducted as standard descriptive statistics. Test-retest reliability for dichotomous and categorical items was computed using the kappa statistic (k) for categorical data [25]; for ordinal and continuous measures, test-retest reliability was computed using the intraclass correlation coefficient (ICC). Reliability was rated as modest (.30 to .49), moderate (.50 to .69), or high (.70 to 1.00) for the purposes of comparison. Some demographic differences were examined with the chi-square test. Differences in test-retest reliability in men and women were compared using a 2-tailed *t* test of equality of means applied to point estimates of reliability and their asymptotic variance estimates.

We did not employ weighting techniques to match the US population or the US tobacco-user population since we were not attempting to describe population characteristics with this convenience sample.

The responses "don't know" and "unsure" were included in some analyses (indicated in table footnotes) where those responses were potentially informative about difficulty of recall, such as a "don't know" response to a question about age of first alcohol use.

Results

Median time to completion of the questionnaire was 13.7 minutes at both Time 1 and 2. Median was a more useful measure than mean because of the likelihood that respondents left the questionnaire while engaging in other activities.

Data Integrity

Responses to scaled grid questions were evaluated for the presence of straight-line responding other than "unsure/don't know" options and examined for excessively short response times. A total of 24 out of 1253 respondents at both Time 1 and Time 2 were excluded for multiple mismatches and other indices of inadequate responding [26]. Five data-point outliers excluded in data analyses (indicated in table footnotes) ranged from 40 to 2582 standard deviations from the mean and appeared to be inaccurate responses to single questions rather than intentionally incorrect responses as part of a pattern of inadequate responding (see Multimedia Appendix 2).

Response Rate

Respondents at Time 1 (N = 3142) were re-invited at Time 2; those responding to the Time 2 retest invitation and completing the LTUQ (N = 1229, 39.1% response rate, see [27]) were included in the analyses. Nonresponse due to changes in email address, Internet access, or other factors could not be determined.

Differences between Time 2 responders and nonresponders were examined on several dimensions of demographics and tobacco use. Time 1 respondents not responding at Time 2 were more likely to be slightly younger (mean = 42.7 years, SD = 12.2 for nonresponders vs mean = 44.0 years, SD = 11.9 for responders; $t = -2.93$, $P = .003$), more likely to be female (55.1%, 1068 of 1914 nonresponders vs 51.6%, 632 of 1229 responders; $\chi^2 = 9.6$, $P = .008$), less likely to report race as white (85.8%, 1642 of 1914 nonresponders vs 87.6%, 1077 of 1229 responders; $\chi^2 = 6.5$, $P = .04$), and more likely to have smoked at least 100 cigarettes in their lifetime (98.9%, 1888 of 1910 nonresponders vs 96.0%, 1175 of 1224 responders; $\chi^2 = 27.4$, $P < .001$).

Test-Retest Reliability Estimates

Most reliability estimates calculated on the test-retest sample were statistically significant, although reliability was modest for some measures.

Respondent Characteristics

Respondents included in the test-retest analyses (Time 2, N = 1229) ranged in age at Time 1 from 18 to 78 years (Table 1). Less than 1% (5 of 1229) reported never using tobacco or nicotine; their data were included if questions did not require exposure to self-administered tobacco or nicotine. About 84% (926 of 1102) of respondents reported having used cigarettes either daily or weekly, with minimal use of cigars, smokeless tobacco, or pipe tobacco. Subjects self-reported demographic information regarding education and race/ethnicity with high reliability (Table 1).

Tobacco Use

Some measures relating to lifetime tobacco use and specifically to cigarette use showed high reliability (Table 1 and Table 2). This included smoking more than 99 cigarettes in lifetime, current cigarette use, age at first use, age at first weekly and daily use, age at and duration of first or only quit attempt, and lifetime pack-years.

Although reliability of test-retest self-report of the age of first tobacco use (mean = 15.5 years reported at Time 1 and Time 2) was high, other aspects of first use reflected modest to moderate reliability. Test-retest reliability was moderate for type of tobacco first used, which reportedly was a cigarette for about 94% (1092 of 1162) of participants. Subjective responses to first use had modest to moderate reliability (Table 2).

Separate sets of questions asked about the age of first weekly smoking and the amount used at that time, and the age of first daily smoking and the amount used at that time. Reliability was higher for age at onset of weekly or daily use than for the number of cigarettes used, which had moderate reliability. Dependence-related questions regarding the time to first cigarette in the morning at the onset of daily cigarette use had moderate reliability (Table 2). Age at first or only quit attempt of at least 3 months' duration exhibited high reliability, as did the duration of that quit attempt and the use of a cessation aid (Table 2).

Test-retest calculation of pack-years, a common metric for evaluating tobacco use across the lifespan, was evaluated in the August/October group only because a question added mid-study made the pack-years calculations possible. Reliability of the pack-years calculation was high (Table 2).

Table 1. Test-retest reliability of respondents' self-report of demographics and tobacco use

	Time 1	Time 2	No.	ICC or κ[a]	95% CI or SE
Demographics					
Age, years, mean (SD)	44.0 (11.9)	44.1 (11.9)	1229		
Female, %	51.6	51.5	1229		
Education: > high school, %	88.9	89.0	1229	κ = 0.88	0.86, 0.90
Ethnicity: white, %	87.6	87.9	1229	κ = 0.86	0.82, 0.90
Lifetime use of cigarettes					
Smokers reporting using > 99 cigarettes/lifetime, %	96.22	96.14	1217	κ = 0.81	0.72, 0.90
Total cigarettes smoked if < 100/lifetime, mean (SD)	19.3 (25.1)	22 (23.7)	38	ICC = 0.70	0.08
Frequency of tobacco use					
Cigarettes daily, % (no.)	83.7 (922)	84.0 (926)	1102	κ = 0.51	0.44, 0.57
Cigars, % ever weekly or daily (no.)	6.6 (73)	5.5 (61)	1102	κ = 0.66	0.63, 0.70
Smokeless tobacco, % ever weekly or daily (no.)	2.8 (31)	2.7 (30)	1102	κ = 0.71	0.66, 0.76
Pipe tobacco, % ever weekly or daily (no.)	2.5 (28)	2.0 (22)	1102	κ = 0.70	0.66, 0.74
Other tobacco/nicotine, % (no.)	6.2 (68)	5.3 (58)	1102	κ = 0.50	0.38, 0.61
Current tobacco use					
Number of cigarettes/week, mean (SD)	103.7 (88.1)	103.7 (89.4)	859	ICC = 0.83	0.01
Number of cigars/week, mean (SD)	0.91 (8.8)	1.19 (9.9)	858	ICC = 0.30[b]	0.03
Number of tins of smokeless tobacco/week[b], mean (SD)	0.07 (0.6)	0.09 (0.7)	858	ICC = 0.90	0.01
Number of pipe tobacco uses/week, mean (SD)	0.14 (2.2)	0.15 (2.2)	858	ICC = 0.98[b]	0.001

[a] Reliability was not calculated for age and sex because those variables were used for screening.

[b] ICC calculations excluded one outlier.

Table 2. Test-retest reliability of self-report details of tobacco use history

	Time 1	Time 2	No.	ICC or κ	95% CI or SE
First use of tobacco					
Age first tried tobacco, years, mean (SD)	15.5 (3.5)	15.5 (3.9)	1162	ICC = 0.81	0.01
First tobacco was cigarette, %	94.2	93.4	1162	κ = 0.51	0.38, 0.64
Experience at first use of tobacco (1–5 scale), mean (SD)					
Dizzy	3.14 (1.3)	3.19 (1.3)	1162	ICC = 0.51	0.02
Lightheaded	3.10 (1.3)	3.15 (1.3)	1162	ICC = 0.49	0.02
Nauseated	2.21 (1.3)	2.28 (1.3)	1162	ICC = 0.54	0.02
Enjoyed it	3.00 (1.1)	2.87 (1.1)	1162	ICC = 0.50	0.02
Coughing/choking	2.91 (1.3)	2.97 (1.3)	1162	ICC = 0.51	0.02
Liked taste	2.56 (1.2)	2.51 (1.1)	1162	ICC = 0.51	0.02
Felt bad	2.19 (1.2)	2.24 (1.2)	1162	ICC = 0.51	0.02
Relaxed/calm	2.70 (1.2)	2.61 (1.1)	1162	ICC = 0.36	0.03
Irritated throat	2.64 (1.3)	2.73 (1.3)	1162	ICC = 0.49	0.02
Head rush/buzz	3.36 (1.3)	3.47 (1.2)	1162	ICC = 0.52	0.02
Felt good	2.58 (1.2)	2.51 (1.1)	1162	ICC = 0.38	0.03
Difficulty inhaling	2.59 (1.4)	2.69 (1.4)	1162	ICC = 0.41	0.03
Liked smell	2.44 (1.2)	2.46 (1.2)	1162	ICC = 0.51	0.02
Weekly use of cigarettes					
Age first smoked cigarettes at least weekly, years, mean (SD)	17.3 (4.2)	17.3 (4.1)	913	ICC = 0.85	0.01
Number of cigarettes/week when started weekly use, mean (SD)	34.3 (35.1)	34.2 (39.3)	554	ICC = 0.52	0.03
Daily use of cigarettes					
Age first used daily, years, mean (SD)	17.6 (4.5)	17.6 (4.3)	857	ICC = 0.82	0.01
Number of cigarettes/day when started daily use, mean (SD)	9.64 (7.2)	9.50 (7.3)	515	ICC = 0.54	0.03
Smoked < 1 hour after waking when started daily use, %	26.3	26.5	515	κ = 0.53	0.44, 0.62
Minutes to first cigarette of day when started daily use, mean (SD)	160.4 (148)	148.6 (137)	411	ICC = 0.57	0.03
First or only cigarette quit attempt of ≥ 3 months' duration					
Age, years, mean (SD)	28.3 (9.7)	28.2 (9.8)	546	ICC = 0.86	0.01
Number of months, mean (SD)	13.5 (17.6)	12.9 (16.6)	430	ICC = 0.79	0.02
Used cessation aid, %	25.1	24.0	546	κ = 0.71	0.64, 0.78
Pack-years					
Pack-years, mean (SD)	19.1 (18.1)	19.9 (18.7)	504	ICC = 0.76	0.02

Risk and Protective Factors for Tobacco Use

Reliability was moderate or high for questions about the four risk/protective categories: family history of tobacco use, secondhand smoke exposure, alcohol use, and religiosity (Table 3). Reliability of family history reports of parental, sibling, and offspring tobacco use were high. Questions about exposure to secondhand smoke indicated moderate to high reliability. Questions about alcohol use ranged in reliability from moderate to high. Respondents indicated at both Time 1 and Time 2 that when they drank alcohol, they also used tobacco about 60% of the time. Among questions about religiosity, those indicating highest reliability were regarding seeking divine guidance in decision making, and serving God (Table 3). Reliability of other questions regarding religiosity ranged from modest to high.

Table 3. Test-retest reliability of measures of risk for and protection against tobacco use[a]

	Time 1	Time 2	No.	ICC or κ	95% CI or SE
Family history of tobacco use					
Mother used tobacco, %	47.8	48.6	1188	κ = 0.86	0.83, 0.89
Father used tobacco, %	72.4	74.0	1148	κ = 0.75	0.71, 0.79
Number of siblings who used tobacco, mean (SD)	1.38 (1.5)	1.40 (1.5)	392	ICC = 0.84	0.02
Number of offspring who used tobacco, mean (SD)	0.68 (1.0)	0.63 (0.9)	243	ICC = 0.87	0.02
Exposure to secondhand smoke					
Smoking currently allowed inside home, % no	56.0	57.2	1213	κ = 0.80	0.77, 0.83
Smoking currently allowed in car, % no	38.2	38.4	1213	κ = 0.83	0.80, 0.86
Children currently exposed to smoke inside home, % no	85.6	87.0	1229	κ = 0.65	0.58, 0.71
Alcohol use					
Ever used alcohol, %	91.2	89.3	522	κ = 0.41	0.29, 0.54
Drink alcohol currently, %	80.4	75.9	515	κ = 0.59	0.50, 0.67
Age first tried alcohol, years, mean (SD)	15.9 (2.8)	15.8 (3.1)	444	ICC = 0.70	0.02
Used any form of tobacco when first used alcohol[b], %	68.5	62.9	143	κ = 0.64	0.47, 0.80
Used cigarettes when first tried alcohol[b], %	73.0	69.7	337	κ = 0.63	0.54, 0.72
How often drink, % at least several times per week	27.2	28.1	442	ICC = 0.80	0.02
Number of drinks when use alcohol, mean (SD)[c]	3.10 (2.3)	3.08 (2.3)	407	ICC = 0.76	0.02
Use tobacco now when drink alcohol, % of time	59.6	58.6	917	ICC = 0.78	0.01
Religiosity (scale 1–5) [d]					
My faith involves all of my life, mean (SD)	2.67 (1.5)	2.70 (1.5)	454	ICC = 0.72	0.02
One should seek God's guidance when making every important decision, mean (SD)	3.05 (1.6)	3.05 (1.6)	1035	ICC = 0.84	0.01
In my life I experience the presence of the divine, mean (SD)	2.81 (1.5)	2.83 (1.5)	434	ICC = 0.70	0.02
My faith sometimes restricts my actions.	2.39 (1.5)	2.47 (1.5)	462	ICC = 0.63	0.03
Nothing is as important to me as serving God as best I know how, mean (SD)	2.64 (1.5)	2.68 (1.5)	452	ICC = 0.82	0.02
I try hard to carry my religion over into all my other dealings in life, mean (SD)	2.62 (1.5)	2.66 (1.4)	463	ICC = 0.70	0.02
My religious beliefs are what really lie behind my whole approach to life, mean (SD)	2.57 (1.5)	2.60 (1.4)	455	ICC = 0.75	0.02
It doesn't matter so much what I believe as long as I lead a moral life, mean (SD)	3.57 (1.5)	3.53 (1.5)	462	ICC = 0.47	0.04
Do you participate or believe in one specific religion or belief system?, % yes	52.8	52.5	1224	κ = 0.71	0.67, 0.74
How frequently do you attend church meetings or gatherings associated with this religion or belief system?, % daily or weekly	18.1	17.6	1224	κ = 0.55	0.51, 0.58
How often do you pray or meditate in an effort to communicate with deity, or with what some people call a "higher power"?, % daily or weekly	49.5	50.7	1224	κ = 0.52	0.48, 0.55
Have you ever participated in a religious or social group that discourages or prohibits tobacco use?, % yes	12.9	12.1	703	κ = 0.53	0.45, 0.60

[a] Some questions were added to the LTUQ between first and second waves of test-retest administration, resulting in lower cell sizes.

[b] Findings include responses of "yes," "no," "unsure," or "decline to state."

[c] ICC calculations excluded two outliers.

[d] Adapted from Hoge and colleagues [22,23]. Scale: 1 ("disagree") to 5 ("agree") plus "unsure" or "decline to state," unless indicated otherwise in parentheses.

Sex, Age, and Income

Reliability estimates of several questions about the frequency of tobacco use and the age of first use differed between men and women. Statistically significant results were as follows: Reliability of self-reported age at first use was higher for women (0.84) than for men (0.78; $P < .001$). However, men's self-reported age at onset of weekly smoking had higher reliability (0.92) than that of women (0.79; $P < .001$). Women reported pack-years with higher reliability (0.81) than did men (0.66; $P < .001$). Women also recalled the level of first-use head rush/buzz (0.60 vs 0.45 for men; $P < .001$) and difficulty inhaling (0.47 vs 0.35 for men; $P = .02$) with higher reliability (see Multimedia Appendix 3, Supplementary Table 1).

The reliability estimates of several questions varied by age group. Statistically significant results were as follows: Younger respondents (18 to 37 years) reported the age at first tobacco use less reliably (ICC = 0.77) than their older counterparts (middle, 38 to 50 years, ICC = 0.81; and older, 51 to 78 years, ICC = 0.85; $P = .01$). However, the younger group's reporting of the age at first daily use (ICC = 0.91) showed higher reliability than that of middle (ICC = 0.82) and older (ICC = 0.79) respondents ($P < .001$). Reliability was high among all age groups for both questions. Younger respondents' reporting of pack-years was high (ICC = 0.89), whereas the middle (ICC = 0.66) and older (ICC = 0.68) groups' response reliability was moderate ($P < .001$). For two subjective responses to first tobacco use (irritated throat [$P = .01$], felt good [$P = .04$]), the younger group's responses had modest and moderate reliability, while those in older groups either had modest reliability or did not meet criteria for modest reliability (see Multimedia Appendix 3, Supplementary Table 2).

The only statistically significant tobacco-use difference based on median household income was the amount of tobacco used the first time, which was reported somewhat more reliably by middle-income respondents ($P < .001$), although the reliability for all three income groups was modest (see Multimedia Appendix 3, Supplementary Table 3).

Discussion

Research Goals

These findings paralleled our 2-year CASI reliability study of a similar Web-based sample [13]. The present findings supported the supposition that key questions retrospectively asking about tobacco use and elements of risk and protection can be recalled with moderate to high reliability in a Web-browser environment. Potentially salient events and aspects of risk can be recalled more reliably than less memorable events (eg, age of first use of tobacco elicited higher reliability than type of tobacco first tried or amount used at first try). Sex, age, and approximated income effects in both studies indicated few reliability variations based on those characteristics.

Reliability estimates in the present 2-month study were generally higher than those reported in the 2-year test-retest reliability study of an earlier version of the LTUQ [13]. In the earlier CASI study, also conducted on a closed, invitation-only, randomly selected Web-panel convenience sample, the apparent salience of events affected reliability to a greater extent than in the present study. A 2-month test-retest administration may be more subject to carryover effects from persistence of memory, although the risk is smaller than for the shorter time intervals common in psychometric analyses of substance use questions. Some questions with low reliability in the 2-year study were not included in this study because of their apparent psychometric inadequacy.

The reliability of questions about subjective response to first use of tobacco was more modest in the 2-year study, although the present findings of scaled responses showed only modest to moderate reliability. When responses were dichotomized to *any* versus *none* in the 2-year reliability study, reliability was higher. The modest reliability of these measures suggests the advisability of neither expecting nor requiring fine-tuned recall of early events.

The relative strength of the pack-years reliability measure was comparable to that reported by Bernaards and colleagues [28], indicating that reliability of retrospective recall of pack-years can approximate that of prospective measurement, with some limitations.

The findings also provided support for exploring and expanding the use of the Web for questionnaire self-administrations. The rapid expansion of Internet access across the US population has made panel participation feasible for an increasingly broader range of respondents. A lingering question, however, is whether Internet penetration remains so linked to income and education levels that ascertaining a sufficiently broad or representative Web sample is possible. Recent findings from the Pew Internet & American Life Project [17] indicate that Internet access is no longer the domain of the young, but now crosses all age boundaries. Some 87% of those aged 30-34 use the Internet, with 83% of those aged 40-44, 80% of those aged 35-39, 80% of those 45-49, and 78% of those aged 50-54. Internet use is growing most rapidly in the 70-75 age group, with 45% currently online. Between 70% and 80% of all those online have home broadband access.

Internet access also is no longer the domain of only the wealthy and educated. As early as 2004, a commissioned research study [29] found that computer use was more than 72% for those with a high school diploma, and exceeded 86% for all other education level groups. Education attainment information collected annually through the US Census Bureau's American Community Survey and the Current Population Survey [30] indicated that 84% of US adults older than age 25 had at least a high school diploma or equivalent. Some 54.4% had at least some college. Education attainment figures reflect some racial and ethnic disparities, such as higher education levels among Asians and lower education levels among individuals not born in the United States.

Also of concern is whether a Web sample can be representative of US smokers. A 2007 Centers for Disease Control and Prevention (CDC) report [31] indicated that smoking prevalence varied by education, with higher smoking rates among those with less than 12 years of schooling (33.3% of smokers) and those with a diploma equivalent (44.0%). Smoking rates were lower among those with more education. They were also lower

among those 65 years or older (8.3%), compared with smoking rates between 21.0% and 22.8% for younger groups. Smoking rates were higher (28.8%) among those below the federal poverty level than among those above that level (20.3%). The most current census report [30] does not delineate nativity, which influences education level and could affect smoking rates.

Limitations

Sample

The findings were not intended for extrapolation to the general US population or US tobacco users. Since this study was conducted on a convenience sample, its representativeness and generalizability relative to the US population were undetermined. A quota-cell, weighted, or other population-based study was beyond the scope of this research. Education level of respondents was higher than that of the US population. Race did not approximate national statistics. Additionally, the sample may have underrepresented groups still lagging in education and in Internet access, such as individuals with disabilities, those born outside the United States, and those below the federal poverty level.

The median-income approximation should be interpreted with caution because the measure, based on ZIP codes, was an indirect determination.

Validity

Although reliability does indicate repeatability and stability of responses, acceptable levels of reliability do not establish the validity of responses. It is possible that subjects responded consistently but inaccurately. The investigators currently are examining LTUQ validity in two longitudinal samples.

Pack-Years Calculation

The process of estimating pack-years from LTUQ data may have underestimated or overestimated actual total consumption. It also did not take into account the use of other types of tobacco, which would have been feasible but would have required a considerably more complex calculation. The validity of our estimation approach depended on the assumption of a linear change in the number of cigarettes smoked between any two time points for which cigarette consumption was stated. This assumption may have been particularly questionable when the individual had never succeeded in quitting for 3 or more months, and when many years separated the questionnaire administration date and the date when the individual first smoked weekly. If the ramp-up were more rapid than linear, we would have tended to underestimate pack-years. Also, if the individual temporarily reduced cigarette consumption prior to starting a quit attempt, our estimation approach would have tended to underestimate cigarette consumption for the interval ending on the date that quit attempt started. Finally, missing information about other quit attempts (of any duration) may have resulted in overestimation of pack-years.

Conclusions

This study reinforced the expectation that retrospectively collected self-report data about the development and maintenance of addictive behaviors can be a valuable and reliable source of information about lifetime substance use [4]. The present results add to the evidence indicating that this relatively economical approach can yield reliable reports of behaviors that have not been captured in real time. The findings thus provide support for exploring and expanding the use of the Web for questionnaire self-administrations.

As Internet penetration breaks through demographic boundaries, sampling can more readily include those with less education, lower income levels, and those in older age ranges. Even so, in spite of greater relative ease of access, accurate Web-based research will continue to require appropriate sampling and analytic procedures, as well as cautious interpretation and extrapolation.

Acknowledgments

The authors express appreciation to Nancy Chong for providing programming and Web hosting expertise. Psychometric work on the LTUQ has been funded by NIH grant DA018019 to Gary E Swan. Initial development work on the LTUQ was under subcontract to University of Michigan, NIH grant CA75581 to Ovide F Pomerleau. Development of the computerized precursor of the LTUQ was funded by NIH grant DA11795 to Janet Brigham. The sponsors exerted no influence in the study design, data collection, analysis, interpretation of data, writing of the report, and decision to submit the paper for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screen views from the LTUQ

[PDF file (Adobe PDF), 427 KB - jmir_v11i3e35_app1.pdf]

Multimedia Appendix 2

Evaluating data integrity in retrospective recall of lifetime tobacco use

[PDF file (Adobe PDF), 35 KB - jmir_v11i3e35_app2.pdf]

Multimedia Appendix 3

Supplementary tables

[PDF file (Adobe PDF), 124 KB - jmir_v11i3e35_app3.pdf]

References

1. Kahler CW, Strong DR, Papandonatos GD, Colby SM, Clark MA, Boergers J, et al. Cigarette smoking and the lifetime alcohol involvement continuum. Drug Alcohol Depend 2008 Jan 11;93(1-2):111-120 [FREE Full text] [Medline: 17964082] [doi: 10.1016/j.drugalcdep.2007.09.004]
2. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med 2008 May 22;358(21):2249-2258 [FREE Full text] [Medline: 18499567] [doi: 10.1056/NEJMsa0706154]
3. Sussman S. Risk factors for prevention of tobacco use. Pediatric Blood & Cancer 2005;44(7):614-619. [doi: 10.1002/pbc.20350]
4. Kenkel D, Lillard DR, Mathios A. Smoke or fog? The usefulness of retrospectively reported information about smoking. Addiction 2003 Sep;98(9):1307-1313. [Medline: 12930218] [doi: 10.1046/j.1360-0443.2003.00445.x]
5. Johnson TP, Mott JA. The reliability of self-reported age of onset of tobacco, alcohol and illicit drug use. Addiction 2001 Aug;96(8):1187-1198. [Medline: 11487424] [doi: 10.1080/09652140120060770]
6. Huerta M, Chodick G, Balicer RD, Davidovitch N, Grotto I. Reliability of self-reported smoking history and age at initial tobacco use. Prev Med 2005 Aug;41(2):646-650. [Medline: 15917064] [doi: 10.1016/j.ypmed.2005.01.011]
7. Post A, Gilljam H, Bremberg S, Galanti MR. Maternal smoking during pregnancy: a comparison between concurrent and retrospective self-reports. Paediatr Perinat Epidemiol 2008 Mar;22(2):155-161. [Medline: 18298690] [doi: 10.1111/j.1365-3016.2007.00917.x]
8. Matt GE, Hovell MF, Zakarian JM, Bernert JT, Pirkle JL, Hammond SK. Measuring secondhand smoke exposure in babies: the reliability and validity of mother reports in a sample of low-income families. Health Psychol 2000 May;19(3):232-241. [Medline: 10868767] [doi: 10.1037/0278-6133.19.3.232]
9. Grant BF, Dawson DA, Stinson FS, Chou PS, Kay W, Pickering R. The Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV): reliability of alcohol consumption, tobacco use, family history of depression and psychiatric diagnostic modules in a general population sample. Drug Alcohol Depend 2003 Jul 20;71(1):7-16. [Medline: 12821201] [doi: 10.1016/S0376-8716(03)00070-X]
10. Ruan WJ, Goldstein RB, Chou SP, Smith SM, Saha TD, Pickering RP, et al. The alcohol use disorder and associated disabilities interview schedule-IV (AUDADIS-IV): reliability of new psychiatric diagnostic modules and risk factors in a general population sample. Drug Alcohol Depend 2008 Jan 1;92(1-3):27-36 [FREE Full text] [Medline: 17706375] [doi: 10.1016/j.drugalcdep.2007.06.001]
11. Timberlake DS, Rhee SH, Haberstick BC, Hopfer C, Ehringer M, Lessem JM, et al. The moderating effects of religiosity on the genetic and environmental determinants of smoking initiation. Nicotine Tob Res 2006 Feb;8(1):123-133. [Medline: 16497606] [doi: 10.1080/14622200500432054]
12. Leigh JT, Bowen S, Marlatt GA. Spirituality, mindfulness and substance abuse. Addict Behav 2005;30(7):1335-1341. [doi: 10.1016/j.addbeh.2005.01.010]
13. Brigham J, Lessov-Schlaggar CN, Javitz HS, McElroy M, Krasnow R, Swan GE. Reliability of adult retrospective recall of lifetime tobacco use. Nicotine Tob Res 2008 Feb;10(2):287-299. [Medline: 18236293] [doi: 10.1080/14622200701825718]
14. Litwin MS. How to Measure Survey Reliability and Validity. Thousand Oaks, CA: Sage; 1995.
15. Thompson B. Guidelines for authors reporting score reliability estimates. In: Thompson B, editor. Score Reliability: Contemporary Thinking on Reliability Issues. Thousand Oaks, CA: Sage; 2003:93-95.
16. Thompson B. Understanding reliability and coefficient alpha, really. In: Thompson B, editor. Score Reliability: Contemporary Thinking on Reliability Issues. Thousand Oaks, CA: Sage; 2003:5.
17. Jones S, Fox S. Generations online in 2009. Washington, DC: Pew Internet & American Life Project; 2009. URL: http://www.pewinternet.org/Reports/2009/Generations-Online-in-2009.aspx?r=1 [WebCite Cache ID 5fqTivxpl]
18. Thomas AJ, Eberly LE, Davey Smith G, Neaton JD. ZIP-code-based versus tract-based income measures as long-term risk-adjusted mortality predictors. Am J Epidemiol 2006 Sep 15;164(6):586-590 [FREE Full text] [Medline: 16893922] [doi: 10.1093/aje/kwj234]
19. National Cancer Institute. Dictionary of cancer terms. URL: http://www.cancer.gov/Templates/db_alpha.aspx?CdrID=306510 [WebCite Cache ID 5fqTaQx6k]
20. Wood DM, Mould MG, Ong SBY, Baker EH. "Pack year" smoking histories: what about patients who use loose tobacco? Tob Control 2005 Apr;14(2):141-142 [FREE Full text] [Medline: 15791025] [doi: 10.1136/tc.2004.009977]
21. Drobes DJ, Munafò MR, Leigh F, Saladin ME. A family smoking index to capture genetic influence in smoking: rationale and two validation studies. Nicotine Tob Res 2005 Feb;7(1):41-46. [Medline: 15804676] [doi: 10.1080/14622200412331328510]
22. Hoge DR. A validated intrinsic religious motivation scale. J Sci Stud Relig 1972;11(4):369-376. [doi: 10.2307/1384677]

23. Hoge DR, Carroll JW. Determinants of commitment and participation in suburban Protestant churches. J Sci Stud Relig 1978;17(2):107-127. [doi: 10.2307/1386155]
24. Bassett RL. Intrinsic Religious Motivation Scale. In: Hill PC, Hood RWJ, editors. Measures of Religiosity. Birmingham, AL: Religious Education Press; 1999:135-137.
25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571] [doi: 10.2307/2529310]
26. Brigham J, Lessov-Schlaggar CN, Krasnow RE, Javitz HS, Swan GE. Evaluating data integrity in retrospective recall of lifetime tobacco use. Presented at: Proceedings of the 13th Annual Meeting of the Society for Research on Nicotine and Tobacco; February 21-24, 2007; Austin, TX POS114 URL: http://srnt.org/meeting/2007/pdf/onsite/2007SRNTAbstracts-FINAL.pdf [WebCite Cache ID 5fqSoUR7h]
27. ; American Association for Public Opinion Research. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 5th edition. Lenexa, KS: AAPOR; 2008. URL: http://aapor.org/uploads/Standard_Definitions_07_08_Final.pdf [WebCite Cache ID 5fv0x5cxK]
28. Bernaards CM, Twisk JW, Snel J, Van Mechelen W, Kemper HC. Is calculating pack-years retrospectively a valid method to estimate life-time tobacco smoking? A comparison between prospectively calculated pack-years and retrospectively calculated pack-years. Addiction 2001 Nov;96(11):1653-1661. [Medline: 11784461] [doi: 10.1080/09652140120080778]
29. ; Microsoft Corporation. Accessible Technology in Computing—Examining Awareness, Use, and Future Potential. Microsoft Corporation. 2004. URL: http://www.microsoft.com/enable/research/phase2.aspx [WebCite Cache ID 5fqTW8x7D]
30. ; U.S. Census Bureau. Educational Attainment in the United States: 2007. Washington, DC: US Department of Commerce, U.S. Census Bureau; 2009 Jan. URL: http://www.census.gov/prod/2009pubs/p20-560.pdf [WebCite Cache ID 5fqSOFdQn]
31. Thorne SL, Malarcher A, Maurice E, Caraballo R. Cigarette smoking among adults -- United States, 2007 [published erratum appears in MMWR 2008 Nov;57(45):1281]. MMWR 2008;57(45):1221-1226 [FREE Full text] [WebCite Cache ID 5fqTrWjrX] erratum: [FREE Full text] [WebCite Cache ID 5fqTrWjrX]

Abbreviations

CASI: computer-assisted self-interviewing
LTUQ: Lifetime Tobacco Use Questionnaire

Edited by G Eysenbach; submitted 13.04.09; peer-reviewed by S Viehbeck, B Bock; comments to author 08.05.09; revised version received 18.05.09; accepted 21.07.09; published 11.08.09.

Please cite as:
Brigham J, Lessov-Schlaggar CN, Javitz HS, Krasnow RE, McElroy M, Swan GE
Test-Retest Reliability of Web-Based Retrospective Self-Report of Tobacco Exposure and Risk
J Med Internet Res 2009;11(3):e35
URL: http://www.jmir.org/2009/3/e35/
doi:10.2196/jmir.1248
PMID:19674962

Original Paper

Recruitment to a Randomized Web-Based Nutritional Intervention Trial: Characteristics of Participants Compared to Non-Participants

Melanie A Stopponi[1], MPA, CHES; Gwen L Alexander[2], PhD; Jennifer B McClure[3], PhD; Nikki M Carroll[1], MS; George W Divine[2], PhD; Josephine H Calvi[4], MPH; Sharon J Rolnick[5], PhD; Victor J Strecher[6], PhD, MPH; Christine Cole Johnson[2], PhD; Debra P Ritzwoller[1], PhD

[1]Institute for Health Research, Kaiser Permanente, Denver, Colorado, USA

[2]Henry Ford Hospital, Detroit, Michigan, USA

[3]Group Health Center for Health Studies, Seattle, Washington, USA

[4]Kaiser Permanente Georgia, Center for Health Research Southeast, Atlanta, Georgia, USA

[5]HealthPartners Research Foundation, Minneapolis, Minnesota, USA

[6]Center for Health Communications Research, University of Michigan School of Public Health, Ann Arbor, MI, USA

Corresponding Author:
Melanie A Stopponi, MPA, CHES
Institute for Health Research
Kaiser Permanente, Colorado
PO Box 378066
Denver, CO 80237-8066
USA
Phone: +1 303 614 1207
Fax: +1 303 614 1205
Email: Melanie.A.Stopponi@kp.org

Abstract

Background: Web-based behavioral programs efficiently disseminate health information to a broad population, and online tailoring may increase their effectiveness. While the number of Internet-based behavioral interventions has grown in the last several years, additional information is needed to understand the characteristics of subjects who enroll in these interventions, relative to those subjects who are invited to enroll.

Objective: The aim of the study was to compare the characteristics of participants who enrolled in an online dietary intervention trial (MENU) with those who were invited but chose not to participate, in order to better understand how these groups differ.

Methods: The MENU trial was conducted among five health plans participating in the HMO Cancer Research Network in collaboration with the University of Michigan Center for Health Communication Research. Approximately 6000 health plan members per site, between the ages of 21 and 65, and stratified by gender with oversampling of minority populations, were randomly selected for recruitment and were mailed an invitation letter containing website information and a US$2 bill with the promise of US$20 for completing follow-up surveys. Administrative and area-based data using geocoding along with baseline survey data were used to compare invitees (HMO members sent the introductory letter), responders (those who entered a study ID on the website), and enrollees (those who completed the enrollment process). Generalized estimating equation multivariate and logistic regression models were used to assess predictors of response and enrollment.

Results: Of 28,460 members invited to participate, 4270 (15.0%) accessed the website. Of the eligible responders, 2540 (8.9%) completed the consent form and baseline survey and were enrolled and randomized. The odds of responding were 10% lower for every decade of increased age ($P < .001$), while the likelihood of enrolling was 10% higher for every decade increase in age ($P < .001$). Women were more likely to respond and to enroll ($P < .001$). Those living in a census tract associated with higher education levels were more likely to respond and enroll, as well as those residing in tracts with higher income ($P < .001$). With a 22% (n = 566) enrollment rate for African Americans and 8% (n = 192) for Hispanics, the enrolled sample was more racially and ethnically diverse than the background sampling frame.

Conclusions: Relative to members invited to participate in the Internet-based intervention, those who enrolled were more likely to be older and live in census tracts associated with higher socioeconomic status. While oversampling of minority health plan members generated an enrolled sample that was more racially and ethnically diverse than the overall health plan population,

additional research is needed to better understand methods that will expand the penetration of Internet interventions into more socioeconomically diverse populations.

Trial Registration: Clinicaltrials.gov NCT00169312; http://clinicaltrials.gov/ct2/show/NCT00169312 (Archived by WebCite at http://www.webcitation.org/5jB50xSfU)

(J Med Internet Res 2009;11(3):e38) doi:10.2196/jmir.1086

KEYWORDS

Recruitment; Web-based interventions; making effective nutritional choices; Cancer Research Network; CRN; fruits and vegetables; research subject selection; selection; patient; mass screening; Internet; motivation; cultural diversity; health maintenance organizations

Introduction

Web-based behavioral programs can efficiently and effectively disseminate health education to a broad population [1] and allow people to access multimedia information on their own schedule, at their own pace and location, without geographic limitations or the expense of face-to-face interactions [2]. Online tailoring may increase the effectiveness of these programs [3]. Indeed, many health plans and for-profit wellness companies are now offering Web-based health behavior change programs [1], and more attention is now being given to understanding effective recruitment for such Web-based programs [4]. The future effectiveness of Internet-delivered health promotion programs as population-based interventions will depend on their ability to reach and engage a broad variety of users.

While Internet-based programs have the potential to reach millions of people, potential reach is not actual reach [5]. Health plans (health maintenance organizations, HMOs) maintain comprehensive electronic data, making the recruitment process efficient since individuals can be readily identified and targeted by age, gender, diagnoses, and procedures. These databases also allow the research teams to track participation and eventually outcomes related to health and medical care [6,7].

We previously reported on the importance of combining pre- and post-enrollment financial incentives for recruiting and maintaining participation in an online health promotion program [8]. A recent study that compared response rates by gender and racial subgroups to determine incentive combinations that would better recruit a diverse group of healthy adults who were members of a Midwestern HMO to an online health program [8] determined that a combination of a US$2 bill and the promise of US$20 for completing an online follow-up survey was most effective across gender and broad racial/ethnic subgroups. What is not known is the effect of this combination across diverse geographic regions.

Furthermore, to our knowledge, no one has examined who enrolls in online health promotion programs and how the characteristics of these people differ from those who do not participate. This information has important implications for those designing and marketing online health promotion programs; however, data on nonresponders are often unavailable, making such a comparison impossible. Using automated health plan data and geocoding, we were able to examine this issue in the Making Effective Nutritional Choices (MENU) trial, which is an online dietary intervention study funded by the National Cancer Institute and open to members of five US health plans [9].

Based on data from other health promotion programs, both Internet-based and otherwise [10], we hypothesized that women and persons of higher socioeconomic status would be more likely to enroll in MENU [11-13]. This paper reports the results of the MENU recruitment efforts, including a comparison of the demographic characteristics of people who were invited but did not respond (invitees), people who visited the website and entered their log-in ID but did not enroll (responders), and those who were eligible and enrolled (enrollees). We believe our results shed important light on the demographic characteristics of people interested and willing to participate in an online health promotion intervention study and will help inform this burgeoning area of interest.

Methods

Setting

We developed an Internet-based program to promote increased intake of fruits and vegetables, “Making Effective Nutritional Choices (MENU),” and tested it among diverse members of five health plans. MENU was a randomized trial conducted in conjunction with the HMO Cancer Research Network (CRN) [9]. The CRN consists of the research programs, enrollee populations, and legacy databases of the participating integrated health care organizations. The goal of the CRN is to conduct research on cancer prevention, early detection, treatment, long-term care, and surveillance [14]. Five of the CRN affiliated health care delivery systems—Group Health in Seattle, Kaiser Permanente Colorado in Denver, HealthPartners in Minneapolis, Henry Ford Health System in Detroit, and Kaiser Permanente Georgia in Atlanta—collaborated with the University of Michigan’s Center for Health Communications Research for the MENU study. The Institutional Review Boards from all participating institutions approved this study prior to data collection.

Recruitment

For the remainder of the paper, the following terms with associated definitions will apply: Invitees are defined as members of one of the five health plans noted above who were mailed a recruitment letter; responders are defined as any invitee who entered a study ID on the MENU website; enrollees are defined as responders who completed all enrollment steps in

the study, which included an online eligibility survey, consent form, and baseline survey.

Invitees

Using administrative databases, each site identified a random sample of individuals aged 21-65 years at the beginning of the study enrollment period (September 2005) who were current members with at least 1 year enrollment in the respective health plan with no enrollment gaps greater than 90 days at one time. We used diagnostic codes in the health plan databases to exclude from the sample anyone with a medical condition that could be negatively affected by increasing intake of fruits and vegetables. These conditions included current cancer treatment, gastroparesis, neurological conditions, mental health conditions, and use of anticoagulant medications. From potentially eligible members, we drew a random sample of approximately 6000 individuals from each participating health plan, stratified by gender, with a 10% response goal for study enrollment.

At three sites, minority racial/ethnic groups (African American at Sites 2 and 5, and Hispanic at Site 4) were oversampled in order to enhance the enrollment of these populations. The demographic characteristics of the respective health plan memberships are shown in Table 1. At the time this study began, health plans did not prospectively collect data on the race and ethnicity of their members. Race and ethnicity of the respective health plan members were estimated based on various health plan surveys, census data, etc. Approximately 35% of the membership of Site 2 was African American. Only at Site 2 could the race/ethnicity category for members be ascertained from administrative databases; therefore, the sample pool included 60% of the invitees from the African American strata and 40% from "all others"; both categories were stratified by gender. Approximately 32% of the Site 5 enrollment population was African American, and two of its medical clinics had an estimated African American population of 90%. Site 5 oversampled for African Americans by pulling 61% of its sample from these two predominately African American clinics. Part way through recruitment, in response to an overall lower enrollment rate, this site added an additional 887 invitees, using the same oversampling methodology, to the recruitment pool. Approximately 16% of Site 4's population was Hispanic. This site weighted 50% of its total recruitment by a probable Hispanic indicator. Probable Hispanics were identified with a Latino surname algorithm based on the Passel-Word Spanish surname list utilized by the 1990 Census [15-18] (also Carroll NM et al, unpublished data, 2009).

Table 1. Characteristics of the entire membership of participating MENU health plans

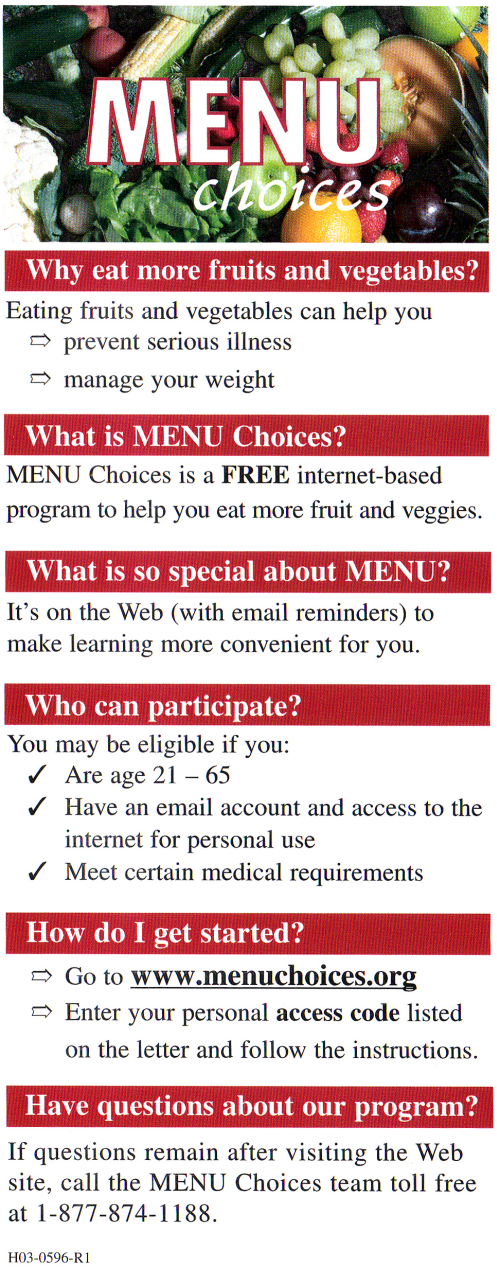
	Site 1	Site 2[a]	Site 3	Site 4	Site 5
Total enrollment (× 1000)	480	250	669	373	271
Age (years), %					
≤ 24	35	39	37	34	32
25-44	32	35	39	28	33
45-64	22	20	22	26	28
65-74	12	6	1	9	4
≥ 75	11	–	1	6	2
Female, %	53	53	52	51	53
Race, %					
White	89	60	85	75	58
African American	4	35	6	6	32
Asian American	5	< 1	5	2	4
Native American	1	< 1	1	1	< 1
Hispanic	1	< 1	3	16	3
Other	0	2	0	1	2
Web access, %	75	73	83	78	79

[a] Includes only members assigned to staff model medical group (physicians who are employed by the health plan).

Invitees were mailed a single introductory letter from each respective site that described the study, was signed by each health plan's respective investigator, and was postmarked locally using metered postage. This local focus was intended to increase members' confidence in the study invitation [19,20]. Study-related information, a unique log-in ID, and the name and telephone number of the health plan's contact person were included in the letter. The letter was accompanied by a colorful 3-inch by 8-inch flyer (Figure 1) inviting recipients to participate in the study and a pre-paid monetary incentive [8]. Both the letter and flyer invited the recipient to "check out" the website and provided the MENU study's URL and toll-free telephone number. An incentive combination was used based on results from previous work investigating optimal incentives for Web-based health program participation and retention [8]. Each introductory letter contained a US$2 bill as an enrollment incentive and described the US$20 promised incentive that would be sent for completion of each of the three online

follow-up surveys, over the course of 1 year. All study materials were provided in English only. Since the population at Site 4 may have included a large number of non-English-proficient invitees, the introductory letter from this site included Spanish text explaining that the MENU program was being offered in English only.

Figure 1. MENU flyer



Responders

Those interested in participating logged on to the study website and were asked to answer questions (pre-consent) and then to complete a short survey that confirmed eligibility status. The eligibility survey included 9 to 12 questions, depending on personal tailoring, and included questions to confirm health plan membership status, age, accessibility to the Internet for personal use, frequency of use of a personal email address, and history and treatment of certain health conditions. Eligibility was restricted to those who reported having access to the Internet for personal use, who had a working email account that they used at least once a week, and who did not have a health condition that conflicted with eating fruits and vegetables.

Responders found to be eligible were presented with an online informed consent form. The process asked members to read the consent information and click "I agree" after each page before being able to move to the next page of the consent form. The toll-free telephone help number and a "help" link appeared at the bottom of each page. The final screen provided the option to print a paper copy of the consent form.

Enrollees

Following completion of the consent process, the website prompted the responders to provide email and postal contact

information so that incentives and email reminders could be sent as part of the intervention protocol. The email address was verified through a reply email sent automatically from the study website server. Once the email address was verified, the online baseline survey was available for completion. It was comprised of approximately 70 questions and took approximately 25 minutes to complete. Participants were considered "enrolled" after they completed the survey. The Web programming allowed participants to complete the survey over several sessions, if interrupted. Responders were given 28 days to begin the baseline survey and 28 days to complete the survey once started. During this time, up to four automated email reminders were sent, one every three or four days, to persons with incomplete surveys. After the final 28 days expired, the survey was closed to those not completing the enrollment process. Enrollees were those participants completing all steps of enrollment.

Measures

Age, gender, and residential address were obtained from administrative databases for each invitee. We employed geocoding techniques to the residential address to create area-based proxies for income and education as socioeconomic variables for each invitee. Geocoding was performed using MapMarker Plus and 2000 Census data. Each participant's address was mapped to a census tract and the corresponding median household income and proportion of the census tract attaining various educational levels. Indicator variables were created for each invitee, with cut points at median household income and post–high school versus less educational levels. Race, ethnicity, and other demographic and socioeconomic variables were obtained by self-report from enrollees in the baseline survey.

Statistical Analysis

The numbers of invitees, responders, and enrollees were tabulated. Descriptive statistics were computed to characterize demographics and geocoded information for each group. Statistical significance of differences was tested using the Wilcoxon rank sum or the chi-square test. The protected least significant difference approach to multiple comparisons was used to compare enrollment rates among the sites.

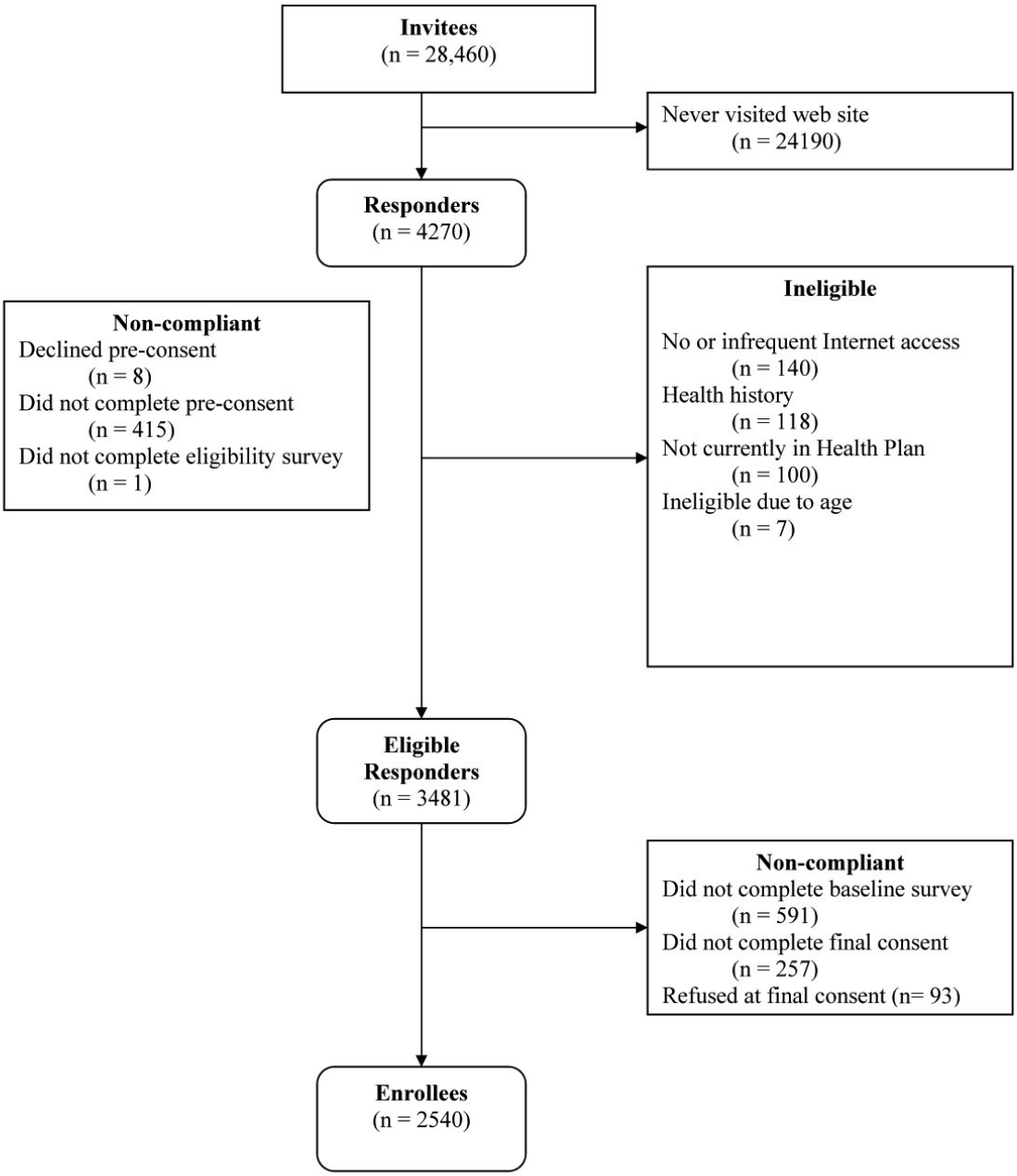
The associations between age, gender, and census-derived household income and education indicators, and between response and enrollment, were assessed using generalized estimating equation (GEE) multivariable logistic regression. GEE was used to take into account clusters defined by site. Customary residual and influential statistics were examined to assess model fit and evaluate outliers. Analyses were conducted using SAS 9.1 [21].

Results

Characteristics are presented for each participating health plan population (Table 1). Across the five study sites, the estimated proportion of African Americans ranged from 4% to 35%, with an overall average of 12.5%. The proportion of Hispanic members ranged from < 1% to 16%, with an overall average of 4.5%.

Figure 2 indicates the flow of participants from invitation to enrollment. A total of 28,460 adults from each of the five health plans were invited to participate; 15% (n = 4270) responded to the introductory letter, visited the website, and entered their unique log-in ID (responders). Of these, 18.5% (n = 789) failed to complete the enrollment process. Reasons for dropoff included not completing the pre-consent or eligibility questions (n = 424, 54%), ineligibility due to no or infrequent Internet access (n = 140, 17.7%), ineligibility due to medical history (n = 118, 15.0%), or ineligibility due to not being currently enrolled in the health plan (n = 100, 12.7%). Out of the eligible responders (n = 3481), a total of 941 (27.0%) did not enroll. Additional details on the reasons for dropout are provided in Figure 2. In total, 2540 (8.9% of invitees) enrolled.

Figure 2. Recruitment flowchart for MENU



Characteristics of the invitees, responders, and enrollees derived from geocoding, administrative data, or self-report are summarized in Table 2. While 47.8% of the invitees lived in a census area defined as having greater than 63% completing high school, 56.3% of the responders and 58.7% of the enrollees lived in such areas. Similarly, 40.1% of invitees mapped to a census area median household income of at least US$52,250, and half of the responders and enrollees mapped to this higher income level. Comparing those enrolled to those invited, enrollees were more likely to be from households with higher

education (10.9% vs 7%, $P < .001$) and income (11.1% vs 7.4%, $P < .001$). Women represented 50.2% of invitees and made up 61.4% of responders and 64.8% of enrollees. On average, responders were slightly older than invitees and enrollees (46.0 years vs 44.3 years and 45.5 years, respectively.) The percentage of enrollees compared to invitees varied by site, with the highest enrollment seen in Site 1 (13.1%, $P < .001$). Finally, the enrollees were comprised of 22.3% (n = 566) African Americans and 7.6% (n = 192) Hispanics.

Table 2. Descriptive characteristics of invitees, responders, and enrollees in MENU[a]

Characteristic	Invitees (n = 28,460) No. (%)	Responders (n = 4270) No. (%)	Enrollees (n = 2540) No. (%)	Enrolled/ Invited[c] %
Census-Derived Data [b]				
Education				
< 63% high school or greater	13,622 (52.2)	1705 (43.7)	957 (41.3)	7
≥ 63% high school or greater	12,461 (47.8)	2195 (56.3)	1360 (58.7)	10.9
Median household income				
<US$52,250	15,621 (59.9)	1994 (51.1)	1154 (49.8)	7.4
≥ US$52,250	10,462 (40.1)	1906 (48.9)	1163 (50.2)	11.1
Administrative Data				
Gender				
Female	14,298 (50.2)	2621 (61.4)	1645 (64.8)	11.5
Male	14,162 (49.7)	1649 (38.6)	895 (35.2)	6.3
Health plan site				
Site 1	4053 (14.2)	845 (19.8)	532 (20.9)	13.1
Site 2	6372 (22.4)	930 (21.8)	520 (20.5)	8.2
Site 3	4332 (15.2)	712 (16.7)	456 (18.0)	10.5
Site 4	5751 (20.2)	805 (18.9)	514 (20.2)	8.9
Site 5	7952 (27.9)	978 (22.9)	518 (20.4)	6.5
Survey Data				
Race				
White/other	--	--	1935 (77.4)	--
African American	--	--	566 (22.6)	--
Hispanic ethnicity				
Yes	--	--	192 (7.6)	--
No	--	--	2325 (92.4)	--
Highest level of education				
≤ High school or vocational tech	--	--	397 (15.7)	--
Some college	--	--	855 (33.8)	--
College degree	--	--	663 (26.2)	--
Post-grad	--	--	617 (24.4)	--

[a] Some variables had missing data; thus, numbers may not equal total.

[b] A total of 4274 (15.0%) invitees, 564 (13.2%) responders, and 333 (13.1%) enrollees had addresses that could not be mapped to the census data.

[c] Enrollee versus invitee associations were all statistically significant ($P < .001$).

Table 3 presents results from a GEE logistic regression model that was used to evaluate the association between invitee age, gender, and indicator variables for census area median household income and education and visiting the MENU website and entering the study ID. Age, gender, and census area household income and education were all statistically significantly

associated with being a responder. For every decade increase in age, a 0.89 decrease in odds of responding was observed (95% CI 0.82-0.96, $P = .002$).

Table 3. Adjusted odds ratios for predicting a response to the MENU letter among invitees by age, gender, and census area income and education

Variable	Odds Ratio[a]	95% CI	*P*
Age (decade)	.89	.82, .96	.002
Female	1.40	1.19, 1.63	< .001
Higher census area income	1.11	1.01, 1.23	.04
Higher census area education	1.17	1.05, 1.29	.004

[a] All odds ratios adjusted for other variables in table.

Table 4 describes the results from a GEE logistic regression model predicting enrollment. All characteristics included in the model were statistically significantly associated with enrollment status, with female gender having the strongest association (95% CI 1.63-2.16, $P < .001$).

Table 4. Adjusted odds ratios for predicting enrollment of invitees by age, gender, and census area income and education

Variable	Odds Ratio[a]	95% CI	*P*
Age (decades)	1.10	1.06, 1.13	< .001
Female	1.88	1.63, 2.16	< .001
Higher census area income	1.32	1.19, 1.46	< .001
Higher census area education	1.36	1.10, 1.68	.004

[a] All odds ratios adjusted for other variables in table.

We did not have race/ethnicity invitee data from most of the participating health plans, and therefore could not use them as a predictor of response or enrollment. Our attempts to oversample minority populations are shown in the enrollment in Table 5, reflecting the underlying health plan minority distribution. While the estimated proportion of African Americans varied across all five CRN health plans (see Table 1), oversampling resulted in the MENU study enrollees being comprised of 22.3% (n = 566) self-identified African Americans. Across all five sites, the proportion of Hispanics who enrolled was 7.6% (n = 192).

Table 5. Race and ethnicity of MENU enrollees

	Site 1 No. (%)	Site 2[a] No. (%)	Site 3 No. (%)	Site 4[a] No. (%)	Site 5[a] No. (%)	Total No. (%)
African American	8 (1.5)	230 (44.2)	13 (2.9)	14 (2.7)	301 (58.1)	566 (22.3)
White	469 (88.2)	248 (47.7)	413 (90.6)	369 (71.8)	169 (32.6)	1668 (65.7)
Other	47 (8.8)	41 (7.9)	29 (6.4)	110 (21.4)	40 (7.7)	267 (10.5)
Unknown	8 (1.5)	1 (0.2)	1 (0.2)	21 (4.1)	8 (1.5)	39 (1.5)
Hispanic	21 (4.0)	7 (1.4)	5 (1.1)	146 (28.4)	13 (2.5)	192 (7.6)
Non-Hispanic	506 (95.1)	508 (97.7)	447 (98.0)	367 (71.4)	497 (96.0)	2325 (91.5)
Unknown	5 (0.9)	5 (1.0)	4 (0.9)	1 (0.2)	8 (1.5)	23 (0.9)

[a] Sites oversampling for diverse populations: Site 2 and Site 5 oversampled for African Americans, and Site 4, for Hispanics.

Discussion

This study describes the results of a recruitment effort to enroll members of five health plans across the United States into a Web-based behavioral intervention trial. A significant mailing volume was needed to achieve our target enrollment, with a final enrollment rate of nearly 9% of invitees. This enrollment rate is consistent with other Web-based enrollment efforts [10]. Site enrollment rates varied, with overall sampling yielding a diverse group of men and women and minorities. Similar to findings of Oenema et al [22], women made up two-thirds of the enrollees. In contrast to other findings, this health promotion program appealed to people equally across the education spectrum. Enrollees were equally divided between college completers, graduate school completers, and those with less than college graduate education. Older invitees were less likely to respond to the invitation to visit the website, but were more likely to enroll in the study. This suggests that younger invitees were open to investigating a Web-based study, but consistent with findings by Verheijden and colleagues [23], the older

responders actually took the step and enrolled into the Web-based behavior change program.

Of those who enrolled, there were significantly more women; enrollees were generally older and non-Hispanic white and resided in census areas of higher educational and income levels. The enrollment of more women than men was consistent with higher female enrollment in other dietary intervention programs. Women tend to be the shoppers and food preparers of families [11-13,24].

By oversampling minority health plan members at three of the five sites, we were able to enroll a diverse cohort that was over 22% African American and almost 8% Hispanic. While the proportion of overall minority enrollees varied by site, oversampling doubled the proportion of participating African American and Hispanic members relative to their underlying populations, as noted in Table 1, improving the generalizability of the final study outcomes.

Almost 85% of invitees never investigated the website in response to the invitation letter. Efforts were made to encourage letter opening and reduce the appearance of "junk mail," including using a more business-style envelope with metered postage and a recognizable affiliation (HMO) in the return address [25]. These results may partially reflect self-selection since eligibility requirements were included in the invitation. These low recruitment numbers may reflect a lack of awareness of inadequate dietary patterns [22] and perceived difficulty of or disinterest in increasing fruit and vegetable consumption [26].

Given the large number of invitees who did not even view the website, we assume that there were significant barriers to study enrollment. The 12-month, longitudinal MENU study might have imposed too much time burden on invitees. Additionally, a proportion of invitees probably did not have convenient access to both the Internet and an email account on a weekly basis, even though the number of Americans using the Internet has grown to 79% [27]. Invitees were initially given a US$2 bill in the introduction letter and told that they would receive a higher dollar amount after completion of each of the 3-, 6-, and 12-month surveys [8]. It is possible that these invitees needed some further incentive to continue on to enrollment, including availability of a brief portion of the website program or limited access to it.

Nearly one-third of those presumably eligible who visited the website to read about the study decided not to pursue participation. A respondent viewing the website was met with requests to complete an eligibility survey, give consent, set up a contact account (which included providing home and email addresses and telephone number), and respond to emailed requests to complete a lengthy baseline survey. Completing the lengthy consent and the baseline survey was required for enrollment and access to the study's intervention website. The consent process required by the Research Ethics Board presented full details on expectations and time frames for participation in this 12-month study. The consent form covered several pages, and skipping pages to the end was not possible. While we simplified the consent form to meet the 7th grade reading level recommendation and used bullets and numbering to aid reading, this Web-based consent "contract" might have proved daunting, particularly to those who had been merely curious about the program requirements and interested in the offered incentives. Future studies need to account for this barrier and consider the enrollment completion rate in determining invitee sample size. In addition, the baseline survey that was also required for enrollment was quite lengthy and could have been a barrier to completing enrollment.

Limitations and Strengths of the Study

Limitations of this study include relatively limited knowledge of the individual characteristics of the target population. Further, the geocoding yielded low resolution of income and education. There are other variables associated with likelihood of enrolling, for example race and ethnicity, which we could not or did not assess. Practical human subject limitations precluded contacting nonrespondents to elucidate further reasons for not enrolling.

Strengths of this study include a large and diverse target population representing five geographic regions and oversampling of minority members. We recruited from a known sample of potential participants, used health system administrative data to identify age and gender of invitees, and used personal ID access codes to track the Web sign-on information for respondents. Another strength includes our ability to measure response to an online health promotion intervention program, acknowledging the relatively low numbers of those programs currently available [28]. Also, by stratifying invitees by gender, we were able to measure the response rate of men and women.

Conclusions

Web-based interventions have vast potential to reach virtually anyone with Internet access [29], either through work or personal computers. While only 9% of those invited went on to enroll in MENU, we demonstrated that a mailed invitation letter and online enrollment provided a efficient and relativity successful mode to recruit diverse participants to a Web-based health promotion study. The majority of our enrollees were women, with proxy variables demonstrating relatively high household socioeconomic status. However, through oversampling, minority members enrolled at a higher rate than the membership percentage across the combined health plans. More research and reporting of response and enrollment rates in Web-based research, defining various age, gender, racial and ethnic groups, is needed to enhance the recruitment to eHealth interventions in order to expand participation across diverse populations. Such reporting for Web-based programs designed for other lifestyle interventions to prevent and treat cancer and chronic diseases would greatly enhance future utilization of this medium by all racial and ethnic populations.

Acknowledgments

National Cancer Institute (NCI) grant U19 CA079689 supported this research, which we performed in affiliation with the Cancer Research Network (CRN).

We acknowledge and thank all site mangers, the MENU teams, the University of Michigan, and Judy Mouchawar, MD.

Conflicts of Interest

None declared.

References

1. Ritterband LM, Gonder-Frederick LA, Cox DJ, Clifton AD, West RW, Borowitz SM. Internet interventions: In review, in use, and into the future. Prof Psychol: Research and Practice 2003;34(5):527-537. [doi: 10.1037/0735-7028.34.5.527]
2. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006;8(2):e10 [FREE Full text] [Medline: 16867965] [doi: 10.2196/jmir.8.2.e10]
3. Noah SA, Abdullah SN, Shahar S, Abdul-Hamid H, Khairudin N, Yusoff M, et al. DietPal: a Web-based dietary menu-generating and management system. J Med Internet Res 2004 Jan 30;6(1):e4 [FREE Full text] [Medline: 15111270] [doi: 10.2196/jmir.6.1.e4]
4. ; Office of Disease Prevention and Health Promotion. Communicating Health: Priorities and Strategies for Progress. Action Plan to Achieve the Health Communication Objectives in Healthy People 2010. Washington, DC: U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion; 2003. URL: http://odphp.osophs.dhhs.gov/projects/HealthComm/ [accessed 2009-01-30] [WebCite Cache ID 5j1UubLdX]
5. McClure JB, Greene SM, Wiese C, Johnson KE, Alexander G, Strecher V. Interest in an online smoking cessation program and effective recruitment strategies: results from Project Quit. J Med Internet Res 2006;8(3):e14 [FREE Full text] [Medline: 16954124] [doi: 10.2196/jmir.8.3.e14]
6. Joinson AN, Ulf-Dietrick R. Personalized salutation, power of sender and response rates to Web-based surveys. Computers in Human Behavior 2007;23(3):1372-1383. [doi: 10.1016/j.chb.2004.12.011]
7. Dillman DA. Survey Implementation. New York: John Wiley and Sons; 2000.
8. Alexander GL, Divine GW, Couper MP. Effect of incentives and mailing features on online health program enrollment. Am J Prev Med 2008;34(5):382-388. [doi: 10.1016/j.amepre.2008.01.028]
9. Alexander GL, McClure J, Calvi J. A randomized clinical trial evaluating online interventions to improve fruit and vegetable consumption. American Journal of Public Health. In press. .
10. Block G, Sternfeld B, Block CH. Development of Alive! (A Lifestyle Intervention via E-mail), and its effect on Health-related quality of life, presenteeism, and other behavioral outcomes: Randomized controlled trial. J Med Internet Res 2008;10(4):e43. [doi: 10.2196/jmir.1112]
11. O'Doherty Jensen K, Holm L. Preferences, quantities and concerns: Socio-cultural perspectives on the general consumption of foods. European Journal of Clinical Nutrition 1999;53(5):351-359. [doi: 10.1038/sj.ejcn.1600767]
12. Wardle J, Haase AM, Steptoe A, Nillapun M, Jonwutiwes K, Bellisle F. Gender differences in food choice: the contribution of health beliefs and dieting. Ann Behav Med 2004 Apr;27(2):107-116. [Medline: 15053018] [doi: 10.1207/s15324796abm2702_5]
13. Watters JL, Satia JA, Galanko JA. Associations of psychosocial factors with fruit and vegetable intake among African-Americans. Public Health Nutr 2007 Jul;10(7):701-711. [Medline: 17381950] [doi: 10.1017/S1368980007662284]
14. Wagner EH, Greene SM, Hart G. Building a research consortium of large health systems: the Cancer Research Network. J Natl Cancer Inst Monogr 2005;35(35):3-11. [doi: 10.1093/jncimonographs/lgi032]
15. Passel JS, Word DL. 1980;census An application of Bayes' Theorem. Presented at: Annual Meeting of the Population Association of America; April 10-12, 1980; Denver, Colorado.
16. Perkins R. Evaluating the Passel-Word Spanish Surname List. Washington, DC: Population Division, U.S. Bureau of the Census; 1993.
17. Word DL, Perkins RC J. Building a Spanish surname list for the 1990's—A new approach to an old problem. Washington, DC: U.S. Bureau of the Census; 1996. URL: http://www.census.gov/population/documentation/twpno13.pdf [accessed 2008-03-05] [WebCite Cache ID 5j1VAivBn]
18. Ritzwoller DP, Carroll N, Gaglio B. Variation in Hispanic self-identification, Spanish surname, and geocoding: implications for ethnicity data collection. Open Health Svs Pol J 2008;1(1):12-18. [doi: 10.2174/1874924000801010012]
19. Koo M, Skinner H. Challenges of internet recruitment: a case study with disappointing results. J Med Internet Res 2005;7(1):e6 [FREE Full text] [Medline: 15829478] [doi: 10.2196/jmir.7.1.e6]
20. Glasgow RE, Nelson CC, Kearney KA. Reach, engagement, and retention in an Internet-based weight loss program in a multi-site randomized controlled trial. J Med Internet Res 2007;9(2):e11. [doi: 10.2196/jmir.9.2.e11]
21. ; SAS Institute Inc.. 2004 SAS/STAT 9.1 Users Guide. Cary, NC: SAS Institute Inc.; 2004.

22. Oenema A, Brug J, Lechner L. Web-based tailored nutrition education: results of a randomized controlled trial. Health Educ Res 2001 Dec;16(6):647-660 [FREE Full text] [Medline: 11780705] [doi: 10.1093/her/16.6.647]
23. Verheijden MW, Jans MP, Hildebrandt VH, Hopman-Rock M. Rates and determinants of repeated participation in a web-based behavior change program for healthy body weight and healthy lifestyle. J Med Internet Res 2007;9(1):e1 [FREE Full text] [Medline: 17478410] [doi: 10.2196/jmir.9.1.e1]
24. Smith KS, Eubanks D, Petrik A, Stevens VJ. Using web-based screening to enhance efficiency of HMO clinical trial recruitment in women aged forty and older. Clin Trials 2007;4(1):102-105. [Medline: 17327250] [doi: 10.1177/1740774506075863]
25. Helgeson JG, Voss KE, Terpening WD. Determinants of mail-survey response: Survey design factors and respondent factors. Psychology and Marketing 2002;19(3):303-328. [doi: 10.1002/mar.1054]
26. Campbell MK, Reynolds KD, Havas EA. Stages of change for increasing fruit and vegetable consumption among adults and young adults participating in the National 5-A-Day for Better Health community studies. Health Educ Behav 1999;26(4):513-534. [doi: 10.1177/109019819902600409]
27. Internet activities. Pew Internet & American Life Project URL: http://www.pewinternet.org/Static-Pages/Trend-Data/Online-Activites-Total.aspx [accessed 2009-10-13] [WebCite Cache ID 5j1VNts0v]
28. Strecher V. Internet methods for delivering behavioral and health-related interventions (eHealth). Annu Rev Clin Psychol 2007;3(1):53-76. [Medline: 17716048] [doi: 10.1146/annurev.clinpsy.3.022806.091428]
29. McNeill LH, Viswanath K, Bennett GG, Puleo E, Emmons KM. Feasibility of using a web-based nutrition intervention among residents of multiethnic working-class neighborhoods. Prev Chronic Dis 2007 Jun 15;4(3):A55 [FREE Full text] [Medline: 17572959]

Abbreviations

CRN: Cancer Research Network
HMO: health maintenance organization
MENU: Making Effective Nutritional Choices

Edited by G Eysenbach; submitted 21.04.08; peer-reviewed by D Tate; comments to author 14.08.08; revised version received 19.05.09; accepted 17.06.09; published 26.08.09.

Please cite as:
Stopponi MA, Alexander GL, McClure JB, Carroll NM, Divine GW, Calvi JH, Rolnick SJ, Strecher VJ, Johnson CC, Ritzwoller DP Recruitment to a Randomized Web-Based Nutritional Intervention Trial: Characteristics of Participants Compared to Non-Participants J Med Internet Res 2009;11(3):e38
URL: http://www.jmir.org/2009/3/e38/
doi:10.2196/jmir.1086
PMID:19709990

Original Paper

What Is My Cancer Risk? How Internet-Based Cancer Risk Assessment Tools Communicate Individualized Risk Estimates to the Public: Content Analysis

Erika A Waters[1], PhD, MPH; Helen W Sullivan[1], PhD, MPH; Wendy Nelson[2], PhD; Bradford W Hesse[1], PhD

[1]Health Communication and Informatics Research Branch, National Cancer Institute

[2]Basic and Biobehavioral Research Branch, National Cancer Institute

Corresponding Author:
Erika A Waters, PhD, MPH
National Cancer Institute
Health Communication and Informatics Research Branch
Executive Plaza North Room 4051B
6130 Executive Blvd MSC 7365
Bethesda, MD 20892-7365
Tel: +1 301 496 8586
Fax: +1 301 480 2198
Email: erika.a.waters@gmail.com

Abstract

Background: Internet-based cancer risk assessment tools have the potential to inform the public about cancer risk and promote risk-reducing behaviors. However, poorly communicated information on these websites may result in unintended adverse health outcomes.

Objective: This study examined whether: (1) Internet-based cancer risk assessment tools use risk communication formats that facilitate comprehension and reduce bias (as identified by the empirical literature); (2) the use of these formats varies by website affiliation; and (3) the websites provided information necessary to evaluate the quality of the risk estimate.

Methods: A content analysis of Internet-based cancer risk assessment tools was conducted. The terms *calculate cancer risk*, *cancer risk calculator*, *estimate cancer risk*, *assess cancer risk*, and *cancer risk assessment* were searched using three search engines. We identified 47 risk assessment tools and coded each according to standardized criteria. We calculated simple frequencies on all coding categories and performed crosstabulations but did not conduct formal statistical analysis due to small cell sizes.

Results: Use of risk communication formats that facilitate comprehension and reduce bias varied widely (eg, 30% of websites [14/47] provided absolute and comparative risk information but 83% [39/47] provided safety messages). Use of formats that facilitate comprehension varied by website affiliation and communication strategy (eg, only 8.3% [1/12] websites affiliated with the health care industry provided absolute and comparative risk information, but 83% [5/6] of websites affiliated with a governmental organization did so). Only 53% (25/47) of websites provided information about the statistical model or the peer-reviewed literature that was used to calculate the risk estimate.

Conclusion: Internet-based cancer risk assessment tools varied in their use of risk communication formats that facilitate comprehension and reduce bias. Formats that are difficult to understand may cause people to misperceive their cancer risk and consequently take inappropriate action.

(J Med Internet Res 2009;11(3):e33) doi:10.2196/jmir.1222

KEYWORDS

Risk; risk assessment; communication; risk communication; perception; risk perception; calculators, programmable; risk calculator; Internet; online; cancer

Introduction

Many laypeople obtain individualized cancer risk estimates from Internet-based risk assessment tools. Data from the 2005 Health Information National Trends Survey (HINTS), which comprises a representative sample of the United States' adult population, reveal that approximately 25% of people have used the Internet to seek information about cancer [1]. Although the proportion of people who sought information about their cancer risk is not known, use of online cancer risk assessment tools is common. For example, the *Your Disease Risk* website averaged nearly 2000 visitors per day in 2006 (personal communication by Graham Colditz, 2006), and the National Cancer Institute's Breast Cancer Risk Assessment Tool averaged over 1200 hits per day in 2007 (personal communication by Rick Manrow, 2008). The purpose of providing people with individualized risk estimates is to encourage them to engage in health-promoting behaviors [2,3], such as using the health care system appropriately, making good medical decisions, engaging in health screening, avoiding tobacco use, engaging in physical activity, and eating a healthy diet. To achieve this, it is crucial that the information be presented in a way that facilitates comprehension and does not bias risk perceptions [4].

Gurmankin Levy and colleagues [5] examined the quality of risk estimates for 13 breast cancer risk assessment websites. Many of these risk assessment tools used different risk factors to calculate risk, and in some cases excluded well-established risk factors, such as age at first live birth. Consequently, it is not surprising that the risk estimates provided for a particular risk profile varied across assessment tools. These authors [5] also found that many Internet-based risk assessment tools did not provide sufficient information to evaluate website quality. For example, some websites did not contain information about institutional affiliations or identify the statistical model used to calculate risk [5,6]. Because millions of people seek cancer risk information from Internet sources [1], the public health implications of inaccurate or inadequate risk communication are clear. Risk assessment tools that provide incorrect information may lead people to experience negative outcomes, such as seeking too much or too little medical care.

Even when risk assessment tools provide accurate risk estimates, the information needs to be presented clearly. People do not always understand probabilistic information [7-11], even though they often report wanting to use it to make medical decisions [12]. The dilemma is exacerbated by the fact that risk perceptions, risk comprehension, and decision making can be affected by the ways in which probability information is presented [4,13-18]. For example, arrays of stick figures (also referred to as "pictographs" and "icon arrays") can make it easier for people to understand the effects of hypothetical medical treatments and can reduce the undue influence that side effects have on hypothetical treatment decisions [19]. Misperceiving one's risk of illness may result in poor health decisions. For example, a woman who overestimates her risk of developing cervical cancer might undergo excessive screening, but she might also be so overwhelmed with anxiety that she avoids screening entirely.

A number of researchers have tested various ways to communicate probabilistic risk information [9,14,16,20-22]. According to these studies, the optimal risk communication format depends on what the communication is expected to accomplish. For example, presenting a drug's benefit only as a relative risk reduction ("reduce the overall mortality by 20.3%") can be more effective in persuading physicians to prescribe a drug than presenting only absolute risk reduction information ("reduce overall mortality from 7.8%...to...6.3%") [23] (page 123). However, if the goal is education rather than persuasion, providing both absolute and relative information is most effective [24]. Thus, no single risk communication format will be appropriate in all situations and under all conditions.

If the goal of Internet-based risk assessment tools is to assist people in making good decisions about their health, they must communicate probabilistic risk information in a way that facilitates comprehension and reduces biased interpretations [4]. A large and growing literature on risk communication permits us to infer which formats are likely to be most useful in helping people understand cancer risk estimates (Table 1). However, it is unknown how these formats interact with each other to influence perceptions of risk. Consequently, Table 1 should not be viewed as a checklist of required criteria, but instead as a list of important factors to take into consideration when developing a tool. Just as readers would not use a statistical test without understanding its underlying principles, they should not develop a risk assessment tool without having some rudimentary understanding of how different risk communication formats might affect risk perceptions and comprehension. Comprehensive reviews of the complex issues surrounding probabilistic risk communication can be found at [21,22,24-27].

Table 1. Communication formats that reduce bias and facilitate comprehension of probabilistic risk estimates[a]

Risk Communication Format and Selected Relevant Citations	Why the Recommended Format Is Important When the Communication Goal Is to Educate and Inform
Describe the risk using words and numbers [20,22].	Using words only is ambiguous because people assign different numeric values to the same label (eg, "small" can mean "2%" to some people and "10%" to others). Using numbers only is problematic due to the population's low levels of numeracy (ie, the ability to use numeric information) and a lack of contextual information (eg, Should a 7% lifetime risk of breast cancer be considered a high risk or a low risk?).
Communicate numeric risk as N in 1000 or as a percentage [16,17,20,22].	Risk comprehension is highest when risk estimates are presented as a percentage or as N in 1000, compared to other formats like the number-needed-to-treat or odds ratios. However, both recommended formats have drawbacks. The N in 1000 format can encourage people to overemphasize risk by "imagining the numerator," but the percentage format is more difficult to use when conducting complex calculations (eg, the probability of a woman having breast cancer given a positive mammogram).
Provide absolute and comparative risk information [20,25,28-30], but see [24].	Providing both absolute and comparative information helps people determine the amount of importance that they should place on the risk and guides them in making informed decisions about their behavior. For example, telling a woman that she has a 5% 5-year risk of developing breast cancer might not be meaningful unless she recognizes that this means that she is at above average risk. However, telling people only that they are at below-average risk might reduce motivation to engage in preventive behavior.
Compare cancer risk to the risk of other hazards [22].	Helping people understand where their risk of cancer falls in relation to other hazards such as heart disease, being struck by lightening, and being in a car accident allows them to place the risk in context and thereby help them determine where to invest their limited time, energy, and economic resources.
Frame the risk in positive and negative terms [18,20,25].	Framing the risk in negative terms only (eg, "Your risk of cancer is 5%") places focus only on the negative outcome and might result in exaggerated risk perceptions. Adding positive framing (eg, "This means you have a 95% chance of not getting cancer) helps participants place the risk in context.
Specify the duration of risk [20,25].	Specifying whether the risk estimate is applicable to the next 5 years, 10 years, or over the visitor's lifetime is essential to help them place the risk in context and determine how much they should be concerned about the event. For example, a 7% risk of breast cancer would be more worrisome if it was applicable to the next 5 years than over one's lifetime.
Provide safety messages and risk reduction strategies [31-33].	Informing people how to reduce their risk is an essential component of risk communication messages, particularly for individuals who have not learned risk reduction strategies previously. Providing risk information without such safety messages may undermine risk communication efforts by encouraging people to control their fear (eg, by trying to ignore the risk) rather than encouraging people to control the danger (eg, by engaging in appropriate health behaviors).
Include a visual display of risk [20,22,26].	Using a visual display can increase comprehension of risk information. However, care must be taken to avoid biasing perceptions of risk (eg, displays that focus attention on the number of people affected by a disease can exaggerate a risk compared to displays that include information about the number of people affected and the number of people who are not affected).
Acknowledge that the risk estimate contains an element of uncertainty [22].	Individualized risk estimates are based on statistical modeling of population-level data. Consequently, they always contain a level of uncertainty. Informing the audience of this fact is essential to prevent them from attributing an unreasonable degree of certainty to the estimate.

[a]These formats can be implemented with varying levels of success and might not be equally effective in all situations. Additional examples of each format are located in Table 5.

This study describes the risk communication formats that Internet-based cancer risk assessment tools use to convey individualized risk information to the public. This study asks two questions that, to our knowledge, have not been addressed previously: (1) Do the tools use risk communication formats that have been empirically shown to facilitate comprehension, and (2) Does the use of these formats vary by website affiliation? Following Gurmankin Levy and colleagues [5], we also examined whether the websites provided the basic information necessary to evaluate the quality of the risk estimate (ie, the statistical model or peer-reviewed citations).

Methods

Overview

During October 2007 we conducted an Internet search to identify websites that provided individualized cancer risk assessment. The search was conducted by entering the search terms *calculate cancer risk, cancer risk calculator, estimate cancer risk, assess cancer risk,* and *cancer risk assessment* into the Google, MSN, and Yahoo! search engines. These search engines accounted for 82.6% of all Internet searches that originated in the United States in 2006 [34]. To locate the Internet-based cancer risk assessment tools, a total of 1500 websites were examined (ie, the first 100 search results for each of the five search terms, for each of the three search engines).

Out of the 1500 websites examined, we identified 51 websites that gave specific cancer risk estimates. Forty-four of the identified websites were unique interactive websites that provided individualized cancer risk estimates. These websites required visitors to enter information about their status on several cancer risk factors. Seven of the identified websites were non-interactive; they stratified risk information by two or three variables (eg, lung cancer risk by smoking status and gender). These non-interactive websites were included because they provided more specific risk information than general population data. Of these 51 websites, four were excluded because they required information seekers to download a software program

(n = 2), provide a mailing address (n = 1), or provide payment (n = 1) before obtaining results. A total of 47 websites were evaluated (Table 2).

Table 2. Websites hosting cancer risk assessment tools (WebCite® links are listed below the original URL)

1.	Breast Link http://www.breastlink.com/default.aspx Archived by WebCite® at http://www.webcitation.org/5gBFlHdu9
2.	CancerRiskInfo.com http://www.cancerriskinfo.com/ Archived by WebCite® at http://www.webcitation.org/5g6prLoiq
3.	Carefirst Blue Cross Blue Shield http://carefirst.staywellsolutionsonline.com/RelatedItems/42,BreastCancerRisk Archived by WebCite® at http://www.webcitation.org/5gBHb6xdn
4.	Center for Cancer Quality Assurance and Professional Education http://qap.sdsu.edu/screening/breastcancer/bda/flowcharts/risk_algo1.html Archival by WebCite® prohibited by website.
5.	Claxton Hepburn Medical Center http://www.chmed.org/breastca.htm Archived by WebCite® at http://www.webcitation.org/5gBHInllZ
6.	Cornell University http://envirocancer.cornell.edu/factsheet/diet/fs49.BCRisk.cfm Archived by WebCite® at http://www.webcitation.org/5gBGdq722
7.	Dermatology Imaging Center http://www.dermatologyimaging.com/skincancertest.html Archived by WebCite® at http://www.webcitation.org/5gBGrY1lI
8.	Divine http://www.divine.ca/en/breast-cancer-corner/breast-cancer-risk-calculator/c_244/ Archived by WebCite® at http://www.webcitation.org/5gBHOBQOw
9.	Dr. Halls MD http://www.halls.md/breast/risk.htm Archived by WebCite® at http://www.webcitation.org/5g6orRIOY
10.	EBSCO Publishing http://calculators.epnet.com/?docid=healthcalculators/breastcancer/precalcdoc&token=b0c3eb60-99e5-4038-bc04-819fded5c1d6&DeliveryContext=healthlibrary&CollectionIID=446&frame=&rooturl= Archived by WebCite® at http://www.webcitation.org/5gBIDVLXM
11.	Estronaut.com http://www.estronaut.com/a/breastInteractive2.htm Archived by WebCite® at http://www.webcitation.org/5g6pLGN9G
12.	Fairview Health Services http://www.fairview.org/staywell/assess_load.aspx?ContentTypeId=42&ContentId=OvarianCancerRisk Archived by WebCite® at http://www.webcitation.org/5gBHTLDA0
13.	Fred Hutchinson Cancer Research Center http://www.compass.fhcrc.org/edrnnci/bin/calculator/main.asp?t=prostate&sub=disclaimer&v=prostate&m=&x=Prostate%20Cancer Clicking this link now redirects visitors to the University of Texas Science Center in San Antonio, Texas: http://deb.uthscsa.edu/URORiskCalc/Pages/uroriskcalc.jsp Archived by WebCite® at http://www.webcitation.org/5gBGIO1yJ
14.	Hotflash! Menopause Matters http://www.families-first.com/hotflash/news/mayoquiz6.htm Archived by WebCite® at http://www.webcitation.org/5g6s0slTC
15.	Imaginis http://imaginis.com/breasthealth/bc_risks2.asp Archived by WebCite® at http://www.webcitation.org/5gBGVk6Tl

16. iVillage.com
http://cancer.health.ivillage.com/tools/assessment_index.cfm
Archived by WebCite® at http://www.webcitation.org/5g6pJnqwO

17. Little Company of Mary Hospital and Health Care Centers
http://pursuingpainfreecancer.com/breast_test.php
Archived by WebCite® at http://www.webcitation.org/5gBHvAIVP

18. McGill University
http://www.mcgill.ca/cancerepi/society/calculate/
Archived by WebCite® at http://www.webcitation.org/5g6p4wfuN

19. MD Anderson Cancer Center
http://www2.mdanderson.org/app/risk/
Archived by WebCite® at http://www.webcitation.org/5gBH8xzEu

20. Merck and Co.
https://www.merck.healthinkonline.com/merckTools/AssessMerckSourceBreastCancer.asp
Archived by WebCite® at http://www.webcitation.org/5gBFra58Q

21. Memorial Sloan Kettering Cancer Center
http://www.mskcc.org/mskcc/html/12463.cfm
Archived by WebCite® at http://www.webcitation.org/5g6oz4wBH

22. Men's Health Forum
http://www.malehealth.co.uk/userpage1.cfm?item_id=117&pop=326
Archived by WebCite® at http://www.webcitation.org/5g6q64w6Y

23. Mesoblog
http://www.mesoblog.org/risk-calculator.php
Archived by WebCite® at http://www.webcitation.org/5gBHQNRgC

24. National Breast and Ovarian Cancer Centre
http://www.nbcc.org.au/risk/yourrisk.html
Archived by WebCite® at http://www.webcitation.org/5gBHLKXwC

25. National Surgical Adjuvant Breast and Bowel Project
http://www.breastcancerprevention.org/raf_source.asp
Archived by WebCite® at http://www.webcitation.org/5g6p0aimY

26. Northeast Health Systems
http://www.nhshealth.org/index.cfm?Action=Education.BreastCancerQuiz
Original link no longer valid. Non-interactive version of the site can be found at:
http://web.archive.org/web/20070826153816/http://www.nhshealth.org/index.cfm?Action=Education.BreastCancerQuiz

27. Norton Healthcare
http://norton.convergencehealth.com/DesktopDefault.aspx?tabid=820&id=653
Archival by WebCite® prohibited by website.

28. Ohio State University Medical Center
http://www.jamesline.com/patientsandvisitors/prevention/cancergenetics/?ref=medicalnews
Archived by WebCite® at http://www.webcitation.org/5gBHhJ5LK

29. Penn State Hershey Cancer Institute
http://www.hmc.psu.edu/cancer/outreach_education/community/cancer_risk_assessments/cancer_risk_assessment.htm
Link no longer valid; WebCite® citation unavailable.

30. Prostate Cancer Research Foundation of Canada
http://www.prostatecancer.ca/english/prostate_owners_manual/risk_factors/risk/
Archived by WebCite® at http://www.webcitation.org/5g6rslb1L

31. Radon Seal
http://www.radonseal.com/radon-health-risks.htm
Archived by WebCite® at http://www.webcitation.org/5gBFg7aJS

32.	Real Age http://www.realage.com/health_guides/BreastCancer/introduction.asp Archived by WebCite® at http://www.webcitation.org/5gBHjAMpu
33.	Shannon Health http://shannon.convergencehealth.com/DesktopDefault.aspx?tabid=1390&id=644 Archival by WebCite® prohibited by website.
34.	Siteman Cancer Center (not the Your Disease Risk website) http://www.siteman.wustl.edu/crc.aspx?id=459 Archived by WebCite® at http://www.webcitation.org/5gBGnKmVx
35.	St. John's Hospital https://www.healthawareservices.com/nahrs/index.htm?hospID=19&moduleName=lungAware Archived by WebCite® at http://www.webcitation.org/5gBHqIAQU
36.	Susan Love http://www.susanlovemd.com/breastcancer/content.asp?L2=2&L3=2&SID=140 Archived by WebCite® at http://www.webcitation.org/5gBGc9yJy
37.	Urology Channel http://www.urologychannel.com/HealthProfiler/healthpro_psaageRace.shtml Archived by WebCite® at http://www.webcitation.org/5g6rqRUJo
38.	US Environmental Protection Agency http://www.epa.gov/radon/risk_assessment.html Archived by WebCite® at http://www.webcitation.org/5g6s2DYZK
39.	US National Cancer Institute (Breast Cancer) http://www.cancer.gov/bcrisktool/ Archived by WebCite® at http://www.webcitation.org/5g6ouFL9F
40.	US National Cancer Institute (Melanoma) http://www.cancer.gov/melanomarisktool/ Archived by WebCite® at http://www.webcitation.org/5g6owaI5h
41.	US National Cancer Institute (Thyroid Cancer) http://ntsi131.nci.nih.gov/ Archived by WebCite® at http://www.webcitation.org/5g6p2gkq6
42.	US National Cancer Institute and the Centers for Disease Control http://www.smokefree.gov/smokersrisk/index.asp Archived by WebCite® at http://www.webcitation.org/5g6pI9GAQ
43.	Vizilite http://www.vizilite.com/patient_site/risk_assessment/ Original link no longer valid. Non-interactive version of the site can be found at: http://web.archive.org/web/20080105051853/http://www.vizilite.com/patient_site/risk_assessment/
44.	Women's Cancer Network http://www.wcn.org/interior.cfm?diseaseid=13&featureid=3 Original link no longer valid. Non-interactive version of the site can be found at: http://web.archive.org/web/20071228025221/http://www.wcn.org/interior.cfm?diseaseid=13&featureid=3
45.	World Information Service on Energy http://www.wise-uranium.org/rdcum.html Archived by WebCite® at http://www.webcitation.org/5g6q2Y2Gf
46.	Wyoming Valley Healthcare System http://www.wvhc.staywellsolutionsonline.com/InteractiveTools/RiskAssessments/42,BreastCancerRisk Archived by WebCite® at http://www.webcitation.org/5g6s7Apal
47.	Your Disease Risk http://www.yourdiseaserisk.org/ Archived by WebCite® at http://www.webcitation.org/5epfwmhdn

Standardized coding criteria were developed (Table 3) and each website was coded independently by two of the authors (EW and HS). The first 15 websites were used to calibrate the coding procedure, and inconsistencies in the remaining 32 websites were resolved through discussion. Inter-rater reliability was high ($\kappa = 86.4$). If information about the statistical model used to calculate the risk could not be found on the website, EW attempted to obtain it by contacting the website's developer. Two attempts to reach the developers were made over a four-week period.

Table 3. Website coding criteria

Coding Category	Example
General website characteristics	
Organ site	Breast
Type of affiliation	Educational institution
Accessibility to lay audiences	
Intended audience	Lay people
Contains undefined terminology	Biopsy
Non-English version	Spanish
Risk communication strategies	
Words	"Your risk is low."
Numbers	"Your risk is 2%."
Format of numeric information	Percent, frequency (n in 1000), frequency (1 in N)
Absolute risk	"Your risk is low." OR "Your risk is 2%."
Comparative risk (other people)	"Your risk is higher than average."
Comparative risk (other hazards)	"Your risk of getting cancer is 12%. The risk of being injured in a car accident is 10%."
Positive framing	"Your risk is 2 in 100. Your chances of not getting cancer are 98 in 100."
Duration of risk	"Your 5-year risk is..."
Safety message/Risk reduction strategy	"Stop smoking."
Visual display	Bar graph, Line graph, Table
Acknowledges uncertainty	"Just because you're at high risk doesn't mean you'll definitely get cancer." OR "This estimate is based on information obtained from the population and your actual risk might be different."
Quality evaluation elements	
Information about the statistical model	"This website uses the Gail Model."
Peer-reviewed citation	"Harvard Report on Cancer Prevention, Volume IV: Harvard Cancer Risk Index, Cancer Causes and Control, Volume 11:477-488, 2000."

Analyses

General website characteristics, risk communication formats, and the presence of information about the quality of the risk estimate were examined by calculating simple frequencies on all coding categories. The number and percentage of websites that used formats that had the most empirical support for facilitating comprehension and reducing bias was recorded (see Table 1 for a list of formats). Affiliation-based differences in the use of formats were examined using crosstabulations. Formal statistical analysis and significance testing was not possible due to small cell sizes.

Results

General Website Characteristics

The general characteristics of the websites are described in Table 4. There are two areas of particular interest. First, the three most common organ sites for which websites provided assessments (breast, lung, and colorectal cancer) coincide with the three leading causes of cancer mortality in the United States [35]. The second item of interest is the widespread use of technical language. Although laypeople were the intended audience for nearly all of the sites, an overwhelming majority did not define medical and technical terms such as "biopsy," "DCIS," "mastectomy," or "radon progeny" (Table 4).

Table 4. General website characteristics (N = 47)

Website Characteristic	Example	n	%
Organ site[a, b]			
Bladder		4	8.5
Breast		27	57.5
Cancer (general)		5	10.6
Cervical		8	17.0
Colorectal		10	21.3
Gastrointestinal		5	10.6
Kidney		5	10.6
Lung		12	25.5
Ovarian		9	19.2
Pancreatic		3	6.4
Prostate		9	19.2
Skin/Melanoma		6	12.8
Other	Thyroid	8	17.0
Website affiliation			
Government	National Cancer Institute	6	12.8
Educational institution[c]	McGill University	3	6.4
Cancer center[c]	Memorial Sloan-Kettering	8	17.0
Health care industry	CareFirst Blue Cross Blue Shield	12	25.5
Advocacy/non-profit	Women's Cancer Network	6	12.8
Health portal[d]	RealAge.com; Imaginis.com	5	10.6
Commercial industry	RadonSeal	3	6.4
Other/unspecified	Dr. Halls; EBSCO publishing	4	8.5
Accessibility to lay audiences[b]			
Intended audience: Lay people		42	89.4
Contains undefined terminology	Biopsy	39	83.0
Non-English version	Spanish	3	6.4

[a]Websites varied in the number of organ sites for which they provided risk assessments. Most provided assessments for only one cancer site, but others provided assessments for more than one organ site (between 1 and 14 additional organ sites, depending on the website).

[b]The total N in *organ site*, *quality evaluation elements*, and *accessibility to lay audiences* categories does not sum to 47 because the individual elements within each category were not mutually exclusive.

[c]Cancer centers are often located within educational institutions, but the objectives and methods of these two types of institutions might differ. For this reason, assessment tools that were developed by cancer centers that were affiliated with educational institutions were coded as cancer centers.

[d]Health portals are websites that contain information about a variety of medical conditions and/or health issues. WebMD.com [36] is an example of a health portal, although it did not host a cancer risk assessment tool at the time of the study.

Formats for Communicating Individualized Cancer Risk Estimates

In general, few websites used risk communication formats that facilitate comprehension of probabilistic information (Table 5). Few websites provided the risk estimate as numbers and words, or described how the information seeker's cancer risk compared to the risk of experiencing other hazards. Only three websites framed the risk positively (eg, "998 chances in 1000 that you will not develop cancer"), and slightly less than half informed participants of the duration of the risk estimate (eg, 5-year risk). Approximately one-third of the websites provided a visual display. However, some risk communication formats were used widely. Seventeen of the 21 websites that provided any numeric information did so using the percentage format, six used the "N in 1000" format, and three used both. Twenty-four websites compared the information seeker's cancer risk to other people's risk, and fourteen of these also provided the absolute risk. An overwhelming majority of the websites provided information seekers with safety messages like "stop smoking." Slightly more

than half of the websites made at least one statement acknowledging that the estimate contained some degree of uncertainty.

Table 5. Risk communication formats used by Internet-based cancer risk assessment tools to communicate individualized risk estimates (N = 47)

Risk Communication Format	Example	n	%
Words or numbers			
Words only	Your risk is low.	24	51.1
Numbers only	Your risk is 2%.	16	34.0
Both [a]	*Your risk is 2%. This is a low risk.*	*5*	*10.6*
Neither	You may only need to continue screening.	2	4.3
Type of numeric information[b, c]			
Percent	*Your risk is 2%.*	*17*	*81.0*
Frequency (n in 1000)	*Your risk is 20 in 1000.*	*6*	*28.6*
Frequency (1 in N)	Your risk is 10 in 500.	4	19.1
Relative risk ratio	Your risk is 2 times higher than average.	2	9.5
Odds	Your odds of getting cancer are 2:98.	1	4.8
Risk estimate as absolute or comparative information			
Absolute risk only	Your risk is low." OR "Your risk is 2%.	21	44.7
Comparative risk only	Your risk is higher than average.	10	21.3
Absolute and comparative risk	*Your risk is 2%. This is below average. OR Your risk is 2%. The average risk is 3%.*	*14*	*29.8*
Neither absolute nor comparative risk	You may only need to continue screening.	2	4.3
Types of comparative risk information			
Compared to other people only (not hazards)	*Your risk is higher than average.*	*21*	*44.7*
Compared to other people and hazards	*Your risk of getting cancer is 12%, which is higher than average. The risk of being injured in a car accident is 10%.*	*3*	*6.4*
No comparison information		23	48.9
Contextual information[b]			
Positive framing	*Your risk is 2 in 100. This means your chances of not getting cancer are 98 in 100.*	*3*	*6.4*
Duration of risk	*Your 5-year risk is...*	*23*	*48.9*
Safety messages			
At least one	*Stop smoking*	*39*	*83.0*
None		8	17.0
Visual display			
At least one	*Bar graph, line graph, table*	*18*	*38.3*
None		29	61.7
Acknowledgment of uncertainty: Estimate is...[b]			
only an estimate	Your actual risk might be different.	14	29.8
probabilistic	High risk doesn't mean you'll get cancer.	15	31.9
based on population	This estimate is based on data from large clinical trials.	8	17.0
Any acknowledgment		*25*	*53.2*

[a]In general, the formats printed in italics are associated with increased comprehension and reduced bias of risk information. For comprehensive reviews see [21,22,24-27].

[b]The individual elements within the categories *type of numeric information, additional information, and acknowledgment of uncertainty* were not mutually exclusive.

[c]This category is restricted to the 21 websites that provided numeric risk information.

The use of risk communication formats that facilitate comprehension and reduce bias varied by website affiliation (Table 6). For example, websites affiliated with the health care industry were the least likely to communicate risk as percentages or as frequencies (1 of 12 websites) and to provide both absolute and comparative risk information (1 of 12 websites). Websites affiliated with cancer centers were the least likely to provide information about the duration of the risk estimate (1 of 8 websites) and to make at least one statement acknowledging that the estimate contained uncertainty (1 of 8 websites). Health portals were the least likely to provide a safety message or risk reduction strategies (2 of 5 websites). As mentioned in the Analyses section, formal statistical analyses and significance testing could not be performed due to small cell sizes (ie, out of 72 possible cells, only eight contained five or more websites and fifteen included no websites; see Table 6).

Table 6. Website affiliation-based variations in the use of risk communication formats that facilitate comprehension and reduce bias (N = 47)

Supported Format	---------- Affiliation ----------							
	Government (N = 6) n (%)	Educational Institution (N = 3) n (%)	Cancer Center (N = 8) n (%)	Health Care Industry (N = 12) n (%)	Advocacy/ Non-profit (N = 6) n (%)	Health Portal (N = 5) n (%)	Commercial (N = 3) n (%)	Other (N = 4) n (%)
Risk estimate as numbers and words[a]	1 (16.7)	0 (0.0)	0 (0.0)	1 (8.3)	0 (0.0)	1 (20.0)	0 (0.0)	2 (50.0)
Risk estimate as percent or N in 1000[b]	5 (83.3)	2 (66.7)	2 (25.0)	1 (8.3)	2 (33.3)	4 (80.0)	1 (33.3)	3 (75.5)
Absolute and comparative risk information	5 (83.3)	2 (66.7)	1 (12.5)	1 (8.3)	1 (16.7)	2 (40.0)	1 (33.3)	1 (25.0)
Risk compared to other hazards	0 (0.0)	1 (33.3)	0 (0.0)	0 (0.0)	0 (0.0)	1 (20.0)	1 (33.3)	0 (0.0)
Positive framing	2 (33.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (20.0)	0 (0.0)	0 (0.0)
Duration of risk	6 (100.0)	2 (66.7)	1 (12.5)	4 (33.3)	2 (33.3)	4 (80.0)	1 (33.3)	3 (75.0)
Safety messages	5 (83.3)	2 (66.7)	7 (87.5)	12 (100.0)	4 (66.7)	2 (40.0)	3 (100.0)	4 (100.0)
Visual display	3 (50.0)	2 (66.7)	2 (25.0)	3 (25.0)	2 (33.3)	3 (60.0)	2 (66.7)	1 (25.0)
Any acknowledgment of uncertainty	6 (100.0)	2 (66.7)	2 (25.0)	6 (50.0)	2 (33.3)	3 (60.0)	1 (33.3)	3 (75.0)

[a]Percentages are the percent of websites within a given affiliation that contain a particular element (eg, 1 of 6 websites affiliated with government agencies provided risk estimates as numbers and words).

[b]Includes websites that provided risk as numbers only and as numbers and words.

Information Necessary to Evaluate Website Quality

The number of websites that provided information to help visitors evaluate the quality of the risk estimate was limited. Only 25 of the 47 (53.2%) websites provided either information about the statistical model used to calculate the risk or peer-reviewed citations, where such information could be found. Only 13 websites (27.7%) provided both a description of the model and peer-reviewed citations. One of the authors (EW) attempted to contact all 22 of the websites that did not provide any information about the model. Of the 15 websites that provided a valid email address or phone number, only 8 responded to either the first or second inquiry, and only 2 were able to provide the information.

Discussion

Internet-based cancer risk assessment tools can provide cancer risk and prevention information to millions of people worldwide. Effective risk communication has the potential to reduce cancer morbidity and mortality by motivating people to engage in healthy behaviors. However, poorly communicated risk information could mislead people or frighten them unduly, resulting in maladaptive health behaviors [37], such as over- or under-utilization of health care services.

In October 2007, 47 Internet-based cancer risk assessment tools provided individualized cancer risk information to the public. Almost half (20 out of 47) of these websites were developed by cancer centers (n = 8) and the health care industry (n = 12). The three most common organ sites for which assessment tools

were available (breast, lung, colorectal) corresponded to the most common causes of cancer mortality among men and women in the United States [35]. This suggests that the assessment tools are responding to a public health need. However, three factors suggest that some of these websites might not be as useful to the public as they seem at first glance. First, many of the websites did not identify the statistical model used to calculate the risk estimate. Thus, visitors to these sites would have no way to verify that the risk assessment model had been scientifically vetted. The second factor that raises questions about the utility of some of these tools is the extensive use of undefined scientific terminology. Terms like "biopsy" and "radon progeny" may be confusing for individuals with limited literacy, thereby limiting their usefulness.

The third factor that raises concerns about these tools is the fact that they varied widely in their use of risk communication formats that facilitate comprehension of probabilistic information. For some risk communication formats, most websites followed experts' recommendations (eg, 39 of 47 sites provided safety messages). For other formats, few websites followed recommendations (eg, 3 of 47 sites used positive and negative framing). It is unclear whether it is necessary to include all of the risk communication formats identified in Table 1 in a single communication effort. Indeed, it is possible that doing so would overwhelm people with too much information and bias their risk perceptions [38]. However, risk perceptions can also be biased if people receive too little information. For example, seven websites provided information seekers with a simple count of risk factors or a "risk score." One such website then informed people, "You should do something to reduce your risk," but it did not provide specific safety messages, such as "stop smoking." Telling people to "do something" without providing specific recommendations does not facilitate comprehension [33] and might be counterproductive [32,37].

Two websites underscore the potential pitfalls of communicating risk using only numbers or only words. One of these websites provided two numerical risk estimates: one represented the risk if the information seeker continued the health-damaging behavior, and the other represented the risk if the individual stopped the behavior. However, the absolute risk was low in both cases (2% if the behavior continued and 1% if the behavior stopped), which might discourage people from changing their behavior [24,33,39]. The second website provided risk information as words only. It described cancer risk as "moderate," "increased," "high," and "highest." There was no "low risk" category. Thus, an 18-year-old with no risk factors for the type of cancer addressed on this website would be informed that he has a "moderate" risk of cancer. The combination of using only words to describe a risk and using the word "moderate" to describe the lowest-risk category might exaggerate risk perceptions among people at lowest risk [40].

It is also important to note that online risk assessment tools often have different objectives. For example, some risk assessment tools are designed to educate people about healthy lifestyles, whereas others are designed to persuade people to use their services, purchase their products, or engage in a particular health behavior. While it may be acceptable to attempt to persuade people to purchase necessary goods or services, or to engage in appropriate health behaviors, it is incumbent on developers of cancer risk assessment tools to consider the point at which persuasion may infringe upon an individual's ability to make an informed decision. It is beyond the scope of this study to assess the underlying motivations of the various risk assessment tool developers, but website developers, clinicians, researchers, and public health officials should consider the ethical implications of the tools they design.

Despite these concerns, cancer risk assessment tools have the potential to play an important role in cancer prevention and control. However, in order to be acceptable and effective, they should use language that is appropriate for low-literacy populations, communicate risk estimates using formats that increase comprehension and reduce biased interpretations of risk, and provide information about the model used to calculate the estimate.

As mentioned previously, the communication formats chosen will depend upon the goals of the communication. For example, if a developer's goal is to educate women about how the risk of breast cancer increases with age, the ideal website might use a line graph to portray how her risk changes over time [27]. It might describe in words how the risk changes and add the probability estimates at 5-year intervals. It might also include lines that depict the "average woman's" risk over the same time period and the risk of heart disease over the same time period. The website would include accurate information about how often women should undergo mammography screening, how to obtain a mammogram, and how to maintain a healthy weight. A statement would inform visitors that the risk estimate contains some uncertainty, and it would identify the statistical model that was used to generate the risk estimate. Finally, before the website is released to the public, it would be tested for readability and comprehensibility among people with low literacy and to ensure that it elicited relatively unbiased risk perceptions. If it was found to be confusing, the website developer might consider removing some of the information, such as how the woman's risk of breast cancer compared to her risk of heart disease. It might also test whether providing users with control over the amount of information displayed—such as being able to add or remove the 5-year probability estimates—results in better outcomes.

Strengths, Limitations, and Future Directions

This content analysis fills a gap in existing research on Internet-based cancer risk assessment tools. This is the first study to describe how these tools communicate individualized risk estimates to the public and to examine the relationship between risk communication format and developer affiliation. Furthermore, this study replicated the finding that information needed to evaluate the quality of the risk estimate is often missing from these tools [5].

Our extensive search methodology increases the likelihood that most of the Internet-based cancer risk assessment tools that were available in mid-October 2007 were included in this analysis. However, since then, some tools might have added or removed features, additional websites might have been activated, and existing websites might have been deactivated. The small number of websites precluded formal statistical analysis of

affiliation-based differences in communication formats. Furthermore, we only evaluated the presence or absence of specific risk formats. We did not evaluate whether the websites implemented the risk communication formats appropriately. To the best of our knowledge, empirical research has not identified the relative importance of each of the communication formats examined in this study. Consequently, we did not evaluate the overall quality of each website, nor can we say that using a greater number of risk communication formats results in better comprehension than using a lower number of formats.

Future research should examine the demographic and cancer risk profiles of people who use Internet-based cancer risk assessment tools to identify how these tools influence cancer-related cognitions, emotions, and behaviors. This will be increasingly important as we enter the second generation of the Internet [41]. Web 2.0 is eroding the conventional boundaries between information providers and information seekers. The move from static personal websites to blogs, from online encyclopedias to Wikipedia, and from one-player video games to complex interactive virtual worlds exemplifies this transformation. The Centers for Disease Control and the American Cancer Society already have virtual outposts in the online world Second Life [42]. The interactive capabilities of Web 2.0 to help people understand cancer risk and cancer risk reduction strategies may be an important means to promote a healthy lifestyle.

Implications

Internet-based cancer risk assessment tools have the potential to reach a wide audience and motivate people to engage in cancer preventive behaviors. However, because the tools do not always communicate risk in ways that facilitate comprehension, patients may misperceive or be confused about their cancer risk. This confusion may result in unintended negative consequences, such as failing to seek appropriate medical care. Researchers, organizations and clinicians who wish to provide risk assessment services to the public or their patients should refer to several excellent reviews [20,22,25] for specific recommendations on communicating risk in ways that facilitate comprehension, rather than foment confusion.

Acknowledgments

Drs. Erika Waters and Helen Sullivan conducted this research as Cancer Prevention Fellows, Office of Preventive Oncology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland. The National Cancer Institute reviewed and approved this manuscript for publication.

Conflicts of Interest

None declared.

References

1. Finney Rutten LJ, Moser RP, Beckjord EB, Hesse BW, Croyle RT. Cancer Communication: Health Information National Trends Survey. NIH Pub. No. 07-6214. Washington, DC: National Cancer Institute; 2007.
2. Ajzen I. From intentions to action: a theory of planned behavior. In: Kuhl J, Beckman J, editors. Action Control: From Cognitions to Behaviors. New York, NY: Springer; 1985:11-39.
3. Rosenstock IM. The health belief model. In: Glanz K, Lewis F, Rimer B, editors. Health Behavior and Health Education: Theory, Research and Practice. San Francisco, CA: Jossey-Bass; 1990:39-62.
4. Wills CE, Holmes-Rovner M. Patient comprehension of information for shared treatment decision making: state of the art and future directions. Patient Educ Couns 2003 Jul;50(3):285-290. [Medline: 12900101] [doi: 10.1016/S0738-3991(03)00051-X]
5. Gurmankin Levy A, Sonnad SS, Kurichi JE, Sherman M, Armstrong K. Making sense of cancer risk calculators on the web. J Gen Intern Med 2008 Mar;23(3):229-235 [FREE Full text] [Medline: 18188653] [doi: 10.1007/s11606-007-0484-x]
6. Ekman A, Hall P, Litton JE. Can we trust cancer information on the Internet?—A comparison of interactive cancer risk sites. Cancer Causes Control 2005 Aug;16(6):765-772. [Medline: 16049816] [doi: 10.1007/s10552-005-1722-3]
7. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. Med Decis Making 2001;21(1):37-44. [Medline: 11206945] [doi: 10.1177/0272989X0102100105]
8. Schwartz LM, Woloshin S, Welch HG. Can patients interpret health information? An assessment of the medical data interpretation test. Med Decis Making 2005;25(3):290-300. [Medline: 15951456] [doi: 10.1177/0272989X05276860]
9. Reyna V, Brainerd CJ. The importance of mathematics in health and human judgment: numeracy, risk communication, and medical decision making. Learn Individ Differ 2007;17(2):147-159. [doi: 10.1016/j.lindif.2007.03.010]
10. Yamagishi K. When a 12.86% mortality is more dangerous than 24.14%: implications for risk communication. Appl Cogn Psychol 1997;11(6):495-506. [doi: 10.1002/(SICI)1099-0720(199712)11:6<495::AID-ACP481>3.0.CO;2-J]
11. Zikmund-Fisher BJ, Smith DM, Ubel PA, Fagerlin A. Validation of the Subjective Numeracy Scale: effects of low numeracy on comprehension of risk communications and utility elicitations. Med Decis Making 2007;27(5):663-671. [Medline: 17652180] [doi: 10.1177/0272989X07303824]
12. Woloshin S, Schwartz LM, Welch HG. Patients and medical statistics. Interest, confidence, and ability. J Gen Intern Med 2005 Nov;20(11):996-1000 [FREE Full text] [Medline: 16307623]

13. Edwards A, Elwyn G, Gwyn R. General practice registrar responses to the use of different risk communication tools in simulated consultations: a focus group study. BMJ 1999 Sep 18;319(7212):749-752 [FREE Full text] [Medline: 10488001]
14. Rothman AJ, Martino SC, Bedell BT, Detweiler JB, Salovey P. The systematic influence of gain- and loss-framed messages on interest in and use of different types of health behavior. Pers Soc Psychol Bull 1999;25(11):1355-1369. [doi: 10.1177/0146167299259003]
15. Waters EA, Weinstein ND, Colditz GA, Emmons K. Formats for improving risk communication in medical tradeoff decisions. J Health Commun 2006 Mar;11(2):167-182. [Medline: 16537286] [doi: 10.1080/10810730500526695]
16. Cuite CL, Weinstein ND, Emmons K, Colditz G. A test of numeric formats for communicating risk probabilities. Med Decis Making 2008;28(3):377-384. [Medline: 18480036] [doi: 10.1177/0272989X08315246]
17. Gigerenzer G, Hoffrage U. How to improve Bayesian reasoning without instruction: frequency formats. Psychol Rev 1995;102(4):684-704. [doi: 10.1037/0033-295X.102.4.684]
18. McNeil BJ, Pauker SG, Sox HC, Tversky A. On the elicitation of preferences for alternative therapies. N Engl J Med 1982 May 27;306(21):1259-1262. [Medline: 7070445]
19. Waters EA, Weinstein ND, Colditz GA, Emmons KM. Reducing aversion to side effects in preventive medical treatment decisions. J Exp Psychol Appl 2007 Mar;13(1):11-21. [Medline: 17385998] [doi: 10.1037/1076-898X.13.1.11]
20. Fagerlin A, Ubel PA, Smith DM, Zikmund-Fisher BJ. Making numbers matter: present and future research in risk communication. Am J Health Behav 2007;31(Suppl 1):S47-S56. [Medline: 17931136]
21. Feldman-Stewart D, Brundage MD, Zotov V. Further insight into the perception of quantitative information: judgments of gist in treatment decisions. Med Decis Making 2007;27(1):34-43. [Medline: 17237451] [doi: 10.1177/0272989X06297101]
22. Lipkus IM. Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. Med Decis Making 2007;27(5):696-713. [Medline: 17873259] [doi: 10.1177/0272989X07307271]
23. Forrow L, Taylor WC, Arnold RM. Absolutely relative: how research results are summarized can affect treatment decisions. Am J Med 1992 Feb;92(2):121-124. [Medline: 1543193] [doi: 10.1016/0002-9343(92)90100-P]
24. Fagerlin A, Zikmund-Fisher BJ, Ubel PA. "If I'm better than average, then I'm ok?": Comparative information influences beliefs about risk and benefits. Patient Educ Couns 2007 Dec;69(1-3):140-144 [FREE Full text] [Medline: 17942271] [doi: 10.1016/j.pec.2007.08.008]
25. Schwartz LM, Woloshin S, Welch HG. Risk communication in clinical practice: putting cancer in context. J Natl Cancer Inst Monogr 1999;25(1):124-133. [Medline: 10854468]
26. Ancker JS, Senathirajah Y, Kukafka R, Starren JB. Design features of graphs in health risk communication: a systematic review. J Am Med Inform Assoc 2006;13(6):608-618 [FREE Full text] [Medline: 16929039] [doi: 10.1197/jamia.M2115]
27. Lipkus IM, Hollands JG. The visual communication of risk. J Natl Cancer Inst Monogr 1999;25(1):149-163.
28. Klein WM. Comparative risk estimates relative to the average peer predict behavioral intentions and concern about absolute risk. Risk, Decision and Policy 2002;7(2):193-202. [doi: 10.1017/S1357530902000613]
29. Klein WM. Objective standards are not enough: affective, self-evaluative, and behavioral responses to social comparison information. J Pers Soc Psychol 1997 Apr;72(4):763-774. [Medline: 9108694] [doi: 10.1037/0022-3514.72.4.763]
30. Lipkus IM, Klein WM. Effects of communicating social comparison information on risk perceptions for colorectal cancer. J Health Commun 2006 Jun;11(4):391-407. [Medline: 16720537] [doi: 10.1080/10810730600671870]
31. Weinstein ND, Sandman PM. The precaution adoption process model. In: Glanz K, Rimer B, Lewis F, editors. Health Behavior and Health Education: Theory, Research, and Practice. 3rd edition. San Francisco, CA: John Wiley and Sons; 2002:121-143.
32. Leventhal H, Diefenbach M, Leventhal EA. Illness cognition: using common sense to understand treatment adherence and affect cognition interactions. Cognit Ther Res 1992;16(2):143-163. [doi: 10.1007/BF01173486]
33. Weinstein ND. What does it mean to understand a risk? Evaluating risk comprehension. J Natl Cancer Inst Monogr 1999;25(1):15-20. [Medline: 10854451]
34. Sullivan D. Nielsen NetRatings Search Engine Ratings. SearchEngineWatch.com. 2006. URL: http://searchenginewatch.com/_subscribers/topics/&bname=stats_netratings [accessed 2009 Jul 14]
35. ; American Cancer Society. Cancer Facts & Figures. American Cancer Society. 2007. URL: http://www.cancer.org/downloads/STT/CAFF2007PWsecured.pdf [accessed 2009 Jul 14]
36. WebMD - Better Information. Better Health. WebMD. 2009. URL: http://www.webmd.com/default.htm [accessed 2009 Jul 17] [WebCite Cache ID 5iK8dq3ct]
37. Rogers RW, Prentice-Dunn S. Protection motivation theory. In: Gochman DS, editor. Handbook of Health Behavior Research I: Personal and Social Determinants. New York, NY: Plenum Press; 1997.
38. Hay J, Coups E, Ford J. Predictors of perceived risk for colon cancer in a national probability sample in the United States. J Health Commun 2006;11(Suppl 1):71-92. [Medline: 16641075] [doi: 10.1080/10810730600637376]
39. Slovic P, Monahan J, MacGregor DG. Violence risk assessment and risk communication: the effects of using actual cases, providing instruction, and employing probability versus frequency formats. Law Hum Behav 2000 Jun;24(3):271-296. [Medline: 10846372] [doi: 10.1023/A:1005595519944]
40. Berry DC, Michas IC, Bersellini E. Communicating information about medication side effects: effects on satisfaction, perceived risk to health, and intention to comply. Psychol Health 2002;17(3):247-267.

41. O'Reilly T. What is Web 2.0? Design Patterns and Business Models for the Next Generation of Software. O'Reilly Media Inc. 2005. URL: http://www.oreilly.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html [accessed 2009 Jan 28] [WebCite Cache ID 5eAjY4gTw]
42. Linden Research, Inc. Second Life. www.secondlife.com. 2009. URL: http://secondlife.com/ [accessed 2009 Jul 17] [WebCite Cache ID 5gCt3BksZ]

Abbreviations

HINTS: Health Information National Trends Survey

Edited by G Eysenbach; submitted 24.02.09; peer-reviewed by J Hay, I Lipkus; comments to author 16.04.09; revised version received 29.04.09; accepted 17.06.09; published 31.07.09.

Please cite as:
Waters EA, Sullivan HW, Nelson W, Hesse BW
What Is My Cancer Risk? How Internet-Based Cancer Risk Assessment Tools Communicate Individualized Risk Estimates to the Public: Content Analysis
J Med Internet Res 2009;11(3):e33
URL: http://www.jmir.org/2009/3/e33/
doi:10.2196/jmir.1222
PMID:19674958

Original Paper

The Effect of Credibility-Related Design Cues on Responses to a Web-Based Message About the Breast Cancer Risks From Alcohol: Randomized Controlled Trial

Peter R Harris[1], PhD; Elizabeth Sillence[2], PhD; Pamela Briggs[2], PhD

[1]Psychology Department, University of Sheffield, Western Bank, Sheffield, UK

[2]School of Psychology and Sports Sciences, Northumbria University, Newcastle upon Tyne, UK

Corresponding Author:
Peter R Harris, PhD
Department of Psychology
University of Sheffield
Western Bank, Sheffield S10 2TP
UK
Phone: +44 114 222 6627
Fax: +44 114 276 6515
Email: p.harris@sheffield.ac.uk

Abstract

Background: Internet sites typically contain visual design elements that are unrelated to the quality of the health information presented but that could influence credibility judgments and responses to health advice. To assess the effects of such design elements, or credibility cues, experimentally, we exposed women with different levels of weekly alcohol consumption to a website containing high quality but unpalatable information about a related health risk (breast cancer). The information was presented alongside either positive or negative credibility cues unrelated to information content.

Objectives: We explored four research questions: (1) Did the cues influence how the women engaged with the site? (2) Did they influence how the women responded cognitively and emotionally? (3) Did they influence whether the women subsequently acted on the advice? (4) Did the impact of the cues vary with how much alcohol the women reported drinking?

Method: A total of 85 women were randomly assigned to view one of two versions of a website containing the same high-quality content but different cues. One version had positive credibility cues (trustmarks), the other had negative ones (adverts, pharmaceutical sponsorship, and a donation button). Objective measures included visual attention (using eye-tracking equipment), time studying the material, and recall. Subjective measures included cognitive and affective responses and intention to change. Measures of subsequent behavior were taken 1 week later.

Results: First, the cues did not affect how long the women spent on the site or how long they spent reading the text. However, women in the negative cues condition spent more time looking at a donation button than those in the positive cues condition spent looking at a TRUSTe seal ($\beta = -.43$, $P < .001$) but less time looking at a logo ($\beta = .43$, $P < .001$) or at certain other features of the site. Those in the negative cues condition also recalled more site content ($\beta = -.22$, $P = .048$). Second, there were no effects of the cues on any of the measures of cognition, affect, vulnerability, or intentions. However, third, at follow-up, the positive cues had promoted greater alcohol reduction than the negative cues among those women who had previously reported drinking more heavily ($\beta = -.22$, $P = .02$). So, fourth, the responses to the cues did vary with how much alcohol the women typically drank.

Conclusions: Content-irrelevant images and logos can influence the behavioral response to quality health-risk information. These effects may be subtle, changing with time.

(J Med Internet Res 2009;11(3):e37) doi:10.2196/jmir.1097

KEYWORDS

Alcohol; breast cancer risk; health risk information; Internet; Web design

Introduction

The Internet has become an important source of health information and advice, with between 40% and 80% of those with Internet access in the United States and Europe using it for health care purposes [1,2]. However, the quality of the available information is highly variable. For example, in a 2002 meta-analysis, Eysenbach and colleagues noted that 70% of website reviews expressed concern about the quality of the health-related information provided on the Internet [3]. In the face of such variable quality, how do health consumers decide whether or not to trust the information and advice they find online?

Briggs and colleagues addressed this issue in a series of studies of trust in eHealth that led to the development of a staged model of trust and a set of guidelines [4-6]. They found that users rapidly rejected health sites on the basis of superficial cues capable of influencing consumer trust. These included advertising, complex layout, inconsistent design, or the presence or absence of reputable brand markers.

Sillence et al [5,6] noted that a great deal of high-quality health information was lost to the consumer through this process and that drug company sites in particular—often highly rated by health experts in terms of health content [7]—were frequently rejected because of the presence of commercial cues.

Health consumers, then, do not always choose the best-quality health sites or follow the best advice. Indeed, they can show a marked reluctance to trust advice they perceive to be inconsistent with their important prior beliefs, even displaying a "defensive" response to such information [8-11] that can take the form of close critical scrutiny and subsequent rejection of good-quality information [12,13]. As a result, those most at risk can be the most averse to change and may be the most difficult to persuade. Clearly, given this propensity to respond defensively, it is counter-productive to present consumers with cues that might inadvertently trigger a negative response to good health advice.

This is particularly problematic in the eHealth context, where consumers can often find conflicting advice and are able to navigate to websites that tell them what they would prefer to hear. In one study, for example, a population struggling to modify their drinking behavior sought pro-drinking rather than anti-drinking material when left free to choose via the Internet [14]. In our own study, we are interested in the fact that Internet material often contains visual design elements, or cues, that are unrelated to the quality of the health information presented but that could be used to influence credibility judgments about the site and that may also subsequently influence acceptance and adherence to health advice. What happens to an individual who spends time reading important high-quality health advice on a website that coincidentally displays negative credibility cues (such as a drug company site containing high-quality health information alongside advertising)? Can these content-irrelevant cues affect acceptance or rejection of important health advice?

We present an experimental study in which women with different levels of weekly alcohol consumption were exposed to high-quality, uncongenial, Internet-based information concerning a related health risk (breast cancer). This unpalatable information was presented alongside positive or negative trust cues unrelated to the information content. We explored four research questions: First, do the cues influence direct engagement with the material on the site? We examined how long the women spent on the site, their pattern of eye movements when examining the site's pages, and their ability to recall site content. Second, do the cues influence how they respond cognitively and affectively to the site? We examined their subsequent message acceptance, emotional and risk perceptions, and their intentions to adopt the recommended behavior. Third, do the cues influence the extent to which they subsequently act on the advice given? We examined reports of behavior at follow-up 1 week after exposure to the site. Finally, does the impact of the cues vary with how much alcohol the women typically drink? We tested whether baseline alcohol consumption moderated the effects of condition on any of the above outcome measures. That is, we tested whether any cue effects were most pronounced among those who drank most (i.e., those who might display the biggest defense motivation) [10,15].

To test these research questions we developed a two-page website containing information about the link between alcohol consumption and the risk of breast cancer. All material was based on a recent definitive study showing such a link [16]. At the time of testing this was (and in the UK largely remains) a relatively unknown association, so for most women the information on the site was novel. The text contained strong arguments that were not easily denigrated and was identical in both content and layout in both experimental conditions. Participants—women aged 18 to 46 who varied in their level of alcohol consumption—were randomly exposed to this information on a website containing either positive credibility cues (i.e., cues that had been positively associated with credibility in previous research [6]) or negative credibility cues (also based on previous findings [17]). The cues were selected to be unrelated to the content of the site. In summary, the aim of this study was to assess whether design-based credibility cues promote acceptance or rejection of important health advice.

Methods

Participants

The sample consisted of 85 women, all regular Internet users and mainly students of psychology (mean age = 22.9 years, SD = 6.5 years). Of these, over a quarter (n = 22) reported drinking more than the UK Government's recommended limit of 14 units of alcohol per week. (In the UK, a unit of alcohol is 8 g, which is approximately half a pint of beer, a standard measure of spirits, or a glass of wine.) All participants were paid £5.

Design

The study had a between-participants experimental design, with one independent variable: participants were randomly assigned to the positive (n = 42) or negative (n = 43) cues conditions. The experiment had a prospective component, with a follow-up after 1 week.

Materials and Measures

Pre-Manipulation Measures

Along with baseline alcohol consumption, we also measured age, sex, Internet use, attitudes toward alcohol [15], dispositional optimism [18], self-esteem [19], and breast cancer risk [20,21] (Table 1). Baseline alcohol was measured using the procedures employed by Harris and Napper [15], adapted from Dawson and Room [22]. Participants were asked how much alcohol they had consumed in the last 7 days, how much they consumed in a typical week, and how much they had consumed in the last 24 hours. Responses were given in terms of pints of beer/lager/cider, shots, glasses of wine, and bottles (e.g., of beer), with illustrative examples of all types, and later translated into units of alcohol by the experimenter. Del Boca and Noll [23] have shown that self-reported alcohol consumption is at least as accurate as biomarker data for measuring drinking patterns in adult general populations. Reports of alcohol consumption in the past week and in a typical week were strongly correlated (Pearson r = .70, P < .001) and were combined into a mean score for analyses. Alcohol consumption varied from 0 to 42.5 units (mean = 10.48, SD = 9.41).

Health Message

The health message was on two Web pages and contained information about the link between alcohol consumption and the risk of breast cancer. The text of the Web pages was closely based on that used by Harris and Napper [15]. The information was taken from a press release from Cancer Research UK [24] and a newspaper article [25]. All statements it contained were true.

Cues

Two different designs of the Web pages were created (see Multimedia Appendix 1 and 2). Both designs contained exactly the same health message but varied in terms of the presence of adverts, sponsors, and donation requests. The negative condition contained adverts, pharmaceutical sponsorship, and a donation button, whereas the positive condition contained a TRUSTe seal and a Health On the Net foundation (HON code) certificate. Websites can apply to be accredited by these two organizations if their site meets a number of trust indicators, including privacy, transparency, author qualifications, attribution, and justifiability. These cues were chosen from the list of credible and non-credible cues elicited in earlier phases of the research program [5,6] and were selected for their lack of direct relation to content. The size and location of these different design features were constant across both conditions.

Eye-Tracking Measures

The SensoMotoric Instruments iViewX Eye Tracking System (SensoMotoric Instruments GmbH (SMI), Teltow/Berlin, Germany) was used to record the eye movements of participants as they were viewing the Web pages on a desktop PC (1.2 GHz Pentium 4 with 256 Mb memory and a 17 inch LCD monitor). In accordance with the eye-tracking analysis software, the different design features of interest on each Web page were defined as object areas; 15 such objects were identified (Table 2). For example, object 1 was defined as the donation button in the negative condition and the TRUSTe seal in the positive condition. Some of these objects (e.g., page last updated) were common to both conditions. For each condition, the percentage of time spent looking at each object area was calculated (Table 2).

Post-Manipulation Questionnaire

After viewing the message, participants completed the post-manipulation questionnaire. This opened with the manipulation check items, comprising a measure of mood [15] followed by the trust scale developed in an earlier phase of the research program [17]. The trust scale has four factors:

1. information access: eight items (e.g., “The site told me most of what I needed to know”)
2. personalization: eight items (e.g., “It felt like the advice was tailored to me personally”)
3. credibility through impartiality: four items (e.g., “The advice appeared to be impartial and independent”)
4. credibility through design: four items (e.g., “The site had a professional design”)

Responses were given on a 5-point scale from “strongly disagree” to “strongly agree.”

The subjective measures comprised the following:

1. cognitive responses indicating acceptance of the message: three items (e.g., “I believe that drinking alcohol increases the chances of women developing breast cancer) rated on a 7-point scale from “strongly disagree” to “strongly agree”
2. negative affective responses to the message: six items (e.g., “The material on the website made me feel…”) rated on a 7-point scale from “not at all anxious” to “extremely anxious”
3. perceived vulnerability: three items (e.g., “How likely do you think you will be to get breast cancer as a result of your current level of alcohol consumption) rated on a 10-point scale from “impossible” to “extremely likely”
4. intentions to cut down on alcohol: two items (e.g., “I intend to reduce my alcohol consumption in the next 7 days by at least 2 units”) rated on a 7-point scale from “definitely do not intend to” to “definitely intend to”

Items were taken from previous studies (e.g., [15,26]). All had satisfactory reliabilities (Cronbach alpha from .72 to .98).

The questionnaire closed with a test of recall of message content (three items; e.g., “According to the report how many deaths would be avoided annually in Britain if women stopped drinking?”). Each correct answer scored 1.

Follow-Up

After 1 week, participants received a brief follow-up questionnaire by email, measuring reported alcohol consumption over the previous 7 days and containing the measures of intentions, vulnerability, cognitive response (one item), and affective response (three items) from the post-manipulation questionnaire (Cronbach alpha from .70 to .97). The questionnaire was designed to be brief in order to maximize response rate.

Procedure

Participants were tested individually. Upon arriving at the laboratory, they were told that the study involved an evaluation of health information on the Internet. After completing the pre-measures, they were asked to sit comfortably in front of the eye-tracking monitor while the researcher calibrated the system. Each participant was told that she was about to see a website consisting of two pages. They were instructed to imagine that they had found the website by following a link from a search engine and to look at and visually examine the website as they would any site they had found in this way. They were told that they could spend as little or as much time as they wanted on the website. This phase of the study ended when the participant clicked on a link at the bottom of the second Web page. They then completed the post-manipulation measures. Participants were sent the follow-up measures by email approximately 7 days later.

Statistical Analysis

One-way analysis of variance (ANOVA) with randomized condition as the between-participants independent variable or two-step hierarchical regression analyses were used to test for differences in pre-manipulation measures or other variables between the groups. An alpha level of $P = .05$ was set for all analyses.

Results

Randomization and Manipulation Checks

One-way analysis of variance (ANOVA) with randomized condition as the between-participants independent variable revealed no significant differences between the groups (Table 1; maximum $F_{1,56} = 1.52$ [Halls breast cancer risk index], $P = .22$, $eta^2 = .03$). Thus randomization appears to have been successful.

Consistent with the intended manipulation, participants in the positive cues condition trusted the site more on the two credibility trust factors, seeing the site as being higher in credibility through impartiality ($F_{1,82} = 4.74$, $P = .03$, $eta^2 = .06$) and design ($F_{1,82} = 4.92$, $P = .03$, $eta^2 = .06$; see Table 3). Thus, the manipulation appears to have been successful. Participants did not differ significantly on the access or personalization trust factors or on mood (maximum $F_{1,70} = 3.31$ [positive mood], $P = .08$, $eta^2 = .04$).

Table 1. Mean pre-manipulation measures by condition group[a]

	Negative Cues (n= 43) Mean score (SD)	Positive Cues (n= 42) Mean score (SD)
Baseline alcohol (units)[b]	10.03 (9.10)	10.95 (9.80)
Attitudes toward alcohol[c]	4.23 (0.95)	4.25 (0.92)
Self-esteem	19.33 (3.73)	20.05 (3.87)
Dispositional optimism	14.21 (4.60)	14.56 (4.74)
Breast cancer risk (Harvard risk calculator)	2.56 (0.85)	2.63 (0.80)
Breast cancer risk (Halls risk calculator)	16.72 (3.36)	18.02 (4.72)

[a]Higher scores indicate more alcohol consumption, favorable attitudes, and greater self-esteem, optimism, and risk.

[b]Alcohol measured in UK units (one unit is 8 g).

[c]Attitudes were measured on a 7-point bipolar scale.

Objective Measures

Unless stated otherwise, the remaining data were analyzed using two-step hierarchical regression analyses. This allowed us to analyze the effects of baseline alcohol consumption as a continuous variable and to assess the effects of the predictors in combination as well as individually. Where an individual predictor was significant, we report it (β) below; otherwise, in the interests of space, we report only the statistics for the predictors in combination, both in terms of significance and effect size (R^2). At step one, we assessed the main effects of condition (negative cues = 0; positive cues = 1) and baseline alcohol consumption; at step 2, we added the condition × consumption interaction. In accordance with the recommendations of Aiken and West [27], the independent variables were mean centered.

There were no significant overall effects of the predictors on the amount of time participants spent on the site ($F_{3,84} < 1$, $R^2 = .02$) or on how long they looked at the text ($F_{3,73} = 1.08$, $R^2 = .04$). None of the individual predictors (condition, alcohol consumption, or the interaction) approached significance for these dependent variables. However, there were significant main effects of condition on patterns of eye movement to certain features of the site (see Table 2). Women in the negative cues condition spent more time looking at the donation button than those in the positive cues condition spent looking at the TRUSTe seal ($\beta = -.43$, $P < .001$). In contrast, they spent less time looking at the logo ($\beta = .43$, $P < .001$), menu 1 ($\beta = .38$, $P < .001$), and when the site was last updated ($\beta = .27$, $P = .02$). There was a significant main effect of condition on recall (see Table 3), with those in the negative cues condition having significantly better total recall ($\beta = -.22$, $P = .048$). Baseline alcohol consumption did not affect eye movement or recall. No interaction was significant.

Table 2. Percentage of time participants spent visually examining objects on the website, by condition[a]

Object[b]	Negative Cues (n= 43) % of time (SD)	Positive Cues (n= 42) % of time (SD)
1. Donation/TRUSTe seal	5.03 (4.33)	1.80 (2.29)
2. Logo	5.92 (4.07)	12.82 (9.39)
3. Menu 1 (“You are here”)	5.00 (4.13)	9.80 (7.14)
4. Menu 2 (“Health issues”)	2.75 (2.86)	3.42 (3.31)
5. Advert/HON code	1.44 (1.80)	0.95 (1.68)
6. Page last updated	0.00 (0.00)	0.18 (0.50)
7. Sponsor – Pharmaceutical/NHS	0.25 (0.91)	0.34 (1.34)
8. Photo	1.50 (1.75)	1.39 (1.62)
9. Alcohol and breast cancer subtitle	1.50 (1.82)	1.45 (1.55)
10. Next page link	0.08 (0.28)	0.39 (1.67)
11. Text paragraph 1	6.31 (4.44)	6.32 (4.80)
12. Text paragraph 2	10.47 (7.13)	7.47 (7.05)
13. Text paragraph 3	3.40 (3.83)	2.82 (4.14)
14. Application window[c]	57.25 (28.27)	62.03 (26.00)
15. Entire screen[d]	60.58 (28.15)	65.16 (25.80)

[a]Standard deviations are given in parentheses.

[b] Objects with different cues (i.e., Advert/HON code and Pharmaceutical/NHS sponsor) are negative cue/positive cue.

[c] Application window refers to Web browser.

[d] Entire screen refers to everything visible on the monitor (i.e., application, borders, and task bar).

Subjective Measures

Condition did not affect any of the measures. No interaction was significant. Level of baseline alcohol consumption did not affect cognitive responses to the material, but higher levels of alcohol were associated with more negative affect ($\beta = .48$, $P < .001$), higher perceived vulnerability ($\beta = .31$, $P = .004$), and stronger intentions to reduce alcohol ($\beta = .37$, $P = .001$).

Table 3. Other dependent variables, by condition[a]

	Negative Cues (n= 43) Mean score (SD)	Positive Cues (n= 42) Mean score (SD)	Total (n= 85)
Manipulation check items			
Trustfactor 1 (access)	30.05 (4.60)	31.00 (3.56)	30.51 (4.13)
Trustfactor 2 (personalization)	16.63 (3.70)	16.88 (2.90)	16.75 (3.31)
Trustfactor 3 (impartiality)	12.33 (2.10)	13.51 (2.86)	12.90 (2.55)
Trustfactor 4 (design)	11.56 (2.36)	12.93 (3.24)	12.23 (2.89)
Positive mood	2.90 (1.03)	2.47 (1.05)	2.71 (1.05)
Negative mood	1.00 (1.41)	1.09 (1.53)	1.04 (1.46)
Outcome measures			
Recall[b]	1.07 (0.93)	0.66 (0.85)	0.87 (0.91)
Intentions to cut down alcohol	2.60 (1.64)	2.88 (1.81)	2.79 (1.73)
Cognitive responses	4.29 (1.12)	3.90 (1.12)	4.09 (1.14)
Negative affective response	3.54 (1.34)	3.58 (1.66)	3.56 (1.50)
Perceived vulnerability	2.31 (1.09)	2.44 (1.16)	2.38 (1.12)
Baseline alcohol consumption[c]	10.03 (9.10)	10.95 (9.80)	10.48 (9.41)
Follow-up	(n= 36)	(n = 39)	(n = 75)
Alcohol consumption[c]	10.63 (9.48)	9.68 (9.36)	10.13 (9.37)
Belief in link	4.53 (1.60)	4.64 (1.31)	4.59 (1.44)
Intentions to cut down alcohol	2.60 (1.64)	2.98 (1.81)	2.79 (1.73)
Negative affective responses	3.39 (1.90)	3.72 (1.96)	3.56 (1.93)
Perceived vulnerability	2.18 (0.87)	2.34 (1.15)	2.26 (1.02)

[a]Higher scores indicate more trust, more positive/negative mood, better recall, stronger intentions, more acceptance of the message, more negative affect, higher perceived vulnerability, and more alcohol consumption.

[b]Maximum possible recall is 3.

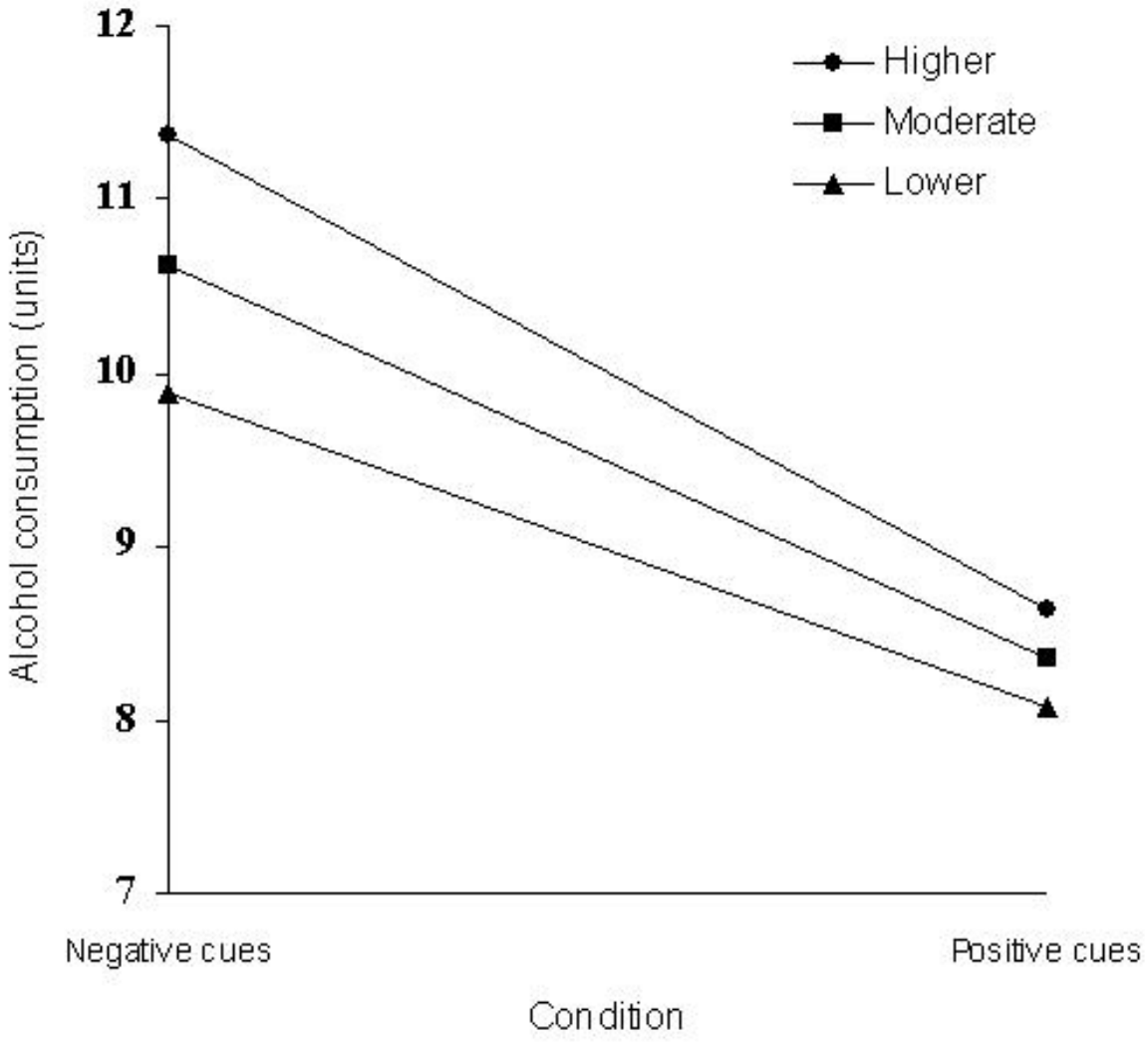
[c]Alcohol measured in UK units (one unit is 8 g).

Follow-Up

A total of 75/85 (88%) participants responded to the follow-up questionnaire. These women did not differ significantly from nonresponders on any of the randomization or manipulation check measures taken at post-manipulation (maximum $F_{1,82}$ = 1.1 [trust factor 3], P = .30, eta^2 = .013). There was no significant association between condition and responding to the follow-up ($\chi^2_{1,85}$ = 1.71, P = .33). There was a significant main effect of baseline alcohol consumption (β = .72, P < .001) but not of condition (β = −.12, P = .17) on reported alcohol consumption over the previous 7 days. More importantly, however, there was a significant interaction between condition and baseline consumption on reported alcohol consumption (β = −.22, P = .02), indicating that how much the women reported drinking at the time they were exposed to the website determined the extent to which they responded to the cues by subsequently reducing their drinking. Inspection of the simple slopes [27] indicated that at higher and moderate levels of baseline alcohol consumption, the positive cues led to greater reductions than did the negative cues (Figure 1). Overall, those in the positive cues condition reported an average decrease of 1.3 units, whereas those in the negative cues condition reported an average increase of 0.6 units.

As for the post-manipulation, higher levels of baseline alcohol in the follow-up were associated with more negative affect (β = .52, P < .001), greater perceived vulnerability (β = .39, P = .001), and stronger intentions to cut down on alcohol (β = .26, P = .04). Moreover, those who drank more at baseline now perceived the evidence linking alcohol and breast cancer to be significantly weaker (β = −.24, P = .02), thus perhaps showing the first signs of a defensive reappraisal of the message. Indeed, condition interacted with baseline alcohol consumption on this measure of belief in the evidence (β = .38, P = .001), with the effect being more pronounced in the positive than in the negative cues condition.

Figure 1. Simple slopes for the interaction between condition and baseline alcohol consumption on alcohol consumption at 1 week follow-up; simple slopes have been calculated at mean (moderate consumption), +1 standard deviation (higher consumption), and −1 standard deviation (lower consumption) levels of baseline consumption



Discussion

Did the cues influence engagement with the material on the site? We examined how long the women spent on the site, their pattern of eye movements when examining it, and their recall of site content and found that, whereas the cues did not affect the overall time the women spent on the site or how long they spent reading the text, they did influence where the women directed their gaze and how much they subsequently recalled. Participants paid more attention to certain negative cues, such as the donation button, than to positive cues in the equivalent location, and those in the negative cues condition were subsequently more accurate in their recall of information, suggesting that they paid closer attention to the site content.

Did the cues influence how the participants responded cognitively and affectively? No. We found no effects of the cues on any of the explicit ratings of cognition, affect, vulnerability, or intentions—at least not initially (see below). However, those who drank more reported more negative affect and vulnerability and expressed stronger intentions to cut down on their drinking, effects that persisted at follow-up. This suggests that the material had been persuasive, even among the heavier drinkers. Interestingly, these effects occurred even though drinking level had not influenced how the women examined the site visually or their recall of its contents.

Did the cues influence the extent to which the women subsequently acted on the advice given? Yes, the important influence of credibility cues was seen most clearly at follow-up, where the positive cues promoted greater alcohol reduction than did the negative cues among those who had previously reported heavier drinking. Thus, the answer to our final question— Does the impact of the cues vary with how much alcohol the women typically drank?—is also yes. Indeed, the cues now interacted with baseline drinking level to change the women's belief in the evidence: specifically, women who had reported drinking the most expressed less belief in the evidence at follow-up, especially the women in the positive cues condition.

Consequently, the results of this study indicate that seemingly superficial design elements of a website can influence responses to health-risk information. As predicted, cues known to be

positively or negatively associated with credibility affected engagement with the site and influenced subsequent health behavior and cognition.

We interpret these data as being consistent with findings in the broader literature on defensive responding to threatening information [8-11]. In essence, we expected the presence of positive (but non-content-related) credibility cues to bolster the message but negative cues to undermine it. We were unsure how immediate or delayed such effects might be, and it is clear that, in the short term, message content predominated. However, with time, the negative cues exerted detrimental effects by reducing the extent to which the heavier-drinking women acted on the advice. Moreover, 1 week later, the heavier drinkers also reported perceiving the evidence linking breast cancer and alcohol to be weaker, thus showing the first signs of message rejection on any of our explicit measures. Perhaps this is because, after a week of reduced consumption, they appreciated more fully the difficulties involved in trying to reduce alcohol consumption. If so, we may have captured them in the early stages of defensively re-evaluating the original website material. This might be the focus for further investigation. Indeed, at this stage there was even the suggestion that the protective effects of the positive cues may have been wearing off, as belief in the link was lower among heavier drinkers in the positive cues condition.

The findings thus present an intriguing picture of a group of participants developing their response to a website containing credible, but unwelcome, health-related information. The eye-tracking data also contribute to our understanding of how people allocate their attention to features of websites. For instance, those in the positive cues condition spent relatively little time looking at the HON code or the TRUSTe seal relative to other elements of the site, such as the logo (see Table 2). Indeed, the presence of such codes and awards has been shown elsewhere to have little effect on the credibility or retention of health information on a Web page [28], and Eysenbach and Köhler [29] noted that consumers failed to click on the HON logo when it was present on websites, despite having suggested that some form of controlling authority or an endorsement by a third party would be a helpful quality marker. There have been claims that consumer expectations of such health seals are often inconsistent with how they respond to them in practice [30]. Certainly, here participants were more inclined either to look at negative cues more than their positive equivalents or to spend relatively little time looking directly at either.

Limitations

We have shown here (to our knowledge for the first time) that seemingly superficial credibility cues embedded in health information and advice can continue to affect responding over time. However, we acknowledge that our study has several limitations. We recruited a mainly student sample to participate in a somewhat artificial environment, under experimental conditions. We cannot rule out the possibility that some participants may have felt themselves “too young” to be at risk of breast cancer or felt obliged to accept the health message contained in the website. We tried to minimize this by explicitly instructing participants to take their time and to explore the site visually as if they had discovered it themselves through a search engine. We also ensured that the health message was aimed at young women in particular. We recognize, however, that the experimental conditions also meant that participants were forced to look at the website and therefore could not choose to “select out” sites as perhaps they would in more natural settings.

Conclusions

Despite these limitations, these findings are potentially very significant since at-risk groups are typically the hardest to persuade [12,13]. Web-based intervention programs can also reach large numbers of heavier drinkers who might otherwise not seek help [31,32]. Our results suggest that there may be a health benefit in combining uncongenial information with positive credibility cues (although the benefits of this over the longer term remain to be researched). The implications in terms of the design are not limited to online presentation of material. Poor design features also occur in non-Internet material, for example, in patient leaflets [33], and negative design cues such as logos, sponsorship, and advertising are commonly found in these materials.

More studies of this kind are needed to explore both the immediate and longer term effects of design cues on health cognition and behavior. The message in this study was strong and persuasive and not easily rejected. It may be that the effects of negative design cues such as advertising are more immediate in evaluations of health risk messages that are weaker. Similarly, future research needs to explore what happens when sites containing credible design cues present incorrect information but noncredible sites present correct information. Longer-term follow-ups are also required, along with studies focusing on other health risk messages and populations.

Written information, often in the form of patient leaflets, has usually been seen as an adjunct to verbal messages provided by the medical profession [34] that may enhance and encourage behavior change [35]. Efforts are being made within the UK health service to evaluate traditional methods of conveying information to patients and to develop and assess new approaches [36]. The impact of design cues on subsequent behavior has implications for those involved in producing useful and effective patient information.

Acknowledgments

The research reported here was supported by ESRC E-Society grant number RES-335-25-0019. We wish to thank Nick Heather for helpful comments on a previous draft of this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshot of the positive cues website

[PDF file (Adobe PDF), 179 KB - jmir_v11i3e37_app1.pdf]

Multimedia Appendix 2

Screenshot of the negative cues website

[PDF file (Adobe PDF), 167 KB - jmir_v11i3e37_app2.pdf]

References

1. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the Internet and e-mail for health care information: Results from a national survey. JAMA 2003 May 14;289(18):2400-2406 [FREE Full text] [Medline: 12746364] [doi: 10.1001/jama.289.18.2400]
2. Fox S. Online health search 2006. Pew Internet & American Life Project. Washington, DC: Pew Internet & American Life Project; 2006 Oct 29. URL: http://www.pewinternet.org/PPF/r/190/report_display.asp [WebCite Cache ID 5eexGWT9f]
3. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002;287(20):2691-2700 [FREE Full text] [Medline: 12020305] [doi: 10.1001/jama.287.20.2691]
4. Briggs P, Burford B, De Angeli A, Lynch P. Trust in online advice. Social Science Computer Review 2002;20:321-332.
5. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? Soc Sci Med 2007 May;64(9):1853-1862. [Medline: 17328998] [doi: 10.1016/j.socscimed.2007.01.012]
6. Sillence E, Briggs P, Harris PR, Fishwick L. Health websites that people can trust - the case of hypertension. Interacting with Computers 2007;19(1):32-42. [doi: 10.1016/j.intcom.2006.07.009]
7. Reed M, Anderson C. Evaluation of patient information Internet web sites about menopause and hormone replacement therapy. Maturitas 2002 Oct 25;43(2):135-154. [Medline: 12385862] [doi: 10.1016/S0378-5122(02)00188-3]
8. Croyle RT, Sun YC, Louie DH. Psychological minimization of cholesterol test results: moderators of appraisal in college students and community residents. Health Psychol 1993 Nov;12(6):503-507. [Medline: 8293735] [doi: 10.1037/0278-6133.12.6.503]
9. Ditto PH, Scepansky JA, Munro GD, Apanovitch AM, Lockhart LK. Motivated sensitivity to preference-inconsistent information. Journal of Personality and Social Psychology 1998;75(1):53-69. [doi: 10.1037/0022-3514.75.1.53]
10. Kunda Z. Motivation and inference: Self-serving generation and evaluation of evidence. Journal of Personality and Social Psychology 1987;53(4):229-241. [doi: 10.1037/0022-3514.53.4.636]
11. Renner B. Biased reasoning: adaptive responses to health risk feedback. Pers Soc Psychol Bull 2004 Mar;30(3):384-396. [Medline: 15030627] [doi: 10.1177/0146167203261296]
12. Eagly AH, Kulesa P, Chen S, Chaiken S. Do attitudes affect memory? Tests of the congeniality hypothesis. Current Directions in Psychological Science 2001;10(1):5-9. [doi: 10.1111/1467-8721.00102]
13. Liberman A, Chaiken S. Defensive processing of personally relevant health messages. Personality and Social Psychology Bulletin 1992;18(6):669-679. [doi: 10.1177/0146167292186002]
14. Joinson AN, Banyard P. Seeking alcohol information on the Internet. Aslib Proceedings 2003;55(5/6):313-319. [doi: 10.1108/00012530310498888]
15. Harris PR, Napper L. Self-affirmation and the biased processing of threatening health-risk information. Pers Soc Psychol Bull 2005 Sep;31(9):1250-1263. [Medline: 16055644] [doi: 10.1177/0146167205274694]
16. ; Collaborative Group on Hormonal Factors in Breast Cancer. Alcohol, tobacco and breast cancer - collaborative reanalysis of individual data from 53 epidemiological studies. British Journal of Cancer 2002;87(11):1234-1245. [Medline: 12439712] [doi: 10.1038/sj.bjc.6600596]
17. Sillence E, Briggs P, Harris PR, Fishwick L. Going online for health advice: Changes in usage and trust practices over the last five years. Interacting with Computers 2007;19(3):397-406. [doi: 10.1016/j.intcom.2006.10.002]
18. Scheier MF, Carver CS, Bridges MW. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. J Pers Soc Psychol 1994 Dec;67(6):1063-1078. [Medline: 7815302] [doi: 10.1037/0022-3514.67.6.1063]
19. Rosenberg M. Society and the Adolescent Self-image. Princeton, NJ: Princeton University Press; 1965.
20. Detailed Breast Cancer Risk Calculator. halls.md. URL: http://www.halls.md/breast/risk.htm [WebCite Cache ID 5WjALECOK]

21. ; Harvard School of Public Health. Disease Risk Index. Cancer—breast cancer. URL: http://www.diseaseriskindex.harvard.edu/update/hccpquiz.pl?lang=english&func=home&quiz=breast [WebCite Cache ID 5Wj9raUkB]
22. Dawson DA, Room R. Towards agreement on ways to measure and report drinking patterns and alcohol-related problems in adult general population surveys: the Skarpö conference overview. J Subst Abuse 2000;12(1-2):1-21. [Medline: 11288465] [doi: 10.1016/S0899-3289(00)00037-7]
23. Del Boca FK, Noll JA. Truth or consequences: the validity of self-report data in health services research on addictions. Addiction 2000 Nov;95(Suppl):S347-S360. [Medline: 11132362]
24. Cancer Research UK. Alcohol, Smoking and Breast Cancer: The definitive answer. Media Release 2002 Nov 12.
25. Boseley S. Daily glass of wine boosts breast cancer risk by 6%. The Guardian 2002 Nov 13.
26. Norman P, Bennett P, Lewis H. Understanding binge drinking among young people: an application of the Theory of Planned Behaviour. Health Educ Res 1998 Jun;13(2):163-169 [FREE Full text] [Medline: 10181015] [doi: 10.1093/her/13.2.163-a]
27. Aiken LS, West RR. Multiple Regression: Testing and Interpreting Interactions. Newbury Park, CA: Sage; 1991.
28. Shon J, Marshall J, Musen MA. The impact of displayed awards on the credibility and retention of Web site information. Proc AMIA Symp 2000:794-798 [FREE Full text] [Medline: 11079993]
29. 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] [doi: 10.1136/bmj.324.7337.573]
30. Burkell J. Health information seals of approval: what do they signify? Information, Communication & Society 2004;7(4):491-509. [doi: 10.1080/1369118042000305610]
31. Linke S, Murray E, Butler C, Wallace P. Internet-based interactive health intervention for the promotion of sensible drinking: patterns of use and potential impact on members of the general public. J Med Internet Res 2007;9(2):e10 [FREE Full text] [Medline: 17513281] [doi: 10.2196/jmir.9.2.e10]
32. Cunningham JA, Humphreys K, Kypri K, van Mierlo T. Formative evaluation and three-month follow-up of an online personalized assessment feedback intervention for problem drinkers. J Med Internet Res 2006;8(2):e5 [FREE Full text] [Medline: 16867968] [doi: 10.2196/jmir.8.2.e5]
33. Fitzmaurice DA, Adams JL. A systematic review of patient information leaflets for hypertension. J Hum Hypertens 2000 Apr;14(4):259-262. [Medline: 10805051]
34. Fitzmaurice DA. Written information for treating minor illness. BMJ 2001;322(7296):1193-1194. [doi: 10.1136/bmj.322.7296.1193]
35. Neil HA, Roe L, Godlee RJ, Moore JW, Clark GM, Brown J, et al. Randomised trial of lipid lowering dietary advice in general practice: the effects on serum lipids, lipoproteins, and antioxidants. BMJ 1995 Mar 4;310(6979):569-573 [FREE Full text] [Medline: 7888933]
36. Graham W, Smith P, Kamal A, Fitzmaurice A, Smith N, Hamilton N. Randomised controlled trial comparing effectiveness of touch screen system with leaflet for providing women with information on prenatal tests. BMJ 2000 Jan 15;320(7228):155-160 [FREE Full text] [Medline: 10634736] [doi: 10.1136/bmj.320.7228.155]

Edited by G Eysenbach; submitted 21.05.08; peer-reviewed by A Joinson; comments to author 03.02.09; revised version received 04.03.09; accepted 17.06.09; published 25.08.09.

Please cite as:
Harris PR, Sillence E, Briggs P
The Effect of Credibility-Related Design Cues on Responses to a Web-Based Message About the Breast Cancer Risks From Alcohol: Randomized Controlled Trial
J Med Internet Res 2009;11(3):e37
URL: http://www.jmir.org/2009/3/e37/
doi:10.2196/jmir.1097
PMID:19709989

Original Paper

Patients' Attitudes Toward Electronic Health Information Exchange: Qualitative Study

Steven R Simon[1], MD, MPH; J Stewart Evans[2], MD; Alison Benjamin[3], MPA; David Delano[3], BA; David W Bates[4,5], MD, MSc

[1]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care, Boston, MA

[2]Division of Clinical Decision Making, Informatics and Telemedicine, Department of Medicine, Tufts-New England Medical Center, Tufts University School of Medicine, Boston, MA

[3]Northern Berkshire eHealth Collaborative, North Adams, MA

[4]Division of General Internal Medicine, Brigham and Women's Hospital, Boston, MA

[5]Department of Clinical and Quality Analysis, Partners HealthCare System, Inc, Boston, MA

Corresponding Author:
Steven R Simon, MD, MPH
Department of Population Medicine
Harvard Medical School and Harvard Pilgrim Health Care
133 Brookline Avenue, Sixth Floor
Boston, MA 02215
USA
Phone: +1 617 509 9938
Fax: +1 617 859 8112
Email: steven_simon@hphc.org

Abstract

Background: In many countries, there has been substantial progress in establishing the electronic transmission of patients' health information between health care providers, but little is known about how best to engage patients in the process.

Objective: We explored patients' views about sharing of electronic health information and their preferences for learning about and participating in this process.

Methods: Patients in one Massachusetts community in the northeastern United States were recruited to participate in focus-group discussions. Prior to discussion, participants completed a written questionnaire that captured their reactions to draft educational materials and a consent form. The discussion moderator and two physicians analyzed the moderator's detailed notes from each session and participants' written comments, using an immersion-crystallization approach.

Results: Three dominant themes emerged: (1) concerns about privacy and security, (2) the potential benefit to a person's health, and (3) the desire for more information about the consent process. On the pre-discussion questionnaire, 55 out of 62 participants (88%) indicated that they would provide consent for their information to be shared electronically among their health care providers, given the materials they had reviewed.

Conclusions: Patients are enthusiastic about electronic health information exchange, recognizing its capacity to improve the quality and safety of health care; however, they are also concerned about its potential to result in breached privacy and misuse of health data. As the exchange of electronic health information becomes more widespread, policy makers will need to ensure that patients have access to concise educational materials and opportunities to engage in conversations about the risks and benefits of participation.

(J Med Internet Res 2009;11(3):e30) doi:10.2196/jmir.1164

KEYWORDS

Regional health information organizations; health information technology; electronic health records; quality of health care

Introduction

The United States has lagged behind other developed nations with respect to adoption of electronic health records (EHRs), especially in primary care, although that appears likely to change. The American Recovery and Reinvestment Act of 2009 will result in an investment of approximately US $19 billion toward the adoption of EHRs and, under the direction of the Office of the National Coordinator for Health Information Technology, an "initial set of standards, implementation specifications, and certification criteria" to enable the electronic exchange of health information [1]. With these developments in the United States, electronic sharing of health information—typically defined as the exchange of personal health information contained in the medical record between at least two different computer networks—is expected to increase exponentially over the next decade [2]. The expansion of electronic data sharing, also known as health information exchange (HIE), is heralded as a key solution to the problems of low quality and high cost of health care [3]. However, HIE has presented substantial challenges, not only in the United States but also in countries such as the United Kingdom [4], Australia [5], and Sweden [6].

The potential benefits of electronic health information exchange include improved health care quality, reduced medical errors and lower health care costs, as well as public health benefits, resulting from early detection of infectious disease and improved tracking of chronic disease management [7]. In the United States, regional health information organizations (RHIOs) have emerged as the leading model to facilitate the electronic exchange of patient-level clinical information between physicians' offices and other health-care organizations that deliver care to patients [8,9]. The overarching goal of these RHIOs is to ensure that physicians and other health care providers have access to the best and most complete information about patients for whom they are caring when they need it most—in real time, when the patient is with them in the examination room, emergency department, or other clinical setting. Not only could awareness of a patient's medical history, such as, for example, a current problem list or known allergies to medications, improve the quality of health care, but it could be, in some instances, life-saving [10]. The model of RHIOs is considerably different from that being used in other nations such as the United Kingdom, which has a single "spine" which is being implemented [11].

To date in the United States, however, the electronic exchange of health information has faced considerable challenges with respect to technical limitations and financial constraints (e.g., is there a sustainable business model?) [12]. Overcoming the lack of interoperability (i.e., the problem of systems on different platforms being unable to exchange information) and the question of who will pay for health information exchange have presented major impediments. In addition, concerns about the privacy and security of personal health information in these systems have also emerged [13,14]. As communities begin to overcome the technical and financial barriers to health information exchange, attention has turned to the importance of engaging community members—the patients—in the process. Specifically, policymakers have begun to focus on patients' perspectives of electronic health information exchange, especially their concerns about confidentiality and data security, and their willingness to provide consent for health information exchange.

In preparation for launching community-wide electronic health information exchange in Massachusetts, we conducted formal and informal discussions among stakeholders to address key questions about the process: How well do patients understand the value of clinical data exchange? To what extent do they endorse the electronic transmission of clinical information among health care providers? What are their concerns and hesitations about the process? How should patients be informed about, and approached for, participation in a community's HIE?

Methods

Study Design

We conducted a qualitative analysis based on transcribed moderators' notes from five focus group discussions and on the free-text and responses from semi-structured questionnaires completed anonymously by focus group discussion participants. The research protocol was approved by the Partners HealthCare Human Research Committee.

Setting

The study was conducted in the Northern Berkshire e-Health Collaborative, one of three communities participating in the Massachusetts e-Health Collaborative. The Northern Berkshire community is a rural, socio-economically diverse region with a population of about 45,000 people, located in the northwestern corner of Massachusetts. It includes the city of North Adams and several smaller towns that surround it. Northern Berkshire community members receive the majority of their health care services from physicians and other health care professionals located in North Adams, Adams, and Williamstown. Focus group discussions were conducted at various locations, including community centers, health care facilities, and restaurants, to solicit a broad spectrum of patients' attitudes and perceptions.

Participants

We recruited adult community members to participate in focus group discussions based on their geographic proximity and affiliation with the site at which each session was being held. For example, for a session conducted at the North Adams Regional Hospital, we recruited hospital employees. We advertised the focus groups at each location, with posted signs, flyers, and email announcements used to recruit participants. We aimed to include a mix of men and women, ranging from young adults to senior citizens, and attempted to include individuals with varying levels of formal education.

Focus Group Discussions

A trained moderator conducted each of the focus group discussions, with one or two additional observers present for each session. The duration of each session ranged from 30 to 120 minutes. The sessions were not audio recorded; however, the moderator and observers took extensive notes and recorded

their collective observations from discussions that followed the sessions.

At the start of each session, prior to any discussion, we asked participants to review the following draft materials as though they were seeing them for the first time at a routine visit to their doctor's office:

- Form seeking consent for patients to allow their doctor's office to share health information with other physicians' offices. (See Multimedia Appendix 1)
- Booklet of supporting information, describing the health information exchange, data security, privacy considerations, and contact information. (See Multimedia Appendix 2)

After reading the materials, participants were asked to complete a one-page questionnaire that ascertained their age, sex, highest education level attained, and health condition (overall healthy, healthy now but with past concerns, some current health concerns, significant current health concerns). In addition, the questionnaire asked whether they would sign the consent form that they had just read.

During each session, the moderator facilitated discussion of individuals' reactions to the documents, including their likes and dislikes, preferences for additional information, reservations about the forms or the HIE itself, and thoughts on improving the documents. Participants were encouraged to ask questions about anything that seemed unclear. Because of scheduling issues, one of the five sessions was conducted as a series of one-on-one interviews between the moderator and individual community members as they arrived at the planned focus group meeting. The content of these interviews was similar to the group discussions and was included without discrimination in this analysis.

Analysis

We performed a content analysis of the moderators' notes and free-text comments from the questionnaires using an immersion-crystallization technique [15]. Three of us (SRS, JSE, AB) read all of the available text and identified the salient themes and principles that emerged. Any discrepancies were settled by consensus. The raw rate of willingness to provide consent for HIE was calculated from the survey forms and stratified by age group, gender, education level, and health condition (healthy vs current health concerns).

Results

Participant Characteristics

Of the 64 study participants, 61 (95%) provided information on their age, sex, and health status. The median age was 50 years (range: 19 to 86 years), and 46 patients (75%) were women. A total of 52 participants (85%) had attended college or some graduate school, reflecting a sample with higher education level than the community at-large. A total of 36 patients (59%) considered themselves generally healthy, while 17 (28%) said that they had some current health concerns.

Salient Themes

The three most common themes that emerged from the focus group discussions and qualitative comments from the written questionnaires were (1) concerns about privacy and security, (2) the potential benefit to a person's health, and (3) the desire for more information about the consent process.

Privacy and Security Concerns

Comments and discussion about privacy ranged from general concerns about privacy to specific concerns about who will have access to the personal health information, what kinds of sensitive health information would be shared, and the risk of unauthorized access to the health information via security breaches. One woman who was "over 65", expressed her acceptance of health professionals' sharing her data but her simultaneous reservation about unauthorized access:

> *I realize that people in the office already can look at my chart. But I'm worried about people that don't need to know—hackers on the outside. What they would do with the information, I don't know, but I still don't like it.*

Others expressed a considerable level of trust in the security of the system. For example, participants were generally nonchalant when informed that some potentially sensitive health information could be shared among physicians' offices. When the moderator noted, for example, that prescriptions for mental health conditions or for erectile dysfunction would be viewable by multiple providers, there was no measurable pushback. One man commented:

> *Yeah, but the doctors [already] ask you about all that stuff anyway, right? This isn't really that different.*

Potential Health Benefits

Across a wide spectrum of participants, the potential for health information exchange to improve health and prevent adverse outcomes was unambiguously endorsed as a rationale for participating. The health benefits of electronic health information exchange were cited with equal frequency among those who had concerns about privacy and security as those who had no reservations about proceeding with the community-wide clinical data exchange. For example, one man with no concerns about security proclaimed:

> *Yeah, I'd sign that [consent form]. It's for my benefit so why not? Nobody can get at the information, so what the heck difference does it make anyway?*

A 29-year-old woman who has some chronic medical conditions said that health information exchange was a "great idea" and would "make the process of seeing a specialist much easier". Another woman, who mentioned that her father had a chronic illness, reported that his paper medical chart had recently been lost at the doctor's office and expressed optimism that an electronic record with health information exchange would avoid that problem:

> *It was extremely important, and we just don't know what happened to it. If this prevents that from happening, it's a good thing.*

Desire for More Information

Some participants expressed high levels of satisfaction with the draft document presented during the focus groups and requested no further explanation or information. In fact, at a focus group held at a local restaurant, several middle-aged men completed the draft forms and were initially under the impression that they were signing the official documents. Nevertheless, participants throughout all demographic groups suggested that the process of obtaining patient consent for opting in to the system of health information exchange would require considerable patience and more extensive information resources. Virtually all patients expressed the sentiment that they should need to provide consent for health information exchange (i.e., an opt-in system); a system that assumed their willingness to participate without obtaining explicit consent (i.e., an opt-out system) would not be acceptable.

There was consensus in all groups that patients should receive information by mail prior to being asked to sign the consent form in the doctor's office. While there were a few individuals who said that they would simply sign the consent form without reading it carefully, most participants said that they would want to take time to read it thoroughly to consider whether to participate. Focus group participants suggested that patients could be warned to arrive 10 - 15 minutes prior to their appointment to enable time for the registration and consent process to occur. Patients agreed that someone in the physician's office needs to be dedicated to the sign-up/consent process:

> *If you send an advance notice, you'll need to be prepared for a lot of calls. You'll need someone who can take the time to answer people's questions, both when the mailing goes out, and when people come in the office.*

One older woman who described herself as ambivalent about whether to opt in to the HIE noted that both of her physicians were included on the list of participating clinicians and then commented:

> *I would have to discuss it very thoroughly with them—what type of personal information they're putting in there.*

Willingness to Provide Consent

In completing the written questionnaires prior to beginning the discussion component of the focus group meetings, 55 out of 62 participants (88%) indicated that they would provide consent for participation in the system of health information exchange, given the materials they had reviewed. While the study was not designed to detect differences in the rates of consent among demographic subgroups, the proportion of participants willing to opt in was qualitatively similar among men and women, younger and older individuals, and those with and without current health concerns, and it was not related to the level of education.

Discussion

Exchanging health-related information electronically to improve clinical practice is central to maximizing the benefit of ongoing efforts to expand electronic health records to physicians' offices everywhere [16]. Because health information exchange involves electronically exchanging patient-identified health information across geographically and commercially separate entities, it raises issues of patient privacy and data security which have resulted in heated debate [17,18]. However, little is known about patients' attitudes toward health information exchange and their preferences for learning about it and giving consent for it. In this qualitative analysis of five focus groups within a community on the eve of launching a regional system of health information exchange, we found that most patients were willing to opt in to a system of health information exchange, although they had concerns about security and privacy and wanted assurances that they would be able to ask questions and obtain more information prior to consenting. Overwhelmingly, patients recognized the potential for the electronic exchange of health information to improve the quality of health care and prevent medical errors. This potential was the driving force behind many patients' enthusiasm for participation.

Patients' concerns about privacy and data security are not surprising, given the attention paid to these issues in both the medical and lay literature [19,20]. Furthermore, there have been some well-publicized breaches, such as an incident in the United Kingdom in which data from 25 million individuals were placed on CDs that were lost in the mail [21]. While some have pointed out that paper-based records systems have long been subject to breaches of privacy and security, the potential for large volumes of electronic data to be accessed in short amounts of time and the ease with which those data can be transmitted from user to user have elevated concerns about potential privacy violations and security breaches in computerized systems [22]. In addition, the high visibility of news stories which depict data loss and security breaches in a variety of business sectors [23], as well as in health care [24,25], seems to have sensitized the public to these concerns in the context of health information exchange. As countries intensify health information exchange efforts, concerns about privacy and data security are likely to increase, as well. For example, in the United Kingdom, where the National Health Service transmits more than 100 million clinical messages electronically every month [26], public concern has been particularly vocal and substantially hampered efforts to advance clinical information exchange programs [27,28]. It is reasonable to expect that public concern in other countries will increase correspondingly as the volume of data exchange expands to these levels.

In the United States, most early data exchange efforts have been at the community level. In this study, the concerns about privacy and security ranged from non-specific, "gut-level" worries to sophisticated, reasoned acknowledgements of the risks in both electronic and paper-based systems. Without question, though, patients across the community in our study expected privacy and security to be addressed and adequately assured.

Among community members, there was unanimous acceptance of the notion that health information exchange would lead to improvements in the quality and safety of health care. Patients cited their own experiences with misplaced paper charts, for example. They easily recognized the potential value of having their health information, such as medication lists or allergies,

immediately accessible to clinicians in practice settings where they are not customarily available. In this study, no one expressed any skepticism about the potential for health information exchange to improve care, and in fact many patients linked this potential benefit to their willingness to provide consent for participation. The promise of improved quality and safety of health care seemed to outweigh patients' deep-seated concerns about confidentiality and data security, as long as the system explicitly and adequately addressed the latter.

Participants in this study also articulated a consistent message that community members must receive clear and concise materials describing the system of health information exchange and have opportunities to ask questions about it before they would be willing to opt in. While nearly all participants indicated that they would provide consent for inclusion in the exchange system, many expressed a need to have information about it, and especially assurances of privacy and security, available for review and consideration well in advance of being asked to "sign on the dotted line". This finding indicates that enrolling a community in health information exchange may take considerable time, as patients want to be able to read and reflect on materials, discuss them with family members, and ask clarifying questions before committing to it. Patients' interest in making an informed decision also likely reflects the fact that they expect to have control over whether their information is included in the system.

The input from the focus group discussions led to considerable revision of the consent form (See Multimedia Appendix 3), the educational / informational brochure (See Multimedia Appendix 4), and the fact sheet (See Multimedia Appendix 5) that were ultimately produced and distributed within the community. In the initial materials and during the discussions, for example, the health information exchange was referred to as a community health record or a shared health record. Focus group discussions suggested that "e-health summary" was a more appealing and meaningful description. In addition, feedback from patients led program leaders to craft a schematic diagram of the multiple sources of information that would populate the e-health summary (See Multimedia Appendix 4, page 5).

Perhaps the most noteworthy outcome of the focus group discussions was the decision to make health information exchange an "opt-in", rather than and "opt-out", experience. At the time of the focus group discussions, regional policymakers had considered establishing the health information exchange system such that the records of all patients of all participating physicians would be included unless the patient actively opted out. However, patients' robust preferences for retaining the authority to provide consent—and their near unanimous expressions of willingness to provide that consent—led to the establishment of an "opt-in" system.

A few prior studies from the United States have explored patients' attitudes toward the use and protection of health information [29-31]. Little is known about how best to approach the process of engaging community members in health information exchange in the United States and abroad. One of the first regional health information exchange systems in the United States, the Santa Barbara County Care Data Exchange, recognized that privacy issues need to be addressed explicitly and early in the process [32]. Santa Barbara leaders also realized the value of engaging community members in the development of policies intended to educate consenting patients before they participated in the health information exchange [33]. These observations are consistent with and reinforced by the salient themes that emerged from our study.

Interestingly, some experts have suggested the potential for streamlining the consent process for health information exchange through electronic communication [34]. While our study did not directly assess the acceptability of an e-consent process, we did find that community members expect information to be presented clearly and concisely, with time provided between the presentation of information and the need to consent in order for there to be opportunities to ask for more or clarified information. To the extent that electronic consent processes can incorporate these community needs, they may enable communities to streamline the enrollment process, though access to paper materials and human information sources will likely remain essential.

The findings of this analysis need to be considered in the context of the study design. We conducted five focus groups in one community in Massachusetts, and the attitudes and preferences of patients in communities elsewhere may differ. On the other hand, the purpose of this study was to identify the breadth of attitudes among community members about consent for health information exchange and, as such, it is likely to have identified the relevant domains of concern for many other communities embarking on similar efforts.

This study provides insight into the ways that patients perceive electronic health information exchange and their willingness to provide consent for participation. Others have already recognized the value of sharing experiences and lessons learned from community implementation of health information technology, including the electronic exchange of health information [9,35]. Future studies should test differing strategies for educating community members about health information exchange and for securing their consent for participation. While there will likely be variability across communities and nations, as well as a need for local programs and policies, each community embarking on the implementation of clinical data exchange should not need to "reinvent the wheel" in terms of engaging patients in the process.

Acknowledgments

We acknowledge the work of Christina Kara, who provided administrative support and assistance with manuscript preparation.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Draft informed consent form

[PDF file (Adobe PDF), 63 KB - jmir_v11i3e30_app1.pdf]

Multimedia Appendix 2

Booklet of supporting information

[PDF file (Adobe PDF), 35 KB - jmir_v11i3e30_app2.pdf]

Multimedia Appendix 3

Patient consent form to allow sharing of medical information via the health information exchange

[PDF file (Adobe PDF), 62 KB - jmir_v11i3e30_app3.pdf]

Multimedia Appendix 4

Educational / informational brochure describing electronic health records and the health information exchange

[PDF file (Adobe PDF), 389 KB - jmir_v11i3e30_app4.pdf]

Multimedia Appendix 5

Information sheet mailed in advance to members of the Northern Berkshire Community

[PDF file (Adobe PDF), 39 KB - jmir_v11i3e30_app5.pdf]

References

1. One Hundred Eleventh Congress of the United States of America. The American Recovery and Reinvestment Act of 2009, Sec. 3004. URL: http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=111_cong_bills&docid=f:h1enr.pdf [accessed 2009 Jul 2] [WebCite Cache ID 5hNa8eGqk]
2. Adler-Milstein J, Bates DW, Jha AK. U.S. Regional health information organizations: progress and challenges. Health Aff (Millwood) 2009;28(2):483-492. [Medline: 19276008] [doi: 10.1377/hlthaff.28.2.483]
3. ; Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Press; 2001.
4. ; United Kingdom Parliament. The National Programme for IT in the NHS: Progress since 2006 — Public Accounts Committee. URL: http://www.publications.parliament.uk/pa/cm200809/cmselect/cmpubacc/153/15304.htm [accessed 2009 Jul 2] [WebCite Cache ID 5hNZdvqxf]
5. de Lusignan S, Teasdale S. Achieving benefit for patients in primary care informatics: the report of a international consensus workshop at Medinfo 2007. Inform Prim Care 2007;15(4):255-261. [Medline: 18237483]
6. Olvingson C, Hallberg J, Timpka T, Lindqvist K. Ethical issues in public health informatics: implications for system design when sharing geographic information. J Biomed Inform 2002 Jun;35(3):178-185. [Medline: 12669981] [doi: 10.1016/S1532-0464(02)00527-0]
7. ; Institute of Medicine. Patient Safety: Achieving a New Standard for Care. Washington, DC: National Academies Press; 2003.
8. Blumenthal D, Glaser JP. Information technology comes to medicine. N Engl J Med 2007 Jun 14;356(24):2527-2534. [Medline: 17568035] [doi: 10.1056/NEJMhpr066212]
9. Yasnoff WA, Humphreys BL, Overhage JM, et al. A consensus action agenda for achieving the national health information infrastructure. J Am Med Inform Assoc 2004;11(4):332-338. [Medline: 15187075] [doi: 10.1197/jamia.M1616]
10. Halamka JD. Health information technology: shall we wait for the evidence? Ann Intern Med 2006 May 16;144(10):775-776 [FREE Full text] [Medline: 16702595]
11. ; National Health Service. Spine Factsheet. NHS Connecting for Health. 2009. URL: http://www.connectingforhealth.nhs.uk/resources/systserv/spine-factsheet [accessed 2009 Jul 2] [WebCite Cache ID 5hNaqvHy4]
12. Adler-Milstein J, McAfee AP, Bates DW, Jha AK. The state of regional health information organizations: current activities and financing. Health Aff (Millwood) 2008;27(1):w60-w69. [Medline: 18073225]
13. Bright B. Benefits of Electronic Health Records Seen as Outweighing Privacy Risks. The Wall Street Journal. 2007. URL: http://online.wsj.com/article/SB119565244262500549.html [accessed 2009 Jul 2] [WebCite Cache ID 5hPHN5DNo]

14. ; Deloitte Center for Health Solutions. 2009 Survey of Health Care Consumers: Key Findings, Strategic Implications. Deloitte. 2009. URL: http://www.deloitte.com/ [accessed 2009 Jul 2] [WebCite Cache ID 5hQlBtebn]
15. Borkan J. Immersion/crystallization. In: Crabtree B, Miller W, editors. Doing Qualitative Research, 2nd edition. Thousand Oaks, CA: Sage Publications; 1999.
16. Chaudhry B, Wang J, Wu S, et al. Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006;144(10):742-752. [Medline: 16702590]
17. Manos D. Privacy concerns remain barrier to health information exchanges. Healthcare IT News. 2007. URL: http://www.healthcareitnews.com/news/privacy-concerns-remain-barrier-health-information-exchanges [accessed 2009 Jul 2] [WebCite Cache ID 5hQstgLjP]
18. Confidentiality as part of a bigger picture: a discussion paper from the BMA. British Medical Association. 2009. URL: http://www.bma.org.uk/ap.nsf/content/ConfidentialityBiggerPicture [accessed 2009 Jul 2] [WebCite Cache ID 5hNavK8fY]
19. Editorial. New York Times. Safeguarding private medical data. The New York Times. 2008. URL: http://www.nytimes.com/2008/03/26/opinion/26wed2.html [accessed 2009 Jul 2] [WebCite Cache ID 5hQlDYMWw]
20. Slutsman J, Kass N, McGready J, Wynia M. Health information, the HIPAA privacy rule, and health care: what do physicians think? Health Aff (Millwood) 2005;24(3):832-842. [Medline: 15886179] [doi: 10.1377/hlthaff.24.3.832]
21. Wintour P. Lost in the post — 25 million at risk after data discs go missing. guardian.co.uk. 2007. URL: http://www.guardian.co.uk/politics/2007/nov/21/immigrationpolicy.economy3 [accessed 2009 Jul 2] [WebCite Cache ID 5hQlF2AYT]
22. Bates DW. Commentary: quality, costs, privacy and electronic medical data. J Law Med Ethics 1982;25(2-3):111-112. [Medline: 11066485] [doi: 10.1111/j.1748-720X.1997.tb01886.x]
23. Dash E. Data breach could affect millions of TJX shoppers. The New York Times. 2007. URL: http://www.nytimes.com/2007/01/19/business/19data.html [accessed 2009 Jul 2] [WebCite Cache ID 5hQl944M7]
24. Stout D, Zeller TJ. Vast data cache about veterans is stolen. The New York Times. 2006. URL: http://www.nytimes.com/2006/05/23/washington/23identity.html [accessed 2009 Jul 2] [WebCite Cache ID 5hQl7xlul]
25. Blankstein A. Digging into celebrity medical records has a long history. Los Angeles Times. 2008. URL: http://articles.latimes.com/2008/may/19/local/me-tabs20 [accessed 2009 Jul 2] [WebCite Cache ID 5hQl4zMqY]
26. ; National Audit Office. The National Programme for IT in the NHS: Project Progress Reports. UK National Audit Office. 2008. URL: http://www.nao.org.uk/publications/nao_reports/07-08/0708484ii.pdf [accessed 2009 Jul 2] [WebCite Cache ID 5hQlGp2x2]
27. Adams T, Budden M, Hoare C, Sanderson H. Lessons from the central Hampshire electronic health record pilot project: issues of data protection and consent. BMJ 2004 Apr 10;328(7444):871-874 [FREE Full text] [Medline: 15073070] [doi: 10.1136/bmj.328.7444.871]
28. Whiddett R, Hunter I, Engelbrecht J, Handy J. Patients' attitudes towards sharing their health information. Int J Med Inform 2006 Jul;75(7):530-541. [Medline: 16198142] [doi: 10.1016/j.ijmedinf.2005.08.009]
29. Nair K, Willison D, Holbrook A, Keshavjee K. Patients' consent preferences regarding the use of their health information for research purposes: a qualitative study. J Health Serv Res Policy 2004 Jan;9(1):22-27. [Medline: 15006236] [doi: 10.1258/135581904322716076]
30. Tracy CS, Dantas GC, Upshur RE. Feasibility of a patient decision aid regarding disclosure of personal health information: qualitative evaluation of the Health Care Information Directive. BMC Med Inform Decis Mak 2004 Sep 10;4(1):13 [FREE Full text] [Medline: 15361257] [doi: 10.1186/1472-6947-4-13]
31. Wilson EV, Lankton NK. Modeling patients' acceptance of provider-delivered e-health. J Am Med Inform Assoc 2004;11(4):241-248 [FREE Full text] [Medline: 15064290] [doi: 10.1197/jamia.M1475]
32. Frohlich J, Karp S, Smith MD, Sujansky W. Retrospective: lessons learned from the Santa Barbara project and their implications for health information exchange. Health Aff (Millwood) 2007;26(5):w589-w591. [Medline: 17670777]
33. Holmquest DL. Another lesson from Santa Barbara. Health Aff (Millwood) 2007;26(5):w592-w594. [Medline: 17670778]
34. Coiera E, Clarke R. e-Consent: The design and implementation of consumer consent mechanisms in an electronic environment. J Am Med Inform Assoc 2004;11(2):129-140 [FREE Full text] [Medline: 14662803] [doi: 10.1197/jamia.M1480]
35. Overhage JM. Health information exchange: ‘lex parsimoniae’. Health Aff (Millwood) 2007;26(5):w595-w597. [Medline: 17670779]

Abbreviations

EHR: electronic health records
HIE: health information exchange
RHIO: regional health information organizations

Edited by G Eysenbach; submitted 14.11.08; peer-reviewed by S Reis, J Marchibroda; comments to author 06.04.09; revised version received 10.06.09; accepted 17.06.09; published 06.08.09.

Please cite as:
Simon SR, Evans JS, Benjamin A, Delano D, Bates DW
Patients' Attitudes Toward Electronic Health Information Exchange: Qualitative Study
J Med Internet Res 2009;11(3):e30
URL: http://www.jmir.org/2009/3/e30/
doi:10.2196/jmir.1164
PMID:19674960

Original Paper

Ability to Generate Patient Registries Among Practices With and Without Electronic Health Records

Adam Wright[1,2], PhD; Elizabeth A McGlinchey[3], BA; Eric G Poon[2], MD, MPH; Chelsea A Jenter[3], MPH; David W Bates[2,3], MD, MSc; Steven R Simon[4], MD, MPH

[1]Clinical Informatics Research & Development, Partners Healthcare System, Boston, MA, USA

[2]Division of General Medicine, Brigham & Women's Hospital, Boston, MA, USA

[3]Clinical and Quality Analysis, Partners Healthcare System, Boston, MA, USA

[4]Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, Boston, MA, USA

Corresponding Author:
Adam Wright, PhD
Division of General Medicine
Brigham and Women's Hospital
75 Francis St
Boston, MA 02115
USA
Phone: +1 781 416 8764
Fax: +1 781 416 8912
Email: awright5@partners.org

Abstract

Background: The ability to generate registries of patients with particular clinical attributes, such as diagnoses or medications taken, is central to measuring and improving the quality of health care. However, it is not known how many providers have the ability to generate such registries.

Objectives: To assess the proportion of physician practices that can construct registries of patients with specific diagnoses, laboratory results, or medications, and to determine the relationship between electronic health record (EHR) usage and the ability to perform registry functions.

Methods: We conducted a mail survey of a stratified random sample of physician practices in Massachusetts in the northeastern United States (N = 1884). The survey included questions about the physicians' ability to generate diagnosis, laboratory result, and medication registries; the presence of EHR; and usage of specific EHR features.

Results: The response rate was 71% (1345/1884). Overall, 79.8% of physician practices reported being able to generate registries of patients by diagnosis; 56.1% by laboratory result; and 55.8% by medication usage. In logistic regression analyses, adjusting for urban/rural location, practice size and ownership, teaching status, hospital affiliation, and specialty, physician practices with an EHR were more likely to be able to construct diagnosis registries (adjusted odds ratio [OR] 1.53, 95% confidence interval [CI] 1.25 - 1.86), laboratory registries (OR 1.42, 95% CI 1.22 - 1.66), and medication registries (OR 2.30, 95% CI 1.96 - 2.70).

Conclusions: Many physician practices were able to generate registries, but this capability is far from universal. Adoption of EHRs appears to be a useful step toward this end, and practices with EHRs are considerably more likely to be able to carry out registry functions. Because practices need registries to perform broad-based quality improvement, they should consider adopting EHRs that have built-in registry functionality.

(J Med Internet Res 2009;11(3):e31) doi:10.2196/jmir.1166

KEYWORDS

Registries; chronic disease; quality assurance, health care; primary health care; family practice; hospital information systems; medical records systems, computerized

Introduction

With the publication of the Institute of Medicine's *Crossing the Quality Chasm* report [1], physicians and health care delivery systems have sharpened their focus on the care of populations and panels of patients. Managing the health of populations demands the ability to identify individuals within the population with specific clinical or sociodemographic characteristics. The core of population management is the patient registry—a list of multiple patients who share some common clinical trait, such as being overdue for a laboratory test, receiving a medication, or having a particular diagnosis.

Registries play a key part in the Chronic Care Model [2,3], a broadly promulgated framework for managing patients with chronic illnesses. In the traditional paradigm of care, a physician reviews the record of, and evaluates, one patient at a time. In the paradigm of the Chronic Care Model and population health management, the physician looks simultaneously at all of his or her patients with a particular diagnosis or combination of diagnoses. This activity is commonly, but not universally, carried out by utilizing an electronic health record (EHR) [4,5].

Laboratory result registries have been used for several purposes, but the most common is detection of patients overdue for screening tests [6-8]. Laboratory result registries can also serve as surrogates for, or adjuncts to, diagnosis-based registries [9,10]. In comparison, medication registries can be used to identify and communicate with patients who are receiving a medication for which some change is recommended or required, as in the case of a safety recall or the new availability of a more effective or less expensive alternative [11-13].

Taken together, this evidence suggests that diagnosis, medication and laboratory registries are essential and effective tools for improving the quality and safety of health care at the population level. A variety of studies at different sites, using different registries and with different disease foci have shown positive results [2-3,6-13]. It is not known, however, how easy or difficult it is for practices in the community to generate such registries. Furthermore, it is apparent that many studies demonstrating the potential for registries to improve care arise in settings with robust electronic health records and computerized provider order entry systems [4,6,8,11,12].

In order to understand better the registry generation capabilities of community ambulatory practices, as well as the relationship between EHR usage and registry capabilities, we undertook the present study. Our goal was to measure physicians' general abilities to perform registry functions in office practice and to explore further the hypothesis that use of electronic health records is associated with the ability to perform registry functions. This study is one aim of a larger study which used a variety of methods, including surveys, focus groups, direct observation, and quality assessment. The goal of the larger study was to measure adoption and use of EHRs in Massachusetts and to compare state-wide adoption to three specific communities in the state that were in the process of implementing community-wide electronic health records with information exchange. Other results of the larger study have been have published previously [14,15].

Methods

Sampling

We carried out a statewide survey of physician practices in Massachusetts between June 2005 and November 2005. We began with a commercial database of physicians in Massachusetts (Folio Associates, Hyannis, MA) which contained contact information for 20,704 physicians practicing at 6308 distinct practice sites in the state. We drew a stratified random sample of practice sites from this database and selected one physician at random from each practice. Our sample was stratified by geography (urban vs nonurban based on county designation, except in the case where there were rural ZIP codes in urban counties, where ZIP codes were used instead), practice size (1 physician, 2 - 3 physicians, 4 - 6 physicians, and 7 or more physicians, exclusive of residents), and practice type (hospital-based primary care, hospital-based specialty/mixed, non-hospital-based primary care, or non-hospital-based specialty/mixed). These sampling characteristics were based on values in the commercial database.

Practices in rural parts of Massachusetts, primary care practices within hospitals, and large practices were oversampled by 100% in our sampling plan to ensure we had adequate representation of these particular practice types. Further details of the sampling plan have been reported previously [14,15].

Survey

Ultimately, we identified a sample of 1884 physician practices across the state of Massachusetts. We mailed a randomly chosen physician at each practice the survey and a US $20 cash incentive to encourage participation. We contacted non-respondents by phone several times and also sent the survey to them two more times (without further cash incentive). The survey contained demographic questions (relating to practice size, teaching, and practice ownership), as well as a variety of questions about quality improvement, practice satisfaction, use of technology, and finances. Some of these questions have been analyzed as part of other aims of this study [14,15]. Two questions (practice size and type) were asked in the survey and were also present in the commercial database used for sampling. In the case that a practice's response differed from the commercial database (eg, because the practice had changed in size), the survey responses were used in analysis as they were more current.

The survey also asked questions about the use of registry functions and availability of an EHR (defined as "an integrated clinical information system that tracks patient health data, and may include such functions as visit notes, prescriptions, lab orders, etc") in the physician's practice. Specifically, physicians were asked to rate the ease of creating lists of patients by diagnosis or health risk (eg, diabetes), by laboratory results (eg, patients with abnormal hematocrit levels), and by medications they currently take (eg, patients on warfarin), using a five-point scale: very easy, somewhat easy, somewhat difficult, very difficult, and cannot generate. Furthermore, physicians were asked if their practice had components of an EHR, specifically defined as "an integrated clinical information system that tracks patient health data, and may include such functions as visit

notes, prescriptions, laboratory orders, etc" and were also surveyed on the availability and use of specific EHR components, such as structured problem or medication lists and electronic reporting and review of laboratory results. The survey instrument and study protocol were reviewed and approved by the Partners Healthcare Institutional Review Board (IRB). The instrument is available as an appendix to this article (See Multimedia Appendix 1).

Analysis

We used SAS 9.1.3 (SAS Institute, Cary, NC), applying weights throughout our analysis to control for both our stratified sampling plan (which included over-sampling of specific groups) and for variable response rates in different strata (specialty, category of practice size, hospital affiliation, and urban/rural location). We used frequency weights (fweights) which are the inverse of the response proportion for each stratum or, equivalently, the weights were determined by taking the population size for each stratum divided by the number of responses. Conceptually, the fweight for a particular response corresponds to the number of physician practices in Massachusetts that this response represents. The ultimate purpose of this weighting strategy was to make our results representative of the population of ambulatory care physician practices in Massachusetts.

We used logistic regression to assess the relationship between the presence of electronic health records in the practice and the ability to create diagnosis registries, laboratory test registries, and medication registries, adjusting each model for the following potential confounding factors:

- urban/rural location
- practice size
- practice ownership (owner, part-owner, non-owner)
- teaching status (whether any students or residents were present in the past year)
- hospital affiliation
- practice type (chosen from solo primary care practice, solo specialty care practice, primary care group/partnership, single specialty group/partnership, multi-specialty group/partnership)

In a secondary analysis limited to practices that had EHRs, we used chi-square tests to examine the relationship between use of key EHR features and the ability to generate each type of registry (diagnosis, laboratory test, medication). For the feature-specific analysis we looked at the effect of problem list use on the ability to generate diagnosis registries, electronic laboratory result review on laboratory test registries, and electronic medication list use on medication registries. In each one of the three cases, the associated feature was chosen for analysis because it was the most directly related feature to the registry type.

Results

A total of 1345 physicians (71%, 1345 of 1884) completed the survey, 1328 by mail and 17 by phone. There were no significant differences between respondents and non-respondents on the sampling characteristics (specialty, practice size, hospital affiliation, and rural practice). The practices reported using a wide variety of commercially available and self-developed EHR systems. Table 1 shows the practice characteristics. Note that some respondents omitted certain questions on the survey so the number of practices does not always add up to 1345. The rural/urban classification was applied based on practice location, so it was available for all practices.

Among the 356 practices which had an EHR and reported its name, a total of 187 (52.5%) used one of the 4 most prevalent systems while the remaining 169 (47.5%) reported using one of 78 other systems that were named. There were also 31 practices that reported having an EHR but did not provide its name—they were still counted as having an EHR for purposes of the analysis.

Table 1. Practice characteristics

	n of practices	% of practices
Practice Location		
Rural	331	24.6
Urban	1014	75.4
Practice Size		
≥ 6 physicians	504	37.5
3 - 5 physicians	280	20.8
1 - 2 physicians	383	28.5
Practice Ownership		
Full owner	460	34.2
Part owner	172	12.8
Non owner	542	40.3
Practice Involvement in Teaching		
Involved in teaching	552	41.0
Not involved in teaching	628	46.7
Hospital Affiliation		
Hospital based practice	360	26.8
Non-hospital based practice	974	72.4
Practice Type		
Solo primary care practice	154	11.4
Solo specialty care practice	192	14.3
Primary care group/partnership	309	23.0
Single specialty group/partnership	338	25.1
Multi-specialty group/partnership	177	13.2
EHR Usage		
Yes	387	28.8
No	794	59.0

Overall, 79.8% of physicians reported that their practices could generate registries of patients with a particular diagnosis; 56.1% could generate registries of patients with a specific laboratory result; and 55.8% could generate registries of patients taking a particular medication. Among physicians who reported that their practices were able to generate registries, the reported ease with which such registries could be generated varied greatly, as shown in Table 2. While 38.9% of physician practices that could generate diagnosis registries said their practice could do it easily or very easily, considerably fewer said that it was easy or very easy to generate registries based on laboratory test results (14.5%) or medications (17.8%).

Table 2. Ease or difficulty of generating registries of patients based on diagnosis, laboratory result and medication use[a]

Ease or Difficulty	Diagnosis registry	Laboratory result registry	Medication registry
Very Easy	15.0%	6.5%	7.4%
Somewhat Easy	23.9%	8.0%	10.4%
Somewhat Difficult	21.7%	16.0%	14.1%
Very Difficult	19.2%	25.6%	23.9%
Cannot Generate	20.2%	43.9%	44.2%

[a]Weighted proportion of physicians reporting that their practice can generate each kind of registry with a particular ease or difficulty.

Table 3 shows the proportion of physician practices that were capable of carrying out registry functions, stratified by pre-specified subgroups of interest. For all three registry types, providers with EHRs were significantly more likely to be able

to perform registry functions than providers using non-electronic record systems ($P < .001$ for all three registry types). Also, larger practices, practices involved in teaching, hospital-based practices, and primary care practices were more likely to be able to generate registries.

In logistic regression analyses controlling for urban/rural location, practice size, practice ownership, teaching status, hospital affiliation, and practice type, the relationship between the presence of EHR and the ability to carry out each registry function remained robust. EHR adopters were more likely than non-adopters to be able to develop registries based on diagnosis (adjusted odds ratio [OR] 1.53, 95% confidence interval [CI] 1.25 - 1.86), laboratory results (OR 1.42, 95% CI, 1.22 - 1.66), and medications (OR 2.30, 95% CI, 1.96 - 2.70). Rural location, practice size, practice ownership, hospital affiliation, and practice type also remained significant correlates of one or more registry capability in the multivariate analyses (Table 4).

Table 3. Ability to perform registry functions according to practice characteristics (these data are weighted but not adjusted for confounding factors)

	Percentage of physicians able to perform function		
	Diagnosis registry	Laboratory result registry	Medication registry
Practice Location [a]			
Rural	78.9%	58.9%	62.6%
Urban	79.9%	55.9%	55.3%
Practice Size [b]			
≥ 6 physicians	82.4%	60.1%	59.2%
3 - 5 physicians	84.8%	63.4%	53.9%
1 - 2 physicians	75.8%	50.6%	54.8%
Practice Ownership [c]			
Full owner	77.8%	51.2%	54.1%
Part owner	81.6%	49.9%	48.3%
Non owner	81.6%	66.6%	61.7%
Practice Involvement in Teaching [d]			
Involved in teaching	85.0%	63.2%	61.8%
Not involved in teaching	76.9%	52.3%	52.5%
Hospital Affiliation [e]			
Hospital based practice	81.9%	69.9%	65.2%
Non-hospital based practice	79.4%	54.5%	54.8%
Practice Type [d]			
Solo primary care practice	74.2%	57.2%	61.7%
Solo specialty care practice	77.1%	47.0%	55.8%
Primary care group/partnership	88.6%	71.4%	67.2%
Single specialty group/partnership	78.5%	56.0%	44.5%
Multi-specialty group/partnership	84.5%	58.7%	59.7%
EHR Usage [d]			
Yes	85.9%	66.7%	71.6%
No	78.0%	52.9%	51.1%

[a] $P = .64$ using a chi-square test for diagnosis registry functions, $P = .25$ for laboratory registry functions, $P = .005$ for medication registry functions.

[b] $P < .001$ for diagnosis and laboratory registry functions, $P = .009$ for medication registry functions.

[c] $P = .004$ for diagnosis registry functions and $P < .001$ for laboratory and medication registry functions.

[d] $P < .001$ for all three registry types.

[e] $P = .153$ for diagnosis registry functions, $P < .001$ for medication and laboratory registry functions.

Table 4. Multivariate correlates of registry function capability

	Diagnosis registry OR (95% CI)	Laboratory result registry OR (95% CI)	Medication registry OR (95% CI)
Practice Location			
Rural	0.95 (0.73 - 1.23)	1.08 (0.86 - 1.34)	1.32 (1.05 - 1.65)
Urban	1 (Ref)	1 (Ref)	1 (Ref)
Practice Size			
≥ 6 physicians	1.09 (0.85 - 1.41)	0.92 (0.74 - 1.14)	1.10 (0.89 - 1.37)
3 - 5 physicians	1.62 (1.29 - 2.03)	1.39 (1.16 - 1.68)	1.20 (1.00 - 1.44)
1 - 2 physicians	1 (Ref)	1 (Ref)	1 (Ref)
Practice Ownership			
Full owner	1.28 (1.04 - 1.58)	0.71 (0.60 - 0.85)	0.65 (0.54 - 0.78)
Part owner	1.19 (0.92 - 1.53)	0.53 (0.43 - 0.65)	0.71 (0.58 - 0.87)
Non owner	1 (Ref)	1 (Ref)	1 (Ref)
Practice Involvement in Teaching			
Involved in teaching	1.42 (1.19 - 1.70)	0.98 (0.85 - 1.13)	1.08 (0.94 - 1.25)
Not involved in teaching	1 (Ref)	1 (Ref)	1 (Ref)
Hospital Affiliation			
Hospital based practice	1.09 (0.85 - 1.39)	1.87 (1.52 - 2.30)	1.35 (1.10 - 1.66)
Non-hospital-based practice	1 (Ref)	1 (Ref)	1 (Ref)
Practice Type			
Solo primary care practice	0.67 (0.47 - 0.96)	1.23 (0.92 - 1.65)	1.97 (1.46 - 2.66)
Solo specialty care practice	0.84 (0.60 - 1.17)	0.90 (0.68 - 1.18)	1.78 (1.35 - 2.34)
Primary care group/partnership	1.47 (1.07 - 2.02)	1.90 (1.50 - 2.42)	1.61 (1.26 - 2.04)
Single specialty group/partnership	0.70 (0.54 - 0.91)	0.91 (0.74 - 1.13)	0.62 (0.50 - 0.77)
Multi-specialty group/partnership	1 (Ref)	1 (Ref)	1 (Ref)
EHR Usage			
Yes	1.53 (1.25 - 1.86)	1.42 (1.22 - 1.66)	2.30 (1.96 - 2.70)
No	1 (Ref)	1 (Ref)	1 (Ref)

OR = adjusted odds ratio; CI = confidence interval
Ref: Reference Category

We also observed a relationship between use of key EHR features and ability to perform related registry functions. Specifically, within the group of physicians who had access to an EHR in their practice, 90.4% of physicians who reported using an electronic problem list at least some of the time had the ability to perform diagnosis registry functions, while only 67.7% of physicians using an EHR without access to an electronic problem list could perform these functions ($P < .001$). Similarly, while 75.2% of physicians who used an electronic medication list could perform medication registry functions, only 53.0% of physicians who used an EHR without a medication list reported they could perform them ($P < .001$). Finally, while 71.5% of physicians who used their EHR to view laboratory results could perform laboratory registry functions, only 33.9% of physicians whose EHR could not be used to view laboratory results reported they could perform these functions ($P < .001$).

Discussion

Principal Results

While many studies have demonstrated the value of being able to perform registry functions for improving the quality and safety of health care [2-3,6-13], few data are available regarding the capability of carrying out registry functions in community-based practices. In this study, about 80% of physicians reported being able to generate registries of patients according to diagnoses, but nearly half of all physicians in ambulatory care practice in Massachusetts could not create registries by medication or laboratory result.

Having EHRs was strongly associated with the reported ability to generate registries based on diagnosis, laboratory test result, or medication, but even among EHR users, 14% could not generate lists of diagnoses, 33% could not do so for laboratory tests, and 28% could not do so for medications. Furthermore,

we found that physicians who reported active use of key EHR functions were considerably more likely to report being able to generate registries. Thus, these data suggest that EHRs appear important for delivering care using registries, and that most but not all EHR users could generate registries using their electronic records.

We are uncertain about why some physician practices with EHRs were unable to create registries. Many of these practices reported using EHRs which we knew to have this capability. It is likely that at least some of the EHR users who reported an inability to generate registries actually have the ability to generate them using their EHR, but are unaware of the feature. This suggests that improvements in documentation, training, and ease of use to help more physicians take advantage of the existing registry capabilities of their EHRs may be useful.

We were also a bit surprised by the relatively high proportion of EHR non-users who reported being able to generate registries. We are uncertain as to the mechanism employed by these users, since our survey did not ask them to explain how they were generating registries. These users may have been using retrospective chart review, prospective tracking, or analysis based on administrative data (such as billing and claims data in a practice management system). Each of these methods has a significant downside. Retrospective chart review is extremely time-consuming and error prone; prospective tracking requires criteria to be developed in advance; and non-clinical data are often less sensitive and specific than clinical data.

Taken together, our findings raise concerns about the ability of many ambulatory care practices, particularly the majority of practices without EHRs, to provide effective care for their patients on a population level. Physicians and practices need to consider population-level care management, not only as an essential component to effective practice within the Chronic Care Model [2,3], but also as a necessary tool for responding to the exigencies of forces driving quality improvement, such as pay-for-performance. Without EHRs, and without active and effective use of key features in robust EHRs [16], physicians and practices will have much greater difficulty in efficiently delivering safe and effective care for patients with acute and chronic problems.

EHRs can and should either include the inbuilt ability to query across patient records by a variety of criteria, or support extracting patient data which can be fed into other applications which do this. The ability to generate registries by diagnosis is common in many commercial EHR systems. In fact, it is a requirement of the 2007 and 2008 Certification Commission for Health Information Technology (CCHIT) criteria for ambulatory EHR certification [17]. However, such certification is voluntary, and many commercial products are not certified, or are certified under the older 2006 criteria, which did not require the availability of registry functions. CCHIT certification is valid for only three years, however, so as products certified under the 2006 criteria are re-certified under the current criteria, this gap will close. Also, although certification is voluntary, there are increasing incentives (or, in some limited cases, mandates) encouraging use of CCHIT-certified products, so the use of non-certified products which may not have registry capabilities is likely to be lessened over time [18]. Furthermore, just because an EHR can be used to generate a registry, the software does not always make it easy or user-friendly for a provider to do so. However, even when EHRs do not come "pre-loaded" with the ability to generate registries, the organization of patient information in structured fields in an electronic format facilitates the creation of registries more easily than in paper-based systems.

It is also worth noting that the ability to create registries in an EHR is generally predicated on the use of structured documentation features within the EHR. For example, if a clinician documents patient problems only in unstructured clinical notes, it is nearly impossible in most commercially available EHRs to build medication registries based on this unstructured information. However, if the clinician uses a structured medication list with a controlled medication vocabulary, generating such a registry becomes much easier.

Finally, our survey was conducted in 2005. Since then, adoption of EHRs in ambulatory practices has increased somewhat [19], though is still far from universal. Since our data suggest that physicians with EHRs are more able to generate registries than non-users, we expect that the current ability of physicians to generate such registries is likely now higher than it was in 2005. Moreover, the capabilities of EHR systems, as a rule, increase over time, so we likewise expect that EHR users are more likely, today, to be able to generate registries than they were in 2005.

Implications

Our findings have several important implications for physicians, for the health care system, and for developers of electronic health records. Because our findings suggest that the ability to generate registries is less than universal, and because generating registries is integral to quality and safety enhancing activities, it may be necessary to take steps to increase these capabilities in office practice. Providing physicians and practices with training and activities to increase awareness of the role of generating registries may be beneficial, but these changes are unlikely to be sufficient. Incentives also likely play an important role; physicians are more likely to adopt and use registry functions if financial incentives are in place to do so [20]. As pay-for-performance initiatives become more prevalent, physicians are likely to embrace the use of registries as a foundation for building practice-level population management capabilities. Our findings also show that EHR users are more likely to be able to generate registries than non-users, so incentives aimed at EHR adoption alone are also likely to have a positive effect on registry capabilities.

Limitations

The study has several limitations. First, our survey was limited to physicians in ambulatory care in Massachusetts, and the results may not be generalizable to other states or regions. However, given that Massachusetts is a state in which more than 45% of physicians have EHRs [15], the large proportions of physicians and practices reporting inability to perform registry functions in this study are likely to be even larger in other states, where EHR adoption has lagged. As such, the need to consider

efforts to adopt EHRs and expand their use may have even greater imperative in other regions.

Another limitation is the self-reported nature of survey studies such as ours. We are, of course, not truly measuring physicians' abilities to generate registries, but instead their self-reports of the ability to generate three specific types of registries. This raises the possibility of social desirability bias influencing physicians' responses to survey questions. However, if this bias were present, then one might expect that physicians overestimated their abilities to perform registry functions, which would mean that even fewer physicians than reported have the ability to generate registries. Also, our survey asked providers how easily they could perform registry functions but did not ask how frequently they actually did perform such functions, or for what purposes and with what results. It is important to note that, among those practices that reported the ability to make these registries, we do not know the frequency with which they did so. It is possible that some practices, although able to create the registries, never actually do. Future qualitative and quantitative studies should explore how physicians and practices are using registries, as well as the barriers to, and facilitators of, effective use of these important tools. Intervention studies will then be able to test strategies for improving physicians' use of registries to improve quality of care and patient safety.

Finally, our survey was limited to registries of diagnoses, medications and lab results. Other types of registries exist, such as registries of patients receiving a particular surgical procedure, which are often used for tracking quality and outcomes, tumor registries, and registries of implanted devices, such as implantable cardiac defibrillators or pacemakers, which are important in the event of a recall. Such registries are generally used in specific specialties, and it would be worthwhile to survey specialists about their use of these special-purpose registries.

Conclusion

While registry functions are available to many physicians, their availability is far from universal. Because generating registries is essential for population health management activities associated with improved quality, safety, and efficiency, it is important that their availability increase. Adoption of EHRs appears to be a useful step toward this end, since practices with EHRs are significantly more likely to be able to carry out registry functions. CCHIT should intensify and expand its requirements for registry function capabilities, and commercial EHR products without these capabilities should be extended to provide them. Health policy makers and health care leaders can then develop and disseminate strategies for using registries for improving patient safety and the quality of health care.

Acknowledgments

This study was funded in part by the United States Agency for Healthcare Research and Quality cooperative agreement #1UC1HS015397-01 and the Massachusetts e-Health Collaborative. The Agency for Healthcare Research and Quality and the Massachusetts e-Health Collaborative had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality or the Massachusetts e-Health Collaborative.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Massachusetts Survey of Physicians and Computer Technology

[PDF file (Adobe PDF), 134 KB - jmir_v11i3e31_app1.pdf]

References

1. ; Committee on Quality of Health Care in America, Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: The National Academies Press; 2001.
2. Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness: the chronic care model, Part 2. JAMA 2002 Oct 16;288(15):1909-1914. [Medline: 12377092] [doi: 10.1001/jama.288.15.1909]
3. Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness. JAMA 2002 Oct 9;288(14):1775-1779. [Medline: 12365965] [doi: 10.1001/jama.288.14.1775]
4. Thomas KG, Thomas MR, Stroebel RJ, et al. Use of a registry-generated audit, feedback, and patient reminder intervention in an internal medicine resident clinic—a randomized trial. J Gen Intern Med 2007;22(12):1740-1744. [doi: 10.1007/s11606-007-0431-x]
5. Wagner EH. Population-based management of diabetes care. Patient Educ Couns 1995 Sep;26(1-3):225-230. [Medline: 7494727] [doi: 10.1016/0738-3991(95)00761-N]
6. Dorr DA, Wilcox A, Donnelly SM, Burns L, Clayton PD. Impact of generalist care managers on patients with diabetes. Health Serv Res 2005 Oct;40(5 Pt 1):1400-1421 [FREE Full text] [Medline: 16174140] [doi: 10.1111/j.1475-6773.2005.00423.x]

7. Jung E, Li Q, Mangalampalli A. Report Central: quality reporting tool in an electronic health record. In: AMIA Annu Symp Proc. 2006 Presented at: p. 971. [Medline: 17238590]
8. Olsha-Yehiav M, Einbinder JS, Jung E, et al. Quality Dashboards: technical and architectural considerations of an actionable reporting tool for population management. AMIA Annu Symp Proc 2006:1052 [FREE Full text] [Medline: 17238671]
9. McCulloch DK, Price MJ, Hindmarsh M, Wagner EH. A population-based approach to diabetes management in a primary care setting: early results and lessons learned. Eff Clin Pract 1998;1(1):12-22 [FREE Full text] [Medline: 10345254]
10. Miller DR, Safford MM, Pogach LM. Who has diabetes? Best estimates of diabetes prevalence in the Department of Veterans Affairs based on computerized patient data. Diabetes Care 2004 May;27(Suppl 2):B10-B21 [FREE Full text] [Medline: 15113777] [doi: 10.2337/diacare.27.suppl_2.B10]
11. Cheetham TC, Chan J, Benson V, Richmond C, Levin E, Campen D. Successful conversion of patients with hypercholesterolemia from a brand name to a generic cholesterol-lowering drug. Am J Manag Care 2005 Sep;11(9):546-552 [FREE Full text] [Medline: 16159044]
12. Jain A, Atreja A, Harris CM, Lehmann M, Burns J, Young J. Responding to the rofecoxib withdrawal crisis: a new model for notifying patients at risk and their health care providers. Ann Intern Med 2005 Feb 1;142(3):182-186 [FREE Full text] [Medline: 15684206]
13. Usher-Smith JA, Ramsbottom T, Pearmain H, Kirby M. Evaluation of the cost savings and clinical outcomes of switching patients from atorvastatin to simvastatin and losartan to candesartan in a Primary Care setting. Int J Clin Pract 2007 Jan;61(1):15-23. [Medline: 17229176] [doi: 10.1111/j.1742-1241.2006.01217.x]
14. Simon SR, Soran CS, Kaushal R. Physicians' usage of key functions in electronic health records from 2005 to 2007: a statewide survey. J Am Med Inform Assoc 2009 Apr 23 [Epub ahead of print. [Medline: 19390104]
15. Simon SR, Kaushal R, Cleary PD, et al. Physicians and electronic health records: a statewide survey. Arch Intern Med 2007;167(5):507-512. [Medline: 17353500] [doi: 10.1001/archinte.167.5.507]
16. Frieden TR, Mostashari F. Health care as if health mattered. JAMA 2008 Feb 27;299(8):950-952. [Medline: 18314438] [doi: 10.1001/jama.299.8.950]
17. ; Certification Commission for Health Information Technology. Final Criteria for 2007 Certification of Ambulatory EHRs. Certification Commission for Health Information Technology. 2007. URL: http://www.cchit.org/files/Ambulatory_Domain/CCHIT_Ambulatory_FUNCTIONALITY_Criteria_2007_Final_16Mar07.pdf [accessed 2008 Feb 19] [WebCite Cache ID 5cOCkQmJL]
18. ; Certification Commission for Health Information Technology. The CCHIT Incentive Index. Certification Commission for Health Information Technology. 2008. URL: http://ehrdecisions.com/wp-content/files/CCHITIncentiveIndex20080925.pdf [accessed 2009 Feb 28] [WebCite Cache ID 5ew7Yky6p]
19. DesRoches CM, Campbell EG, Rao SR, et al. Electronic health records in ambulatory care—a national survey of physicians. N Engl J Med 2008;359(1):50-60. [Medline: 18565855] [doi: 10.1056/NEJMsa0802005]
20. Campbell S, Reeves D, Kontopantelis E, Middleton E, Sibbald B, Roland M. Quality of primary care in England with the introduction of pay for performance. N Engl J Med 2007 Jul 12;357(2):181-190. [Medline: 17625132] [doi: 10.1056/NEJMsr065990]

Abbreviations

CCHIT: Certification Commission for Health Information Technology
EHR: electronic health record
IRB: institutional review board

Edited by K El Emam; submitted 21.11.08; peer-reviewed by V Harrop, S Woods; comments to author 22.02.09; revised version received 04.05.09; accepted 23.05.09; published 10.08.09.

Please cite as:
Wright A, McGlinchey EA, Poon EG, Jenter CA, Bates DW, Simon SR
Ability to Generate Patient Registries Among Practices With and Without Electronic Health Records
J Med Internet Res 2009;11(3):e31
URL: http://www.jmir.org/2009/3/e31/
doi:10.2196/jmir.1166
PMID:19674961

Original Paper

Associations of Leisure-Time Internet and Computer Use With Overweight and Obesity, Physical Activity and Sedentary Behaviors: Cross-Sectional Study

Corneel Vandelanotte[1], PhD; Takemi Sugiyama[2], PhD; Paul Gardiner[2], BSc; Neville Owen[2], PhD

[1]Institute for Health and Social Sciences Research, Central Queensland University, Rockhampton, Australia

[2]Cancer Prevention Research Centre, University of Queensland, Herston, Australia

Corresponding Author:
Corneel Vandelanotte, PhD
Central Queensland University
Institute for Health and Social Sciences Research
Building 18
Bruce Highway
Rockhampton QLD 4702
Australia
Phone: +61 7 49232042
Fax: +61 7 49306402
Email: c.vandelanotte@cqu.edu.au

Abstract

Background: Internet and computer use are increasingly common leisure-time sedentary behaviors, which have the potential to impact negatively on health outcomes. However, little is known about the extent to which adults' Internet and computer use is associated with weight status and time spent in leisure-time physical activity.

Objective: The objective is to examine associations of leisure-time Internet and computer use with overweight and obesity, leisure-time physical activity, and other sedentary behaviors.

Methods: Participants (2650 adults living in Adelaide, Australia) completed a mail-back questionnaire including items on their height and weight, past seven day recall of leisure-time physical activity, Internet and computer use, and other leisure-time sedentary behaviors. Leisure-time Internet and computer use was categorized into no use, low use (less than three hours per week), or high use (three hours or more per week).

Results: Participants with low leisure-time Internet and computer use had the highest levels of educational attainment and employment, and engaged in less other sedentary behaviors when compared to participants with no or high Internet and computer use. Multinomial logistic regression, adjusted for gender, age, employment, education, other sedentary behaviors and physical activity, determined that participants with a high leisure-time Internet and computer use were 1.46 (95% CI = 1.10 - 1.93) times more likely to be overweight (BMI≥25 and < 30 kg/m^2) and 2.52 times more likely (95% CI = 1.82 - 3.52) to be obese (BMI≥30 kg/m^2), compared to those who reported no Internet and computer use in their leisure-time. Adults with high leisure-time Internet and computer use were more likely to be overweight or obese even if they were highly active in their leisure time (OR = 1.86; 95% CI = 1.21 - 2.88), as compared to participants who did not use the Internet or computer. Leisure-time physical activity levels were largely independent of Internet and computer use.

Conclusion: These findings suggest that, apart from nutritional and physical activity interventions, it may also be necessary to decrease time spent in sedentary behaviors, such as leisure-time Internet and computer use, in order to reduce the prevalence of overweight and obesity. Future Internet interventions to reduce weight or increase physical activity may need to differentiate between participants with different levels of Internet use in order to increase their effectiveness. Longitudinal studies are required to examine further the potential causal relationships between the development of overweight and specific sedentary behaviors such as Internet and computer use.

(J Med Internet Res 2009;11(3):e28) doi:10.2196/jmir.1084

KEYWORDS

Internet; computer; overweight; obese; BMI; body mass index; physical activity; human activities; sedentary behavior; leisure-time; leisure activities

Introduction

Many studies have shown that physical inactivity is associated with higher levels of overweight and obesity and that physical activity is essential in the prevention and treatment of overweight and obesity [1,2]. Recently, this evidence has led to the development of specific physical activity guidelines for overweight and obese people [3,4] which state that 60 to 90 minutes of daily moderate to vigorous physical activity are necessary to lose weight or to maintain weight loss.

There are strong adverse associations between time spent in sedentary behaviors and different health indicators [5-7], including the increased likelihood of being overweight or obese [6,8-10]. It is now generally accepted that "sedentariness" and physical (in)activity are two distinct classes of behavior, each with their own determinants [11]. They have independent effects on total energy expenditure, weight, and metabolic variables [10].

However, most of the evidence on associations between sedentary behavior and health outcomes, such as weight status and levels of physical activity, is specific to time spent watching television [5,6,8,10,12], which is the most commonly studied leisure-time sedentary behavior. Associations of health outcomes with other sedentary behaviors such as Internet and computer use remain largely unknown. In Australia, more than 70% of the population has access to the Internet, and this number is still increasing [13]. Internet and computer use are increasingly common leisure-time sedentary behaviors [13,14] which have the potential to impact negatively on health, independent of other sedentary behaviors. Extensive use of the Internet and computers may also displace time spent in leisure-time physical activity.

Several studies have examined the associations between leisure-time Internet and computer use, physical activity, and levels of overweight/obesity in children and adolescents, with inconsistent outcomes. Some studies show that high leisure-time Internet and computer use is associated with higher Body Mass Index (BMI) and lower physical activity levels [15-18]; other studies are not able to confirm this [19-21]. However, to our knowledge no studies have evaluated these relationships in adults.

Further, little is known about how sedentary behaviors relate to each other. In relation to health outcomes, it is important to know whether high leisure-time Internet and computer use is a marker for high levels of other sedentary behaviors. It may be that leisure-time Internet and computer use is related to poor health outcomes due to its association with a broader pattern of sedentary behavior. A study by Sugiyama et al [22] demonstrated that, in women, time spent watching TV was associated positively with time in other sedentary behaviors. To our knowledge, no studies have evaluated how Internet and computer use relate to other sedentary behaviors.

The aim of this study is to examine associations of Internet and computer use, specifically in leisure time (excluding occupational computer use), with overweight and obesity, leisure-time physical activity, and other sedentary behaviors, in a large socially-diverse sample of Australian adults.

Methods

Participants and Procedures

This study is part of an observational epidemiological study (PLACE: Physical Activity in Localities and Community Environments) conducted in urban areas of Adelaide, Australia during 2003 - 2004. Detailed methods of the study have been described elsewhere [23,24]. Briefly, a study sample was drawn from residential addresses within 32 neighbourhoods which are known to vary in socio-economic status. In each neighbourhood, 250 addresses were randomly selected and sent a letter of invitation to participate. Eligible respondents (English speaking, aged between 20 and 65, residing in private dwellings, and able to walk without assistance) who agreed to participate were mailed a survey that included questions about Internet and computer use, other sedentary behaviors, physical activity, body weight, height, and socio-demographic characteristics. Participant recruitment and data collection were handled in a series of waves, between July 2003 and June 2004, in order to obtain data from respondents across the range of seasons. A total of 2650 eligible participants returned the questionnaire. The return rate for those who completed the survey, as a proportion of those known to be contacted was 74.2%. The Behavioral and Social Sciences Ethics Committee of the University of Queensland approved the study.

Measures

Leisure-Time Internet and Computer Use

Participants reported leisure-time Internet and computer use as part of a measurement tool assessing total leisure-time sedentary behavior in the last seven days. For each sedentary activity, the tool asks "How many days did you do this activity in the last 7 days", followed by "On average, how many minutes did you do this activity on the days you did it". This instrument has been shown to have acceptable reliability and validity, especially for Internet and computer use [25]. To evaluate the validity of this measure, three-day sedentary behavior logs were collected from 130 participants. The Spearman rank-order correlations showed that, compared with the three-day log, the last seven day recall measure was acceptable for Internet and computer use ($\rho = 0.6$). Test-retest reliability was evaluated in a sample of 145 participants. Intra-Class Correlations (ICC) indicated acceptable agreement for Internet and computer use (ICC = 0.62). The amount of leisure-time Internet and computer use was split into three categories: no, low (less than three hours per week), and high (three hours or more per week) Internet and computer use.

Overweight and Obesity

Body Mass Index (BMI; kg/m^2) was calculated using self-reported height and weight and was categorized as either normal weight (< 25 kg/m^2); overweight (≥ 25 and < 30 kg/m^2) or obese (≥ 30 kg/m^2). BMI was also used as a continuous variable.

Leisure-time physical activity was assessed using the long-form (31 items) International Physical Activity Questionnaire (IPAQ) [26]. Participants reported the number of days per week and the time spent per day on walking, as well as vigorous-intensity and moderate-intensity leisure-time activities, during the last seven days. The amount of leisure-time physical activity was split into three categories: low (less than one hour per week), medium (between one and three hours per week), and high (three or more hours per week) leisure-time physical activity.

Other Leisure-Time Sedentary Behaviors

The instrument applied to measure Internet and computer use [25] was also applied to measure the total of "other leisure-time sedentary behaviours [*sic*]". This variable included time spent: reading, sitting when talking to friends or listening to music, talking on the phone, playing video games, watching television, and driving or riding in a car; it did not include Internet and computer use. The amount of other sedentary behaviors was split into: low (less than 2.5 hours per day), medium (between 2.5 and 5 hours per day), and high (5 or more hours per day).

Statistical Analysis

One-way ANOVA and Chi-square tests were used for analysing differences in socio-demographic factors according to different categories of Internet and computer use. Multinomial logistic regression analyses were conducted to estimate associations of Internet and computer use with overweight and obesity (model 1), leisure-time physical activity (model 2), and other sedentary behaviors (model 3). The models were adjusted for age, gender, employment, level of education, overweight and obesity (only in models 2 and 3), other sedentary behaviors (only in models 1 and 2), and leisure-time physical activity (only in models 1 and 3). Binary logistic regression was conducted to estimate the odds ratios of being overweight or obese, comparing levels of Internet and computer use (no, low, and high Internet and computer use) and physical activity (low, medium, and high leisure-time physical activity). This model was adjusted for age, gender, education, employment, and other sedentary behaviors. Analyses were conducted using SPSS version 13.0. Significance was accepted at an alpha level of 0.05.

Results

Sample size was 2532 (1554 women, 978 men), after excluding missing values for Internet and computer use (n = 118). Average leisure-time Internet and computer use was 125.3 minutes per week (SD: 273.3). Table 1 shows socio-demographic characteristics for the total sample according to Internet and computer use. Participants with low Internet and computer use had the highest levels of educational attainment and employment, were younger, and participated in less other sedentary behaviors compared to participants with either high Internet and computer use or no use. Participants with high Internet and computer use had the highest BMI compared to the other groups and were more likely to be male.

Table 1. Sample characteristics for total group and according to computer and Internet use categories (mean ± SD or %)[a]

	Total Sample (N = 2650)	No Internet or computer use (N = 1093)	Low Internet and computer use (N = 983)	High Internet and computer use (N= 456)	*P*-value
Sex (% female)	64.0	68.7	65.4	50.1	< .001
Age (yr)	44.5 ± 12.3	45.8 ± 11.8	42.8 ± 12.4	44.1 ± 12.7	< .001
College or university degree (%)	46.3	36.9	55.8	51.0	< .001
Employed (%)	69.2	62.9	77.0	67.6	< .001
Leisure-time physical activity (hrs/week)	3.3 ± 4.5	3.1 ± 4.7	3.6 ± 4.5	3.2 ± 4.3	ns
Body Mass Index (kg/m^2)	26.3 ± 6.4	25.9 ± 5.9	25.9 ± 5.6	27.5 ± 8.3	< .001
Other sedentary behaviours (hrs/week)	27.6 ± 16.9	27.5 ± 18.2	25.6 ± 14.3	32.2 ± 18.2	< .001

[a]Chi-squared and one-way ANOVA were used to examine differences between categories; ns is not significant.

Overweight and Obesity by Internet and Computer Use

As shown in Table 2, leisure-time Internet and computer use was significantly associated with overweight and obesity. Compared to participants that reported no Internet and computer use, participants with low Internet and computer use were 1.3 times more likely to be overweight and 1.4 times more likely to be obese, and participants with high Internet and computer use were 1.5 times more likely to be overweight and 2.5 times more likely to be obese.

Leisure-Time Physical Activity by Internet and Computer Use

Leisure-time physical activity was largely independent of leisure-time Internet and computer use. However, participants with low Internet and computer use were 1.3 times more likely to do more than three hours of leisure-time physical activity, when compared to non-users.

Other Leisure-Time Sedentary Behaviors by Internet and Computer Use

Participants with low and high leisure-time Internet and computer use were respectively 1.8 and 2.5 times more likely to engage in more than five hours of other sedentary behaviors per day, when compared to participants that did not use the Internet and computer.

Table 2. Multinomial logistic regression models predicting overweight or obesity, leisure-time physical activity, and other sedentary behaviors by computer and Internet use[a]

		OR (95% CI)	OR (95% CI)
Model 1: Weight status	**Normal weight (N = 1187)**	**Overweight (N = 783)**	**Obese (N = 442)**
Internet and computer use			
No use	Reference	1.00	1.00
Low use	Category	1.30 (1.01 - 1.56)[b]	1.45 (1.10 - 1.92)[c]
High use		1.46 (1.10 - 1.93)[c]	2.52 (1.81 - 3.51)[c]
Model 2: Leisure-time Physical Activity (LTPA)	**Low LTPA (N = 1375)**	**Medium LTPA (N = 721)**	**High LTPA (N = 429)**
Internet and computer use			
No use	Reference	1.00	1.00
Low use	Category	1.12 (0.87 - 1.44)	1.28 (1.02 - 1.60)[b]
High use		0.81 (0.60 - 1.12)	0.83 (0.63 - 1.11)
Model 3: Other Leisure-time Sedentary Behaviors (LTSB)	**Low LTSB (N = 656)**	**Medium LTSB (N = 1045)**	**High LTSB (N = 557)**
Internet and computer use			
No use	Reference	1.00	1.00
Low use	Category	1.24 (0.99 - 1.55)	1.79 (1.30 - 2.46)[c]
High use		0.99 (0.75 - 1.30)	2.50 (1.75 - 3.57)[c]

[a]Regression models were adjusted for gender, age, employment, educational attainment, other sedentary behaviors, leisure time physical activity, and BMI.

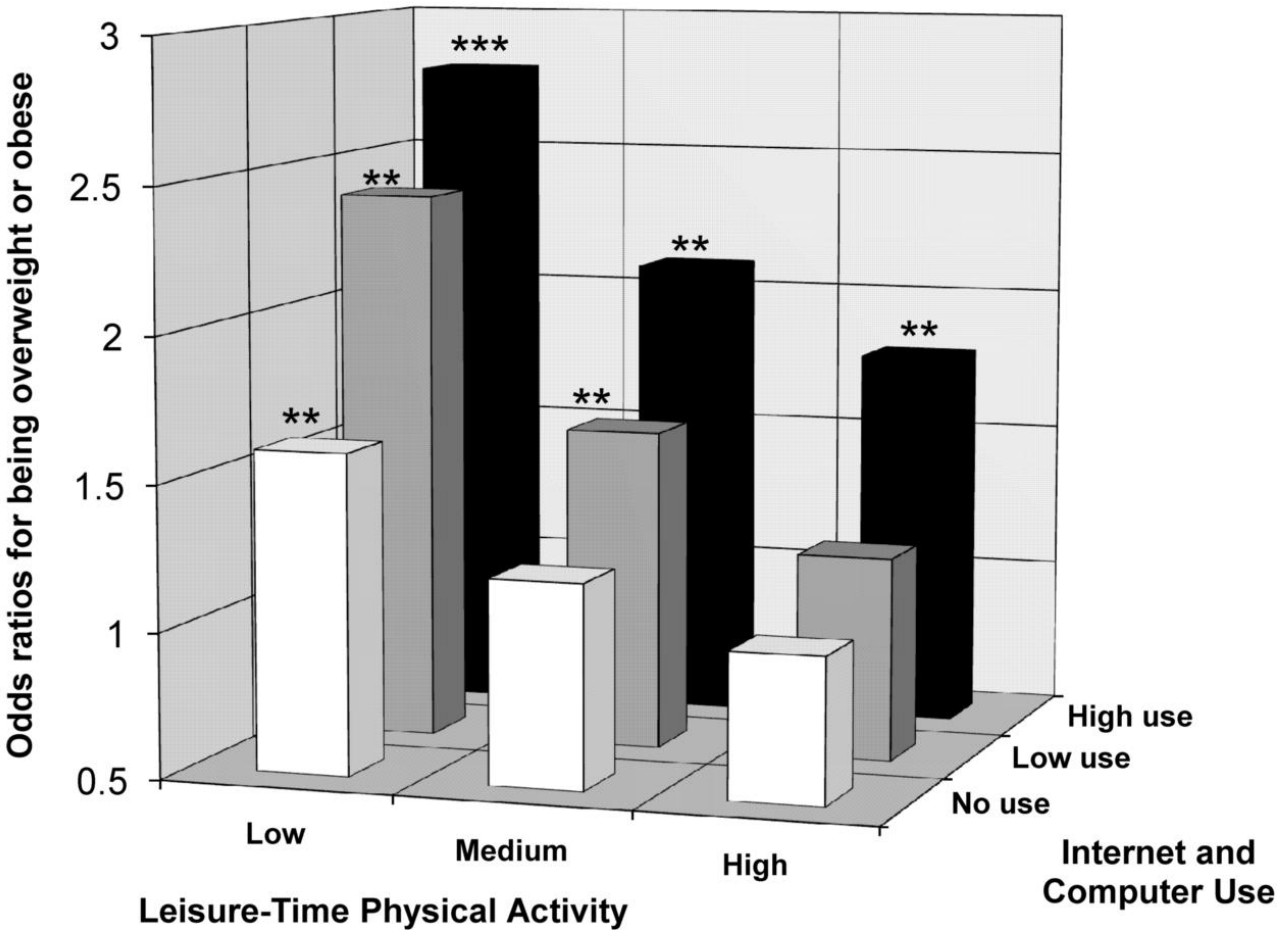
[b]$P < .05$

[c]$P < .001$

Overweight and Obesity by Leisure-Time Physical Activity by Internet and Computer Use

Figure 1 shows the odds ratios for being overweight or obese were higher when leisure-time physical activity was lower and/or when Internet and computer use was higher. For example, participants with low leisure-time physical activity and high Internet and computer use were 2.77 times (95% CI = 1.86 - 4.12) more likely to be overweight or obese, and participants with high leisure-time physical activity and high Internet and computer use were 1.86 times (95% CI = 1.21 - 2.88) more likely to be overweight or obese, as compared to participants with high leisure-time physical activity and participants who did not use the Internet and computer.

Figure 1. The odds ratios for being overweight or obese (BMI ≥ 25), according to combined categories of Internet and computer use (no, low, and high leisure-time Internet and computer use) and physical activity (low, medium, and high leisure-time physical activity). The reference category is having high leisure-time physical activity and not using the Internet and computer, for which the odds ratios are equal to 1. The significance levels on top of the figure bars are differences in relation to the reference category: ** P < .01; *** P < .001



Discussion

The main finding of this study is that leisure-time Internet and computer use is strongly related to being overweight or obese, whereas it is largely independent of leisure-time physical activity. After adjusting for socio-demographic variables, leisure-time physical activity and other sedentary behaviors, participants who used the Internet and computer for three hours or more in the last seven days were 1.5 times more likely to be overweight and 2.5 times more likely to be obese compared to non-users. Although there are no direct comparisons with other studies for these outcomes in adults, they are in line with studies that report that higher amounts of time in sedentary behavior and television viewing are strongly associated with overweight and obesity [5-10,12,27].

The strong associations of leisure-time Internet and computer use with overweight and obesity may in part be explained by the association of leisure-time Internet and computer use with other leisure-time sedentary behaviors. Participants who had high Internet and computer use in their leisure time were 2.5 times more likely to engage in more than five hours of other sedentary behaviors per day. This is consistent with a study by Sugiyama et al [22] which showed that time spent watching TV was positively associated with other leisure-time sedentary behaviors; however, this was only the case for women.

Our results showed that leisure-time Internet and computer use was not strongly associated with leisure-time physical activity. Contrary to what might be expected, participants with low leisure-time Internet and computer use were slightly more likely to be in a higher leisure time physical activity category. While it is difficult to explain this outcome, it might be argued that it could be due to the higher socio-economic profile observed in participants with low Internet use. It is generally the case that those of higher socio-economic status are more physically active [28]. However, the analyses controlled for educational attainment and employment status, so such an interpretation would not apply to our findings. Other than this particular relationship, no associations between leisure-time Internet and computer use and physical activity were observed. This finding is, thus, for the major part consistent with studies that showed non-significant associations between Internet use and physical activity in children and adolescents [19-21] and those between TV viewing time and physical activity in adults [5,6,8-10,12]. From this perspective, the apparent paradox of increasing physical activity using an intervention delivery mode that promotes sedentary behavior (Internet and computer use) appears to be invalid. Our results suggest that the time that spent taking part in Internet interventions is not likely to displace leisure-time physical activity; hence, Internet interventions should be considered as an acceptable method to increase physical activity.

As may be seen in Figure 1, a combination of high Internet and computer use and low leisure-time physical activity was associated with a higher odds ratio of being overweight or obese. This finding is consistent with those of a study by Salmon et al [12] in which higher levels of TV viewing in combination with lower levels of physical activity participation were found to be associated with being overweight or obese. This figure also shows that adults who use the Internet and computer for more than three hours in their leisure time are significantly more likely to be overweight, even if they are highly active in their leisure time. Consistent with what was reported by Salmon et al [12], these findings suggest that, in order to reduce the prevalence of overweight and obesity, it may be important not only to increase participation in physical activity, but also to reduce time spent in sedentary behaviors, such as leisure-time Internet and computer use.

As the level of Internet penetration increases, its users become more representative of the general population; thus, gender, age, and socio-economic differences are diminishing [13]. Nevertheless, interesting socio-demographic profiles emerged when leisure-time Internet and computer use were categorized into different levels of usage. Our findings indicate that participants with low leisure-time Internet and computer use had the highest socio-economic profile, engaged in less time in other sedentary behaviors and were slightly more likely to do more leisure-time physical activity. On the other hand, participants with high leisure-time Internet and computer use had lower socio-economic profiles, engaged in more time in other sedentary behaviors, had a higher BMI, and were more likely to be male. This suggests that different levels of leisure-time Internet and computer use are related to different socio-demographic profiles and health behaviors.

Given the high prevalence of Internet use, and its potential impact on health, it is important to address health issues for Internet users. Internet interventions to reduce weight or increase physical activity are likely to be more effective if they take differences among Internet users into account. Although a substantial number of these Internet interventions have been implemented, no studies reported that participants were targeted differently based on their level of Internet and computer use [29,30]. Our findings suggest that doing so might be important. More efforts should be put in targeting participants with high Internet and computer use as compared to those with a low Internet and computer use (less than 3 hours a week). As indicated above, those with high Internet and computer use have an unfavorable health risk profile, and thus potentially may be more open to participating in such interventions, as they have higher levels of computer use in their leisure time.

The major limitations of this study are that it relies on self-reported measures and a cross-sectional design which does not allow determination of the causal direction of the results. More research, using objective measures and prospective study designs, is needed to evaluate these associations. A further limitation is that this study only investigated leisure-time behaviors, this prevents evaluating the impact of using the Internet and computer at work on physical activity and overweight and obesity. Nevertheless, the associations observed in this study indicate that the impact of Internet and computer use, when only used in leisure time, is strong enough to have an influence on health, and that this impact should be taken into consideration when developing new interventions targeting these leisure-time behaviors.

In summary, high levels of leisure-time Internet and computer use were associated with a higher BMI (even among those engaging in a high level of leisure-time physical activity) and higher levels of other leisure-time sedentary behaviors. However, Internet and computer use was mostly unrelated to leisure-time physical activity. These findings suggest that, in addition to nutritional and physical activity interventions, it may also be necessary to decrease time spent in sedentary behaviors (including leisure-time Internet and computer use) in order to reduce the risk of overweight and obesity. Furthermore, future Internet interventions to reduce weight or increase physical activity may need to differentiate between participants with different levels of leisure-time Internet and computer use, in order to increase their effectiveness. Our study is the first to evaluate these specific associations; hence, more research is needed to confirm these findings. More specifically, longitudinal studies are required to examine further the potential causal relationships between specific sedentary behaviors, such as Internet and computer use, and weight gain.

Acknowledgments

The study was funded by the National Health and Medical Research Council of Australia (NHMRC) Project Grant (#213114) and NHMRC Program Grant (#301200). Vandelanotte is supported by a NHMRC (#519778) and the National Heart Foundation of Australia (NHF) (#PH 07B 3303) post-doctoral research fellowship. Sugiyama is supported by an NHMRC capacity building grant in population health (#252799). Gardiner is supported by a NHF postgraduate research scholarship (#PP 06B 2889).

Conflicts of Interest

None declared.

References

1. Ball K, Owen N, Salmon J, Bauman A, Gore CJ. Associations of physical activity with body weight and fat in men and women. Int J Obes Relat Metab Disord 2001 Jun;25(6):914-919 [FREE Full text] [Medline: 11439308] [doi: 10.1038/sj.ijo.0801622]

2. Hemmingsson E, Ekelund U. Is the association between physical activity and body mass index obesity dependent? Int J Obes (Lond) 2007 Apr;31(4):663-668. [Medline: 16953254] [doi: 10.1038/sj.ijo.0803458]
3. ; US Department of Health and Human Services. Dietary guidelines for Americans, 2005. US Department of Agriculture. URL: http://www.health.gov/dietaryguidelines [accessed 2008 Feb 28] [WebCite Cache ID 5VxfyjhFA]
4. Haskell WL, Lee IM, Pate RR, et al. Physical activity and public health: updated recommendations for adults from the American College of Sports Medicine and the American Heart Association. Med Sci Sport Exer 2007;39(8):1423-1434. [Medline: 17762377] [doi: 10.1249/mss.0b013e3180616b27]
5. Dunstan DW, Salmon J, Owen N, Armstrong T, Zimmet PZ, Welborn TA, et al. Associations of TV viewing and physical activity with the metabolic syndrome in Australian adults. Diabetologia 2005;48(11):2254-2261. [Medline: 16211373] [doi: 10.1007/s00125-005-1963-4]
6. Jakes RW, Day NE, Khaw KT, et al. Television viewing and low participation in vigorous recreation are independently associated with obesity and markers of cardiovascular disease risk: EPIC-Norfolk population-based study. Eur J Clin Nutr 2003 Sep;57(9):1089-1096. [Medline: 12947427] [doi: 10.1038/sj.ejcn.1601648]
7. Patel AV, Rodriguez C, Pavluck AL, Thun MJ, Calle EE. Recreational physical activity and sedentary behavior in relation to ovarian cancer risk in a large cohort of US women. Am J Epidemiol 2006 Apr 15;163(8):709-716 [FREE Full text] [Medline: 16495470] [doi: 10.1093/aje/kwj098]
8. Crawford DA, Jeffery RW, French SA. Television viewing, physical inactivity and obesity. Int J Obes Relat Metab Disord 1999 Apr;23(4):437-440. [Medline: 10340824] [doi: 10.1038/sj.ijo.0800845]
9. Jeffery RW, French SA. Epidemic obesity in the United States: are fast foods and television viewing contributing? Am J Public Health 1998 Feb;88(2):277-280 [FREE Full text] [Medline: 9491022] [doi: 10.2105/AJPH.88.2.277]
10. Hu FB, Li TY, Colditz GA, Willett WC, Manson JE. Television watching and other sedentary behaviors in relation to risk of obesity and type 2 diabetes mellitus in women. JAMA 2003 Apr 9;289(14):1785-1791 [FREE Full text] [Medline: 12684356] [doi: 10.1001/jama.289.14.1785]
11. Owen N, Leslie E, Salmon J, Fotheringham MJ. Environmental determinants of physical activity and sedentary behavior. Exerc Sport Sci Rev 2000 Oct;28(4):153-158. [Medline: 11064848]
12. Salmon J, Bauman A, Crawford D, Timperio A, Owen N. The association between television viewing and overweight among Australian adults participating in varying levels of leisure-time physical activity. Int J Obes Relat Metab Disord 2000 May;24(5):600-606. [Medline: 10849582] [doi: 10.1038/sj.ijo.0801203]
13. Australian Bureau of Statistics. Household use of information technology 2006-2007. Canberra. 2007. URL: http://www.abs.gov.au/Ausstats/abs@.nsf/lookupMF/ACC2D18CC958BC7BCA2568A9001393AE [accessed 2008-02-29] [WebCite Cache ID 5VxhhWMLU]
14. Assael H. A demographic and psychographic profile of heavy Internet users and users by type of Internet usage. J Advertising Res 2005;15(1):93-123. [doi: 10.1017/S0021849905050014]
15. Mota J, Ribeiro J, Santos MP, Gomes H. Obesity, physical activity, computer use, and TV viewing in Portuguese adolescents. Pediatr Exerc Sci 2006;18:113-121.
16. Kautiainen S, Koivusilta L, Lintonen T, Virtanen SM, Rimpelä A. Use of information and communication technology and prevalence of overweight and obesity among adolescents. Int J Obes (Lond) 2005 Aug;29(8):925-933. [Medline: 15925961] [doi: 10.1038/sj.ijo.0802994]
17. Lajunen HR, Keski-Rahkonen A, Pulkkinen L, Rose RJ, Rissanen A, Kaprio J. Are computer and cell phone use associated with body mass index and overweight? A population study among twin adolescents. BMC Public Health 2007;7(1):24 [FREE Full text] [Medline: 17324280] [doi: 10.1186/1471-2458-7-24]
18. Tammelin T, Ekelund U, Remes J, Näyhä S. Physical activity and sedentary behaviors among Finnish youth. Med Sci Sports Exerc 2007 Jul;39(7):1067-1074. [Medline: 17596773] [doi: 10.1249/mss.0b13e318058a603]
19. Kerner MS, Kalinski M, Kurrant AB, Small E, Spatz E, Gropack S. Leisure-time Internet use does not correlate with physical activity or physical fitness level of ninth grade African-American girls. Pediatr Exerc Sci 2001;13:402-412.
20. Wake M, Hesketh K, Waters E. Television, computer use and body mass index in Australian primary school children. J Paediatr Child Health 2003 Mar;39(2):130-134. [Medline: 12603802] [doi: 10.1046/j.1440-1754.2003.00104.x]
21. Feldman DE, Barnett T, Shrier I, Rossignol M, Abenhaim L. Is physical activity differentially associated with different types of sedentary pursuits? Arch Pediatr Adolesc Med 2003 Aug;157(8):797-802 [FREE Full text] [Medline: 12912786] [doi: 10.1001/archpedi.157.8.797]
22. Sugiyama T, Healy GN, Dunstan DW, Salmon J, Owen N. Is television viewing time a marker of a broader pattern of sedentary behavior? Ann Behav Med 2008 Apr;35(2):245-250. [Medline: 18357498] [doi: 10.1007/s12160-008-9017-z]
23. Owen N, Cerin E, Leslie E, et al. Neighborhood walkability and the walking behavior of Australian Adults. Am J Prev Med 2007;33(5):387-395. [Medline: 17950404] [doi: 10.1016/j.amepre.2007.07.025]
24. du Toit L, Cerin E, Leslie E. An account of spatially based survey methods and recruitment outcomes of the Physical Activity in Localities and Community Environments (PLACE) Study. Cancer Prevention Research Centre, School of Population Health, The University of Queensland: Brisbane. 2005. URL: http://www.uq.edu.au/cprc/docs/Place_report_2005_Final.pdf [accessed 2008 Feb 29] [WebCite Cache ID 5VxjKXBmI]

25. Salmon J, Owen N, Crawford D, Bauman A, Sallis JF. Physical activity and sedentary behavior: a population-based study of barriers, enjoyment, and preference. Health Psychol 2003 Mar;22(2):178-188. [Medline: 12683738] [doi: 10.1037/0278-6133.22.2.178]
26. Craig CL, Marshall AL, Sjostrom M, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sport Exer 2003;35(8):1381-1395. [Medline: 12900694] [doi: 10.1249/01.MSS.0000078924.61453.FB]
27. Brown WJ, Miller YD, Miller R. Sitting time and work patterns as indicators of overweight and obesity in Australian adults. Int J Obes Relat Metab Disord 2003 Nov;27(11):1340-1346. [Medline: 14574344] [doi: 10.1038/sj.ijo.0802426]
28. 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. [Medline: 12471307]
29. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [Medline: 17572313] [doi: 10.1016/j.amepre.2007.02.041]
30. van den Berg MH, Schoones JW, Vliet Vlieland TPM. Internet-based physical activity interventions: a systematic review of the literature. J Med Internet Res 2007;9(3):e26 [FREE Full text] [Medline: 17942388] [doi: 10.2196/jmir.9.3.e26]

Abbreviations

ANOVA: analysis of variance
BMI: body mass index
ICC: intra-class correlations
IPAC: International Physical Activity Questionnaire
PLACE: Physical Activity in Localities and Community Environments

Edited by G Eysenbach; submitted 10.04.08; peer-reviewed by R Steele, A McConnon; comments to author 24.11.08; revised version received 23.02.09; accepted 04.04.09; published 27.07.09.

Please cite as:
Vandelanotte C, Sugiyama T, Gardiner P, Owen N
Associations of Leisure-Time Internet and Computer Use With Overweight and Obesity, Physical Activity and Sedentary Behaviors: Cross-Sectional Study
J Med Internet Res 2009;11(3):e28
URL: http://www.jmir.org/2009/3/e28/
doi:10.2196/jmir.1084
PMID:19666455

Original Paper

An Evaluation of Patient-Physician Communication Style During Telemedicine Consultations

Zia Agha[1,2], MD MS; Debra L Roter[3], PhD; Ralph M Schapira[4,5], MD

[1]Veterans Affairs San Diego Health Services Research and Development, San Diego, CA, USA

[2]Department of Medicine, University of California, San Diego, CA, USA

[3]Department of Health Policy and Management, Johns Hopkins School of Public Health, Baltimore, MD, USA

[4]Division of Pulmonary and Critical Care Medicine, Medical College of Wisconsin, Milwaukee, WI, USA

[5]Milwaukee Veterans Affairs Medical Center, Milwaukee, WI, USA

Corresponding Author:
Zia Agha, MD MS
HSRD San Diego Veterans Affairs Medical Center
3350 La Jolla Village Drive (111N-1)
San Diego, CA 92161
USA
Phone: +1 858 552 8585 ext 7674
Fax: +1 858 642 1295
Email: zia.agha@va.gov

Abstract

Background: The quality of physician-patient communication is a critical factor influencing treatment outcomes and patient satisfaction with care. To date, there is little research to document the effect of telemedicine (TM) on physician-patient communication.

Objective: The objectives of this study are to measure and describe verbal and nonverbal communication during clinical TM consultations and to compare TM with in-person (IP) consultations in terms of the quality of physician-patient communication.

Methods: Veteran patients (n = 19) requiring pulmonary medicine consultations were enrolled into the study. The study group included 11 patients from the Iron Mountain Veterans Affairs Hospital (VAMC) remote site. Patients had individual TM consultations with a pulmonary physician at the Milwaukee VAMC hub site. A control group of 8 patients had IP consultations with a pulmonary physician at the Milwaukee VAMC. Video recordings of medical consultations were coded for patient-physician verbal and nonverbal communication patterns using the Roter Interaction Analysis System (RIAS).

Results: There were no differences in the length of TM consultations (22.2 minutes) and IP consultations (21.9 minutes). Analysis of visit dialogue indicated that the ratio of physician to patient talk was 1.45 for TM and 1.13 for IP consultations, indicating physician verbal dominance. Physicians were more likely to use orientation statements during IP consultations *(P* = .047). There were greater requests for repetition from patients during TM consultations (*P* = .034), indicating perceptual difficulties.

Conclusions: The study findings indicate differences between TM and IP consultations in terms of physician-patient communication style. Results suggest that, when comparing TM and IP consultations in terms of physician-patient communication, TM visits are more physician centered, with the physician controlling the dialogue and the patient taking a relatively passive role. Further research is needed to determine whether these differences are significant and whether they have relevance in terms of health outcomes and patient satisfaction with care.

(J Med Internet Res 2009;11(3):e36) doi:10.2196/jmir.1193

KEYWORDS

Telemedicine; remote consultation; physician-patient relations

Introduction

Physician-patient communication problems are a common cause for patient dissatisfaction during in-person (IP) consultations. Dissatisfied patients are less likely to return for physicians' appointments, more likely to switch physicians, and more likely to be noncompliant with recommendations [1]. Such nonadherence contributes to unnecessary diagnostic testing,

results in potentially harmful regimen changes, and leads to wasted health resources [2-4].

Clinical telemedicine (TM) includes applications that use communication technologies to link specialists at a tertiary center to patients and primary providers at a remote site. It is unclear how the physical separation between patient and physician inherent in TM consultations affects physician-patient communication. A number of comprehensive reviews have identified the need to evaluate physician-patient interactions during TM consultations [5-7].

One objective of this pilot study is to describe the use of the "Roter Interaction Analysis System" (RIAS) as a comprehensive content coding methodology for assessing verbal and nonverbal communication during clinical TM consultations [1,8,9]. A second objective is to identify potential differences in physician-patient communication style during TM compared to IP consultations.

Methods

Study Setting and Design

We conducted an observational pilot study to compare the pattern of verbal and nonverbal physician-patient communication during TM and IP consultations. The study was conducted at the Milwaukee VAMC and was approved by the Institutional Review Board. The Milwaukee VAMC, a teaching affiliate of the Medical College of Wisconsin, is a tertiary care facility providing a full range of primary care and specialty services for veterans residing in the greater Milwaukee area. The Milwaukee VAMC provides TM services in medical subspecialties (eg, pulmonary medicine, rheumatology, and infectious disease) to the Iron Mountain VAMC in Michigan [10,11]. Our study population was drawn from a group of veterans referred for pulmonary medicine consultations. Subjects in the TM group (n = 11) were referred for TM consultation from the Iron Mountain VAMC. Those in the control group (n = 8) were referred for pulmonary IP consultation from the Milwaukee VAMC. Informed written consent was obtained from all subjects.

Participating Physicians and Nurses

In the present study, 3 pulmonary specialists participated in conducting TM and IP consultations at the Milwaukee VA. These board certified specialists each had at least two years of experience in conducting TM consultations. To control for physician variability, the same group of physicians (n = 3) conducted both the IP and TM consultations. One of two nurses was always present during the TM consultations at the Iron Mountain remote site. These nurses are trained to assist with TM examinations and are familiar with the equipment and technology used during TM.

Telemedicine Consultation Process

TM consultations were performed between the pulmonary specialists located at the Milwaukee VAMC telemedicine site and the patient and trained nurse located at the Iron Mountain VAMC telemedicine site. Consultations were conducted over a live two-way audio and video conferencing system using a high-speed (384 kbps) telecommunication connection. The Milwaukee VAMC used a Tandberg videoconference system with 27-inch Sony monitors; the Iron Mountain VAMC used a VTEL FRED unit with a 27-inch monitor. The established sound and video quality of TM was near perfect, and both parties were able to change the field of view by zooming in or out as needed. The nurse at the Iron Mountain VA assisted the physician during the physical examination. Electronic medical records of each patient were available to the Milwaukee physicians by accessing the VA's computer network, and an electronic progress note for each consultation was entered into the patient's medical record.

In-Person Consultation Process

Patients undergoing IP care at the Milwaukee VAMC are from the Milwaukee metropolitan area. In our study, IP patients were checked in by a clinic nurse and placed in an examination room previously set up with the video recording equipment. Physicians conducted an IP medical interview and a hands-on physical examination with the patients. No nurse was present in the room during these consultations. Physicians had access to electronic medical records of each participating patient and entered an electronic progress note for each consultation.

Data Collection

For TM examinations, a VCR was used to record the picture and sound as viewed from the TM physician's TV monitor in Milwaukee. This 27-inch monitor displayed the image of the remote site examination room in Iron Mountain, including a full-body frontal view of the seated patient and an upper body frontal view of the nurse seated behind a desk next to the patient. In addition, a 5-inch diagonal picture-in-picture frontal view image of the physician was projected in the upper-left corner of the monitor. The IP consultations were recorded by a digital camcorder equipped with a wide-angle lens mounted on a tripod placed in the examination room. The camera and microphone were placed in an unobtrusive manner, and both the patient and physician were alerted to the presence of these devices. The camcorder was set up in a standardized fashion, and the resulting image included a full body, oblique view of both the seated patient and the seated physician.

Measurement of Physician-Patient Communication

We used the Roter Interaction Analysis System (RIAS) to code the medical dialogue. The RIAS treats a complete thought, defined as a simple sentence, a sentence clause, a sentence fragment, or single word, as the unit of analysis. A complete thought may be categorized as one of 38 mutually exclusive and exhaustive codes. Coding is done directly from video recordings without transcription. In the current study, a nurse was always present during the TM consultations, and her communication was coded separately from that of the TM physician. In 8 of the 19 consultations, the patient was accompanied by a companion, and his or her speech was also coded separately. RIAS coding was performed by trained coders under the supervision of Dr. Roter (co-investigator) at Johns Hopkins University. Dr. Roter trained coders in the use of the RIAS System over several weeks, using a coding manual with detailed definitions and annotated examples, and training tapes

demonstrating standardized techniques. Inter-rater reliability was calculated based on double coding of 20% sample by 2 independent and blinded coders. In the current study, reliability ranged across categories roughly from 0.7 to 0.9, based on Pearson correlation coefficients.

Measurement and Analysis of Nonverbal Communication—Global Affect Ratings

In addition to verbal communication, the RIAS is also used for coding nonverbal communication. In this study, we used RIAS for coding of physician and patient global affect ratings. These ratings were based on the emotional tone of each speaker for the following dimensions: (1) angry or irritated, (2) anxious or nervous, (3) depressed or sad (patient only), (4) emotionally distressed (patient only), (5) dominant or assertive, (6) interested or attentive, (7) friendly or warm, (8) responsive or engaged, (9) sympathetic or empathetic, (10) hurried or rushed (physician only), and (11) respectful. The global affect ratings were performed by trained coders under the supervision of co-investigator Dr. Roter at Johns Hopkins University. The coders assigned an overall Likert score (1 = lowest and 6 = highest) for each dimension for the whole interaction. Inter-rater reliability, calculated as percentage agreement by double coding of 20% random sample by 2 independent and blinded coders, ranged from 87% to 100%.

Results

Descriptive Analyses of the Visit Dialogue

Visit length is associated with communication quantity (number of utterances) and was measured for each IP and TM consultation. There were no significant differences in the length of consultations ($P = 0.96$), with IP visits averaging 21.9 ± 7.6 minutes compared to 22.2 ± 12 minutes for TM visits. Descriptive analyses of the visit dialogue can be presented in several ways. First, an overall "profile" of the visits in terms of physicians', nurses', and patients' communication is presented in terms of counts of utterances ("utterance" is defined as a statement or complete thought). The total number of utterances (ie, "all talk") across all participants (physician, patient, nurse, and patient companion) was similar in the two consultation modes (344 ± 170 utterances during IP and 354 ± 233 during TM). During IP visits, there were an equal number of physician utterances (166 ± 80) and patient utterances (166 ± 88). However, during TM visits, physicians accounted for a higher number of utterances (178 ± 118) as compared to patients (142 ± 127). As noted previously, the nurse was present during TM visits only and, on average, contributed 21 utterances (6%) to the dialogue. The largest categories of nurse contribution included orientations (5 utterances), agreements (3 utterances), biomedical information (3 utterances), and closed medical questions (2.5 utterances). A patient companion was equally likely to be present in both IP and TM consultations: 3 of 8 (38%) IP visits and 5 of 11 (45%) TM visits. The verbal contribution of the companion was similar in the two encounter modes, accounting for an average of 7% (31 utterances) of dialogue during IP visits and 9% (28 utterances) during TM visits ($P = .8$). Most of the companion's contribution to the dialogue was in the provision of information to the physician about the patient's medical symptoms (12 utterances), therapeutic regimen (4.5 utterances), lifestyle and psychosocial status (3 utterances), and agreements (3 utterances).

Verbal Dominance Ratio

Descriptive analyses of the visit dialogue can also be presented as various ratios to capture relative amounts of talk. The ratio of total number of provider utterances to total number of patient utterances is called the "verbal dominance ratio" and is a summary measure of "patient-centered" versus "physician-centered" style of communication. A verbal dominance ratio of 1 indicates equal participation by patient and physician and is indicative of a patient-centered interview style, whereas a ratio of greater than 1 indicates a physician-centered interview style. In this study, the physician dominated the interview more during TM as compared to IP visits (TM = 1.45 vs. IP = 1.13; t = -1.25, $P = .23$). When the second provider (ie, the nurse) was included in the provider portion of the ratio calculation, verbal dominance ratio for TM was even higher (1.7). Due to small sample size, the differences (TM vs. IP) in verbal dominance ratios were not statistically significant.

Analyses of RIAS Content Categories for Physician and Patient Dialogue During TM and IP Visits

For the purposes of this study, each utterance made during the visit by the patient, physician, and nurse was coded into one of 38 mutually exclusive and exhaustive RIAS categories. During analysis, the 38 mutually exclusive content categories are often combined into larger "subsuming categories" that share common meaning [12]. In this study, 38 RIAS content categories were combined into 10 larger subsuming categories. The number of utterances (mean and SD) by the physician and the patient for each subsuming communication category are presented in Table 1.

Table 1. Physician-patient verbal communication during TM and IP visits

Speaker	Physician				Patient			
Type of Visit	TM	IP			TM	IP		
RIAS Subsuming Communication Category	N (SD)	N (SD)	t	*P*	N (SD)	N (SD)	t	*P*
Information-gathering, biomedical	29.1 (18.9)	26.5 (23.1)	-0.26	.80	4.9 (4.3)	4.6 (2.4)	-1.84	.86
Information-gathering, psychosocial	4.9 (5.6)	4.8 (4.4)	-0.07	.95	0.2 (0.6)	0.1(0.4)	-0.26	.80
Information-giving, biomedical	71.1 (56.6)	51 (29.8)	-1.00	.33	86.3 (75.8)	97 (58.9)	0.35	.73
Information-giving, psychosocial	1.4 (2.4)	4.8 (9.2)	1.02	.34	11.8 (15.2)	12.8 (17.3)	0.12	.91
Positive talk	31 (23)	27 (23)	-0.36	.73	24 (26)	32 (19)	0.80	.44
Negative talk	0.2 (0.4)	0.4 (0.5)	0.88	.40	1.5 (2)	0.8 (1)	-.98	.34
Social talk	3.7 (3)	2.1 (1.8)	-1.46	.16	2.6 (2.8)	1.3 (1.4)	-1.32	.21
Rapport building/emotional responsiveness	8.1 (9.1)	10.4 (9.8)	0.52	.61	5.4 (5.4)	6.6 (7.2)	0.42	.69
Partnership building	14.2 (11.5)	12.3 (7.5)	-0.44	.66	1.9 (3.1)	2.9 (2.4)	0.76	.46
Orientation statements	9.7 (8.9)	19 (9.4)	2.17	.047	0.9 (1.25)	1.6 (2.7)	0.78	.46

Information-Gathering, Biomedical and Psychosocial

These RIAS categories include closed- and open-ended questions asked by patient and physician related to (1) medical condition and therapeutic regimen (biomedical), and (2) lifestyle and psychosocial topics. Overall, both patients and physicians engaged in more biomedical than psychosocial information gathering during both TM and IP visits. When comparing the two consultation modes, there was no significant difference in the amount of biomedical information gathered by physicians (TM = 29.1 vs. IP = 26.5; t = -0.26, *P* = .80) or patients (TM =4.9 vs. IP = 4.6; t = -1.84, *P* = .86), or psychosocial information gathered by physicians (TM = 4.9 vs. IP = 4.8; t = -0.07, *P* = .95) or patients (TM = 0.2 vs. IP = 0.1; t = -0.26, *P* = .80).

Information-Giving, Biomedical and Psychosocial

These RIAS categories include information sharing by patient and information sharing plus counseling by physician related to (1) medical condition and therapeutic regimen (biomedical), and (2) lifestyle and psychosocial topics. It was noted that both physician and patient predominantly exchanged information regarding biomedical versus lifestyle and psychosocial topics during both IP and TM consultations. Physicians provided more patient counseling and information sharing regarding biomedical issues during TM as compared to IP visits (TM = 71.1 vs. IP = 51; t = -1.00, *P* = .33). Conversely, physicians made a greater number of counseling statements and provided more information on psychosocial and lifestyle issues during IP versus TM visits (IP = 4.8 vs. TM = 1.4; t = 1.02, *P* = .34). There was no difference between TM and IP for patient information giving for either biomedical (TM = 86.3 vs. IP = 97; t = 0.35, *P* = .73) or psychosocial (TM = 11.8 vs. IP = 12.8; t = 0.12, *P* = .91) categories. While none of the results in the information-giving category reached statistical significance, the data suggest that physician communication was more patient-centered (more dialogue around psychosocial and lifestyle issues) during IP visits and more physician-centered (more biomedical talk and less psychosocial/lifestyle talk) during TM visits.

Positive Talk, Negative Talk, and Social Talk

The category “positive talk” includes laughter, agreement, and approval. Positive talk constituted a considerable portion of total talk for both physicians and patients (TM = 17.9%, SD = 5.3; IP = 18%, SD = 4.8), there was no difference in physician talk (TM = 31 vs. IP = 27; t = -0.36, *P* = .73) or patient talk (TM = 24 vs. IP = 32; t = 0.80, *P* = .44) in this category when comparing the two visit types. “Negative talk” includes statements of disagreement and criticism. Such talk was rarely engaged in, and there was no significant difference in this category for physicians (TM = 0.2 vs. IP = 0.4; t = 0.88, *P* = .40) or patients (TM = 1.5 vs. IP = 0.8; t = -0.98, P = .34) when comparing the two visit types. “Social talk” includes all nonmedical dialogue. Physicians and patients use social talk during medical visits to develop rapport and display interest. Social talk was infrequent overall. Although more social talk occurred during TM visits, there was no significant difference in this category for physicians (TM = 3.7 vs. IP = 2.1; t = -1.46, *P* = .16) or patients (TM = 2.6 vs. IP = 1.3; t = -1.32, *P* = .21) when comparing the two visit types.

Rapport Building/Emotional Responsiveness and Partnership Building

The RIAS category “rapport building/emotional responsiveness” includes instances in which the patient or physician shows concern or asks for or provides an opinion or reassurance. There was no difference between TM and IP visits for physicians (TM = 8.1 vs. IP = 10.4; t = 0.52, *P* = .61) or patients (TM = 5.4 vs. IP = 6.6; t = 0.42, *P* = .69) in this category. “Partnership building” includes verbal communication that indicates understanding, as well as instances in which the patient or physician paraphrases or interprets the other’s talk. There was no difference in partnership building between TM and IP visits for physicians (TM = 14.2 vs. IP = 12.3; t = -0.44, *P* = .66) or patients (TM = 1.9 vs. IP = 2.9; t = 0.76, *P* = .46).

Orientation Statements

This category includes physician statements that tell the patient what is expected during the consultation or what is about to happen (eg, "I am going to check your pulse now."), as well as statements that serve to orient the patient to major topics of discussion or the physical flow of the visit. Physicians used fewer orientation statements during TM (9.7) as compared to IP (19) visits (t = 2.17, *P* = .047). Overall, patients made few orientation statements (eg, requests for instructions related to flow of the visit and physical exam) and there was no difference when comparing the two visit types (TM = 0.9 vs. IP = 1.6; t = 0.78, *P* = .46).

Requests for Repetition and Unintelligible Utterances

Requests for repetition are statements (eg, "What?" "Come again?" "How is that?" "Would you repeat that?") in response to instances in which one participant has not clearly heard or understood another's words or statements. Such requests are indicative of perceptual difficulties. Utterances are coded as "unintelligible" when the coder is unable to understand what a participant (eg, patient or physician) has said. These two categories are not part of the subsuming Roter categories (Table 1) used in this study. However, we coded and analyzed TM vs IP data for these categories to detect any differences in clarity of physician-patient verbal communication, as the effect of TM technology on the quality of communication is an area of concern. Patients made significantly more requests for repetition during TM visits (TM = 1.64 vs. IP = .38; t = -2.33, *P* = .034). Physicians also made more requests for repetition during TM, although the difference was not statistically significant (TM = .45 vs. IP = .00; t = -2.19, *P* = .053). Unintelligible utterances in physician dialogue were more common during IP visits (TM = .27 vs. IP = 3.75; t = 2.66, *P* = .031). There was no significant difference in the number of unintelligible utterances by patients, when comparing the two consultation modes (TM = .91 vs. IP = 2.38; t = 1.16, *P* = .28).

Nonverbal Communication

The RIAS was used for global ratings of the emotional tone (nonverbal communication) of both the patient and physician. Based on vocal qualities, coders captured the global ratings (ie, a single rating for the entire visit) of emotional affect for both the patient and physician on 11 affective dimensions (anger, anxiety, sadness, distress, dominance, interest, friendliness, responsiveness, sympathy, hurried, and respectful) for each IP and TM visit. There were no significant differences noted in global affect ratings for physicians (TM = 2.4 vs. IP = 2.4; t = -0.15, *P* = .88) or patients (TM = 2.3 vs. IP = 2.2; t = -1.17, *P* = .26) during either type of visit.

Discussion

A popular conceptual model used to describe physician-patient communication defines communication styles as either "physician-centered" or "patient-centered." Physician behaviors, such as gathering of information via closed-ended questions, testing hypotheses to make a diagnosis, giving medical directions, and controlling the visit, represent a "physician-centered" style of communication. The patient's role is to listen, follow the physician's directions, and play a passive role during the medical encounter. This type of communication is less successful in addressing the needs of the patient [13-15]. Conversely, a "patient-centered" style of communication is characterized by physician behaviors such as asking open-ended questions, partnership building, shared decision making, information sharing, counseling, and using statements of concern, agreement, and approval. The patient's role is to participate actively in making decisions, to express opinions or concerns about his or her health, and to ask for information. This style is more successful in addressing patient needs and is associated with higher patient satisfaction, better psychosocial adjustment, and improved health outcomes [13-19].

The findings of this study suggest that the use of TM does influence patient-physician communication style. During TM visits, physicians were more likely to dominate the dialogue, as evidenced by a higher verbal dominance ratio (TM = 1.45 vs. IP = 1.13; t = -1.25, *P* = .23). In addition, both physicians and patients were more likely to address biomedical topics during TM visits (topics associated with a physician-centered style of communication), while discussion around psychosocial and lifestyle issues (topics associated with a patient-centered style of communication) was limited. These findings corroborate those reported by Street et al's [6] study in which content analysis of 26 TM consultations between specialists, primary care physicians, and patients showed that the specialists were the dominant communicators in terms of asking questions and displaying controlling behavior.

Clinical TM consultations fundamentally differ from IP encounters due to the physical separation of physician and patient. Direct physical examination and interview are not possible, and a virtual environment replaces the familiar physician's office. It is possible that the physical separation and the lack of the hands-on physical examination reinforce a "physician-centered" interview style observed during TM consultations. Whether such differences in communication, particularly in the absence of the hands-on examination, have an effect on quality of care, health outcomes, and future patient utilization of TM is not known.

We observed patients to be less engaged (less talkative) and more likely to take on a passive role during TM as compared to IP visits. It is possible that poor patient participation and communication during TM visits are due to the lack of familiarity with technology and the perception of physician detachment due to the inherent physical separation. Patient concerns about privacy and confidentiality during TM consultations may further inhibit patient participation, especially if the dialogue involves collecting data on sensitive or personal topics (eg, sexual history).

Clinical TM consultations frequently involve a three-party communication exchange between a specialist at a tertiary site, and a patient and second provider at a remote site. In Street's study, the presence of a referring primary care physician was linked to inhibited patient communication and lack of direct engagement with the consulting physician [6]. While it is possible that the presence of a second provider may result in poor patient participation [6], it is also possible that this presence

helps promote confidence during TM, as patients feel they are getting attention from two health care professionals (two physicians or a physician and a nurse) versus only one physician during IP visits [20]. In our study, a trained clinical nurse assisted the patient during each TM consultation. On average, the nurse contributed only 6% of all utterances during the TM visit and was the least verbally active participant. It appears that the presence of a trained nurse provider instead of a referring physician may be less detrimental to communication between the patient and the TM physician.

The present study results suggest that TM visits are less patient centered than IP visits. In contrast, results of studies that use patient self report to measure patient satisfaction often indicate high patient satisfaction with TM [21]. We believe that a number of patient factors may explain these differences between third-party evaluation of communication during TM, as in the present study, and patient self report of satisfaction. Patients, in general, are likely to view their medical care in a favorable light. In addition, it is possible that patient expectations for quality of communication are different when it comes to TM versus IP care (ie, patients may have a lower expectation from TM consultation and therefore be less critical of shortcomings in communication). Convenience of TM may also play a role. A number of studies have reported high patient satisfaction with TM because it is convenient (reduced travel) and improves patient access to specialist physician care [21]. Patients may also have a positive perception of TM due to the use of latest technology, hence promoting confidence that they are receiving highest quality care. Infatuation with the use of technology during medical care has been reported in the TM literature. Baigent et al [22] and Gammon et al [23] found that patients reported enjoying video consultations and were inspired by the use of technology. In addition, the presence of a second provider during TM may be viewed positively by patients, who perceive it as more attention—two health care professionals (two physicians or a physician and a nurse) versus only one physician during IP visits [20].

Study Limitations

The study is small and exploratory with obvious limitations in terms of experimental design and statistical power. Nevertheless, the study provides a framework for detailed observations and description of patient-physician communication during TM consultations that can be useful in the design of future research in this area. Undoubtedly, the quality of interaction during TM and its potential impact on health care outcomes is an area of growing importance. Further research is needed to help fill the current gaps in the literature and to develop specific interventions that can improve the quality of communication during TM encounters.

Acknowledgments

This material is based upon work supported in part by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development—Health Services Research and Development and by resources and the use of facilities at the VA Medical Center Milwaukee, WI.

Conflicts of Interest

None declared.

References

1. Roter D, Hall J. Doctors Talking With Patients/Patients Talking With Doctors: Improving Communication in Medical Visits. Westport, CT: Auburn House; 1992.
2. Hays RD, Kravitz RL, Mazel RM, Sherbourne CD, DiMatteo MR, Rogers WH, et al. The impact of patient adherence on health outcomes for patients with chronic disease in the Medical Outcomes Study. J Behav Med 1994 Aug;17(4):347-360. [Medline: 7966257] [doi: 10.1007/BF01858007]
3. DiMatteo MR. The physician-patient relationship: effects on the quality of health care. Clin Obstet Gynecol 1994 Mar;37(1):149-161. [Medline: 8194205] [doi: 10.1097/00003081-199403000-00019]
4. DiMatteo MR, Hays RD, Prince LM. Relationship of physicians' nonverbal communication skill to patient satisfaction, appointment noncompliance, and physician workload. Health Psychol 1986;5(6):581-594. [Medline: 3803351] [doi: 10.1037/0278-6133.5.6.581]
5. Whitten PS, Mair F. Telemedicine and patient satisfaction: current status and future directions. Telemed J E Health 2000;6(4):417-423. [Medline: 11242550] [doi: 10.1089/15305620050503898]
6. Street R, Wheeler J, McCaughan W. Specialist-primary care provider-patient communication in telemedical consultations. Telemed J 2000;6(1):45-54. [doi: 10.1089/107830200311842]
7. Miller EA. Telemedicine and doctor-patient communication: an analytical survey of the literature. J Telemed Telecare 2001;7(1):1-17. [Medline: 11265933] [doi: 10.1258/1357633011936075]
8. Roter D. The Roter Interaction Analysis System (RIAS) Coding Manual. Baltimore, MD: School of Hygiene and Public Health, Johns Hopkins University; 1991.
9. Inui TS, Carter WB, Kukull WA, Haigh VH. Outcome-based doctor-patient interaction anaylsis: I. Comparison of techniques. Med Care 1982 Jun;20(6):535-549. [Medline: 7109738] [doi: 10.1097/00005650-198206000-00001]

10. Agha Z, Schapira RM, Maker AH. Cost effectiveness of telemedicine for the delivery of outpatient pulmonary care to a rural population. Telemed J E Health 2002;8(3):281-291. [Medline: 12419022] [doi: 10.1089/15305620260353171]
11. Raza T, Joshi M, Schapira RM, Agha Z. Pulmonary telemedicine--a model to access the subspecialist services in underserved rural areas. Int J Med Inform 2009 Jan;78(1):53-59. [Medline: 18809352] [doi: 10.1016/j.ijmedinf.2008.07.010]
12. Roter DL, Stewart M, Putnam SM, Lipkin M, Stiles W, Inui TS. Communication patterns of primary care physicians. JAMA 1997;277(4):350-356. [Medline: 9002500] [doi: 10.1001/jama.277.4.350]
13. Kinnersley P, Stott N, Peters TJ, Harvey I. The patient-centredness of consultations and outcome in primary care. Br J Gen Pract 1999 Sep;49(446):711-716 [FREE Full text] [Medline: 10756612]
14. Bertakis KD, Roter D, Putnam SM. The relationship of physician medical interview style to patient satisfaction. J Fam Pract 1991 Feb;32(2):175-181. [Medline: 1990046]
15. Krupat E, Rosenkranz SL, Yeager CM, Barnard K, Putnam SM, Inui TS. The practice orientations of physicians and patients: the effect of doctor-patient congruence on satisfaction. Patient Educ Couns 2000 Jan;39(1):49-59. [Medline: 11013547] [doi: 10.1016/S0738-3991(99)00090-7]
16. Stewart M, Brown JB, Boon H, Galajda J, Meredith L, Sangster M. Evidence on patient-doctor communication. Cancer Prev Control 1999 Feb;3(1):25-30. [Medline: 10474749]
17. Roter D. Patient-centered communication. BMJ 2004 Jun 12;328(7453):E303-E304 [FREE Full text] [Medline: 15191997] [doi: 10.1136/bmj.328.7453.E303]
18. Roter DL, Hall JA, Kern DE, Barker LR, Cole KA, Roca RP. Improving physicians' interviewing skills and reducing patients' emotional distress. A randomized clinical trial. Arch Intern Med 1995 Sep 25;155(17):1877-1884. [Medline: 7677554] [doi: 10.1001/archinte.155.17.1877]
19. Roter DL, Hall JA, Merisca R, Nordstrom B, Cretin D, Svarstad B. Effectiveness of interventions to improve patient compliance: a meta-analysis. Med Care 1998 Aug;36(8):1138-1161. [Medline: 9708588] [doi: 10.1097/00005650-199808000-00004]
20. Miller EA. The technical and interpersonal aspects of telemedicine: effects on doctor-patient communication. J Telemed Telecare 2003;9(1):1-7. [Medline: 12641885] [doi: 10.1258/135763303321159611]
21. Mair F, Whitten P. Systematic review of studies of patient satisfaction with telemedicine. BMJ 2000 Jun 3;320(7248):1517-1520 [FREE Full text] [Medline: 10834899] [doi: 10.1136/bmj.320.7248.1517]
22. Baigent MF, Lloyd CJ, Kavanagh SJ, Ben-Tovim DI, Yellowlees PM, Kalucy RS, et al. Telepsychiatry: 'tele' yes, but what about the 'psychiatry'? J Telemed Telecare 1997;3(Suppl 1):3-5. [Medline: 9218364] [doi: 10.1258/1357633971930346]
23. Gammon D, Sørlie T, Bergvik S, Høifødt TS. Psychotherapy supervision conducted by videoconferencing: a qualitative study of users' experiences. J Telemed Telecare 1998;4(Suppl 1):33-35. [Medline: 9640727] [doi: 10.1258/1357633981931353]

Abbreviations

TM: telemedicine
IP: in-person
VAMC: Veterans Affairs Medical Center
VA: veterans affairs
RIAS: Roter Interaction Analyses System

Edited by G Eysenbach; submitted 09.01.09; peer-reviewed by B Wakefield; comments to author 08.05.09; revised version received 04.06.09; accepted 21.07.09; published 30.09.09.

Please cite as:
Agha Z, Roter DL, Schapira RM
An Evaluation of Patient-Physician Communication Style During Telemedicine Consultations
J Med Internet Res 2009;11(3):e36
URL: http://www.jmir.org/2009/3/e36/
doi:10.2196/jmir.1193
PMID:19793720

Original Paper

Text Mining and Natural Language Processing Approaches for Automatic Categorization of Lay Requests to Web-Based Expert Forums

Wolfgang Himmel[1], PhD; Ulrich Reincke[2]; Hans Wilhelm Michelmann[3], PhD

[1]Department of General Practice / Family Medicine, Georg-August-University, Göttingen, Germany

[2]Competence Centre Enterprise Intelligence, SAS Institute GmbH, Heidelberg, Germany

[3]Department of Obstetrics and Gynaecology (Study Group of Reproductive Medicine), Georg-August-University, Göttingen, Germany

Corresponding Author:
Wolfgang Himmel, PhD
Department of General Practice / Family Medicine
University of Göttingen
Humboldtallee 38
37070 Göttingen
Germany
Phone: +49 0 551 39 22648
Fax: +49 0 551 39 9530
Email: whimmel@gwdg.de

Abstract

Background: Both healthy and sick people increasingly use electronic media to obtain medical information and advice. For example, Internet users may send requests to Web-based expert forums, or so-called “ask the doctor” services.

Objective: To automatically classify lay requests to an Internet medical expert forum using a combination of different text-mining strategies.

Methods: We first manually classified a sample of 988 requests directed to a involuntary childlessness forum on the German website “Rund ums Baby” (“Everything about Babies”) into one or more of 38 categories belonging to two dimensions (“subject matter” and “expectations”). After creating start and synonym lists, we calculated the average Cramer’s V statistic for the association of each word with each category. We also used principle component analysis and singular value decomposition as further text-mining strategies. With these measures we trained regression models and determined, on the basis of best regression models, for any request the probability of belonging to each of the 38 different categories, with a cutoff of 50%. Recall and precision of a test sample were calculated as a measure of quality for the automatic classification.

Results: According to the manual classification of 988 documents, 102 (10%) documents fell into the category “in vitro fertilization (IVF),” 81 (8%) into the category “ovulation,” 79 (8%) into “cycle,” and 57 (6%) into “semen analysis.” These were the four most frequent categories in the subject matter dimension (consisting of 32 categories). The expectation dimension comprised six categories; we classified 533 documents (54%) as “general information” and 351 (36%) as a wish for “treatment recommendations.” The generation of indicator variables based on the chi-square analysis and Cramer’s V proved to be the best approach for automatic classification in about half of the categories. In combination with the two other approaches, 100% precision and 100% recall were realized in 18 (47%) out of the 38 categories in the test sample. For 35 (92%) categories, precision and recall were better than 80%. For some categories, the input variables (ie, “words”) also included variables from other categories, most often with a negative sign. For example, absence of words predictive for “menstruation” was a strong indicator for the category “pregnancy test.”

Conclusions: Our approach suggests a way of automatically classifying and analyzing unstructured information in Internet expert forums. The technique can perform a preliminary categorization of new requests and help Internet medical experts to better handle the mass of information and to give professional feedback.

(J Med Internet Res 2009;11(3):e25) doi:10.2196/jmir.1123

KEYWORDS

Text mining; qualitative research; natural language processing; consumer health informatics; Internet; remote consultation; infertility

Introduction

Both healthy and sick people increasingly use electronic media to obtain medical information and advice [1]. Internet users actively exchange information with others about subjects of interest or send requests to Web-based expert forums, or so-called "ask the doctor" services [2,3]. They want to understand specific diseases, to be informed about new therapies, or to ask for a second opinion before they decide on a treatment [4-6]. In addition, these expert forums also represent seismographs for medical and/or psychological needs, which are apparently not met by existing health care systems [5, 7].

In the past, emails, e-consultations, and requests for medical advice via the Internet have been manually analyzed using quantitative or qualitative methods [1-6]. To facilitate the work of medical experts and to make full use of the seismographic function of expert forums, it would be helpful to classify visitors' requests automatically. By doing so, specific requests could be directed to the appropriate expert or even answered semiautomatically, thereby providing comprehensive monitoring. By generating "frequently asked questions (FAQs)," similar patient requests and their corresponding answers could be collated, even before the expert replies. Machine-based analyses could help both the lay public to better handle the mass of information and medical experts to give professional feedback. In addition, this method could be used to help policy makers recognize the health needs of the population [8].

Text mining [9] is a method for the automatic classification of large volumes of documents, which could be applied to the problem at hand. This technique usually consists of finite steps, such as parsing a text into separate words, finding terms and reducing them to their basics ("truncation") followed by analytical procedures such as clustering and classification to derive patterns within the structured data, and finally evaluation and interpretation of the output. Typical text-mining tasks include, besides others, text categorization, concept/entity extraction, sentiment analysis, and document summarization. This technique has been successfully applied, for example, in automatic indexing, ascertaining and classifying consumer complaints, and handling changes of address requests sent to companies by email. Text mining is also used in genome analysis, media analysis, and indexing of documents in large databases for retrieval purposes [8-11].

An automatic classification of lay requests to medical expert Internet forums is a challenge because these requests can be very long and unstructured as a result of mixing, for example, personal experiences with laboratory data. Very often, people simply require psychological help or are looking for emotional reassurance. Such heterogeneous samples of requests appear in the section "Wish for a Child" on the German Rund ums Baby (Everything about Babies) website [12], which provides information for parents, potential parents, and infertile couples.

Although involuntary childlessness is not the focus of this paper, some introductory notes on this condition may be helpful. Infertility leading to involuntary childlessness is defined as the inability of a couple to achieve conception or bring a pregnancy to term after a year or more of regular, unprotected sexual intercourse. Infertile couples may pass through different stages of reactions and feelings, which include shock, surprise, anger, helplessness, and loss of control. Feelings of failure, embarrassment, shame, and stigmatization may lead to social isolation and to a breakdown in communication between the couple, including depressive reactions, anxiety, emotional instability, diminished self-confidence, sexual problems, and conflicts [13].

The vast majority of cases of male infertility are due to a low sperm count, often associated with poor motility and a high rate of abnormal sperm. However, in a large number of patients (25% to 30%), it is not possible to determine the cause of the problem. The main causes of female infertility are ovarian dysfunctions and disorders of the fallopian tubes and uterus. Frequently, two or even all three causes can be found in one patient. Before 1980, infertility due to low sperm quality was treated by performing insemination with the patient's own sperm or donor sperm. This was followed by in vitro fertilization (IVF) in the early 1980s and intracytoplasmic sperm injection (ICSI) in the early 1990s. ICSI only requires one living sperm cell [14].

Like many other conditions, involuntary childlessness is often not caused by just one factor, nor can it always be cured with a single treatment regimen. Patients and doctors alike are often confronted with the fact that they cannot find a reason for childlessness and that a treatment for a particular case is not helpful for a person or couple with a similar problem [15]. In addition to the cause itself, other factors, such as the age of the woman or problems shared by both partners, might also influence the choice of treatment. So it seems consequent that patients/couples suffering from involuntary childlessness use the Internet to get information about their infertility [6].

Requests addressed to medical expert forums such as "Wish for a Child" can be classified according to (1) the subject matter or (2) the sender's expectation (eg, to receive a summary of the current treatment options [second opinion], to get general information about a certain disease or biological process, or to ask for advice about where to seek adequate medical help). While the first aspect is of great importance to medical experts so that they can understand the contents of requests, the latter is of interest to public health experts to allow the analysis of information needs within the population.

We carried out an initial trial to automatically classify these requests using standard text-mining software such as that provided by SAS [16,17]. However, the results of our first trial were rather disappointing since the quality of classification, expressed in terms of precision and recall, did not exceed 60% [18].

To make full use of text mining with complex data, different strategies and a combination of these strategies may refine automatic classification. The aim of this paper is to present a method for an automatic classification of requests to a medical expert forum and to evaluate its performance quality. A special focus of this method should be its flexibility to allow a precise and content-related input of expert knowledge.

Methods

Setting and Data

The analysis is based on a sample of requests collected from the section "Wish for a Child" on the German Rund ums Baby website [12]. In this section, visitors can participate in a medical expert forum and ask questions about involuntary childlessness. Requests and answers are openly published on the website. The structure of these dialogues resembles, for example, The Heart Forum of the Cleveland Clinic Foundation [19].

Visitors to the website ask questions directly to a group of medical experts via a Web-based interface. The expert team consists, at the moment, of eight persons who are experts in gynecology, urology, andrology, and/or embryology. Some of them work in outpatient departments, some in reproductive clinics, and some in university hospitals. So, the expert forum is well equipped to give medical advice in difficult situations, to provide help to make the correct decision, to offer a second opinion, or, in some instances, even to meet psychological needs not covered by doctors. The experts work on an honorary (unpaid) basis.

To date, more than 12,000 requests have been sent to the expert forum and have been published on the site. From these requests, we selected a random sample of 988 and classified them manually to provide a sound basis for training and evaluation.

Manual Classification

Similar to Shuyler and Knight [20], who analyzed questions to an orthopedic website in several dimensions (topics, purpose, relationship), we decided to classify the requests into two dimensions. The first dimension ("subject matter") comprised 32 categories (eg, assessment of pregnancy symptoms or information about artificial insemination). The second dimension ("expectations") comprised six different categories that characterize the goals or the purpose of the sender (eg, emotional reassurance or a recommendation about treatment options).

From the very beginning of the classification process, it became apparent that many requests belong to one subject matter category but fit into more than one category of the second dimension ("expectations"). For example, a visitor asked the experts to comment on the results of a semen analysis and, at the same time, wanted some advice about whether he or she should change doctors. We decided to provide as many categories per request as appropriate. In the first dimension ("subject matter"), this request could be categorized as "semen analysis," and, in the second dimension ("expectations"), as "discussion of results" as well as "treatment options."

Two of the authors (HWM, WH) independently classified the first set of 100 requests manually. Because of a high rate of differing results, we defined the categories more precisely, added and removed some categories, and agreed upon the use of multiple categories. We then classified another 100 requests. This time, strong classification discrepancies, such as each author classifying the text into a different category, occurred in only 12 cases. Some minor discrepancies also occurred, such as agreement in all categories except one additional category that was suggested by one author but not the other. This resulted in a degree of agreement of 0.69 according to the kappa statistic for overlapping categories [21]. Complete agreement was achieved after further discussion and refinement of the categories and then HWM once more manually coded the first 200 and then the remaining 788 requests. The final categories of both dimensions used for classification are shown in Table 2, presented in the Results section.

Table 1. Terms and parents

Terms[a]	Parents
month	month
months	month
monthly	month
all months (eg, January, February)	month
all abbreviations (eg, Jan., Feb.)	month
uterus	uterus
uterus milleu	uterus
utterus	uterus
womb	uterus
in utero	uterus
uterine	uterus
adrenal gland	adrenal gland
temperature	temperature
temperture	temperature
temp.	temperature
body temperature	temperature
thermometry	temperature
all temperature degrees (eg, 37.3°C)	temperature
ultrasound	ultrasound
ultra	ultrasound
ultrasonic	ultrasound
u-sound	ultrasound
sound	ultrasound
scan	ultrasound

[a]Examples for single words, multi-word terms, synonyms, abbreviations, misspellings, etc are translated from the original German data.

Preparation for Automatic Classification

For automatic classification, we created a dataset that contained the text from each request as a separate observation. The text was then parsed into separate words or noun groups. “Parsing” entails several techniques: (1) separation of the text into terms (eg, “uterus”) or multi-word terms (eg, “uterus milleu”), (2) normalization of different formats for dates (eg, 26/02/2008; Feb. 26, 2008) and data (eg, various degrees of temperature), (3) recognition of synonyms, and (4) stemming of verbs, nouns, or (in German) adjectives to their root form (eg, “transfer,” “transferred,” “transferring”). Programs, such as SAS Text Miner, perform this automatically and provide a complete list of all words, noun groups, and so on appearing in the text. The two authors who categorized the requests first by hand formed a detailed starting list [16] of about 10,500 relevant terms in order to include all relevant content words—even misspelled words and abbreviations. Since we focused on these words, greetings and function words such as "hello," "the," or "of" are not included and have therefore no effect upon the classification. As a next step, we clustered similar terms to create 4109 groups of terms called “parents” (for examples, see Table 1). The final dataset was a large table consisting of 998 rows (one row for each document analyzed) and 4109 columns (one column for each parent). The words in each document were analyzed to register how often each parent was represented in the text.

Text-Mining Strategies

To reduce the final dataset consisting of 988 rows and 4109 columns, we used three techniques (as different text-mining procedures): (1) indicator variables on the basis of Cramer’s V, (2) principle component analysis (PCA), and (3) singular value decomposition (SVD). The first strategy was developed by the authors. The second strategy used the indicator variables from the first strategy as input for PCA. The third strategy made use of a standard procedure from statistic software for SVD, SAS Text Miner (SAS, Carey, NJ, USA).

Cramer’s V

We calculated the average Cramer’s V statistic for the association of each of the 4109 “parents” with each category and the subsequent generation of indicator variables that sum for each category all Cramer’s V coefficients over the significant words. Cramer’s V is a chi-square-based measure of association

between nominal variables, with "1" indicating a complete positive association and "0" indicating no association at all. The coefficients were normalized according to the length of the texts (ie, the number of words). The selection criterion for including a parent term's Cramer's V was the error probability of the corresponding chi-square test. Its significance level was alternatively set at 1%, 2%, 5%, 10%, 20%, 30%, and 40%, leading to seven indicator variables per category.

Principle Component Analysis

We conducted PCA to reduce the seven indicator variables of varying significance levels per category into five orthogonal dimensions. PCA transforms a number of correlated variables into a few uncorrelated variables [17]. Each principal component is a linear combination of the original variables, with coefficients equal to the eigenvectors of the correlation. PCA can be used for dimensionality reduction in a dataset by retaining those characteristics of the dataset that contribute most to its variance. The data are transformed to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the principal component), the second greatest variance on the second coordinate, and so on [22].

Singular Value Decomposition

The 500 dimensional SVD was based on the standard settings of the SAS Text Miner software [23]. To understand SVD, the whole text of all requests can be visualized as being a document by term matrix, as described above. The text from each individual request (rows) is divided into its parent terms (columns) by listing the frequency of each term in a given text. Documents are represented as vectors with length m, where m is the number of unique terms indexed in the text. The original document by term matrix is transformed, or decomposed, into smaller matrices, thus, creating a factor space. An SVD projection is a linear combination of the singular values in a row or column of the term × document frequency matrix. A high number of SVD dimensions usually summarizes the data in a better way but requires significant computing resources [24,25].

Statistical Analyses

The sample was split into 75% training data and 25% test data. On the basis of our predictor variables (ie, 38×7 Cramer's V indicators per category, 38×5 principle components per category, and approximately 500 SVDs), we trained logistic regression models to predict the categories. However, if all these predictor variables would be used in a regression model, it would be rather unlikely to detect any significant variables since many of these are highly correlated. Therefore, we chose a more appropriate modelling approach, a stepwise logistic regression. The choice of predictive variables was carried out by an automatic procedure.

To assess the most appropriate model for a classification, we used the following selection methods: (1) Akaike Information Criterion, (2) Schwarz Bayesian Criterion, (3) cross validation misclassification of the training data (leave one out), (4) cross validation error of the training data (leave one out), and (5) variable significance based on an individually adjusted variable significance level for the number of positive cases. For a more detailed description of most of these selection criteria, see Beal [26].

We trained for each target category, each selection criterion, and each type of input variable (Cramer's Vs, principle components, SVDs) one logistic regression. This resulted in 1369 logistic regression models. The detailed notes and the table in the Multimedia Appendix make this procedure more transparent. For the final regression, we used meta-models, which proved the best for each of the 38 categories.

The complete training process produced an automatic method to evaluate both requests from the training sample and new requests. The corresponding software program is called score code. This score code is a function that generates, for any text (request), the probability of belonging to each of the 38 different categories.

To assess the accuracy of our approach, we calculated recall and precision as standard statistics in information retrieval and text mining for each of the 38 categories. Precision is the percentage of positive predictions that are correct (ie, a sort of specificity), whereas recall is the percentage of documents of a given category that were retrieved (sensitivity). We calculated recall and precision at the maximum F-measure [27]. To determine whether our approach yielded better results for precision and recall in the subject matter dimension or the expectation dimension, we compared the macroaverage values for precision and recall between both dimensions [28]. All statistical analyses were performed with SAS 9.1 (SAS, Carey, NJ, USA).

Results

Manual Classification

Table 2 shows the results of our manual classification of the 988 documents. A total of 102 (10%) documents fell into the category "in vitro fertilization (IVF)", 81 (8%) into the category "ovulation," 79 (8%) into "cycle," and 57 (6%) into "semen analysis." These were the four most frequent categories in the subject matter dimension (consisting of 32 categories). The expectation dimension comprised six categories; we classified 533 documents (54%) as "general information" and 351 (36%) as a wish for "treatment recommendations."

Table 2. Quality of automatic classification

Dimension	Requests No.	Training/Validation Ratio	Validation Data	
			Precision%[a]	Recall%[a]
Subject Matter				
abortion	40	30:10	91	100
abrasion	13	9:4	100	100
birth control pill	23	17:6	100	100
charges	25	18:7	100	100
clomifen	26	19:7	100	100
cryo transfer	13	9:4	100	75
cycle	79	59:20	80	86
cysts	16	12:4	100	100
endometriosis	11	8:3	75	100
examination of the oviduct	19	14:5	100	100
habitual abortion	17	12:5	100	100
hormones	36	27:9	78	78
insemination	29	21:8	100	100
intermenstrual bleeding	14	10:4	100	100
IVF	102	76:26	81	88
luteal phase defects	25	18:7	88	100
medical drugs	47	35:12	92	100
menstruation	35	26:9	90	100
multiples	7	5:2	100	100
naturopathy	33	24:9	90	100
nourishment	9	6:3	100	100
oviduct	16	12:4	100	100
ovulation	81	60:21	90	86
PCO	27	20:7	100	100
pregnancy symptoms	36	27:9	100	100
pregnancy test	30	22:8	88	88
pregnancy worries	49	36:13	100	92
semen analysis	57	42:15	88	93
sexual intercourse	14	10:4	100	100
sexual intercourse, problems	5	3:2	100	100
stimulation	40	30:10	63	100
thyroid glands	13	9:4	100	100
Expectations[b]				
current treatment	331	248:83	85	72
discussion of results	310	232:78	86	82
emotions	90	67:23	100	61
general information	533	399:134	92	84
interpretation of own situation	242	181:61	78	69
treatment options	351	263:88	82	81

[a]To calculate recall and precision, we first chose the best model according to the following selection criteria: Akaike's Information Criterion, Schwarz Baysian Criterion, cross validation misclassification of the training data, cross validation error of the training data; then we determined the optimum compromise between recall and precision by the F-measure.

[b]Multiple categories possible.

Automatic Classification

We used different selection criteria to find the best regression models for training and validation. In about half of the categories, the generation of indicator variables based on the chi-square analysis proved to be the best approach for automatic classification. Other categories were best predicted by using either PCA or SVD. Statistical details are shown in the Multimedia Appendix. A 100% precision and 100% recall was realized in 18 out of 38 categories on the validation sample (see Table 2). The lowest rates for precision and recall were 75% and 61%, respectively. The rates for precision and recall were, on average, somewhat lower in the expectations dimension (78.2% and 74.8%, respectively) compared to the subject matter dimension (93.6% and 96.4%, respectively).

Table 3 and Table 4 provide exemplary impressions of the power and the limits of the chi-square analysis. Table 3 lists the most significant words in the category "general information." Interestingly, nearly all of the first 50 words for the category "general information" were negatively associated. This means that the word "injection," for example, is a strong indicator that a document containing this word does not belong to this code. The 51st word ("fertile") was the first one with a positive Cramer's V; it represents a typical question about the fertile days of the menstrual cycle. For nearly all other categories, the most predictive words were positively associated with the respective category.

Table 4 lists the most significant words for the categories "oviduct" and "examination of the oviduct." These categories have been listed separately because "oviduct" was mainly associated with lay requests about reproductive medicine in general while "examination of the oviduct" was used in conjunction with questions about specific treatments or treatment options. Some of the predictive words (eg, "tube," "fallopian tube," "level") are the same in both categories. For example, the word "tube" appears in all requests that we categorized by hand as "oviduct" (n = 16), showing a strong predictive value (Cramer's V of 0.44). However, this word also appears in 79% of the requests that were categorized as "examination of the oviduct" (n = 19). Again, Cramer's V was high (0.37), signalling also the strong predictive value of this word for "examination of the oviduct". In this situation, only the summary of the Cramer's V statistic as an indicator variable guaranteed high precision and recall, and not a single word alone.

For some categories, the input variables also included variables from other categories, most often with a negative sign. For example, the meta-model for "pregnancy test" included a sample of words (as an indicator variable) predictive for the category "menstruation" with a negative sign. This means that absence of words predictive for "menstruation" was a strong indicator for the category "pregnancy test".

For other categories, consideration of a sender's expectation also contributed to a better classification of requests. For example, the meta-model for "hormones" included a sum of relevant terms (on the basis of Cramer's V) as well as significant terms demonstrating the expectation to learn more about one's own situation or to have laboratory data interpreted (both with negative signs, meaning that the absence of these expectations were, besides others, indicators for "hormones").

Table 3. Most predictive words for the category "general information"

Word	Frequency, No. (%)		Cramer's V	*P*
	In "General Information"	In Other Categories		
X-chromosome	70 (13)	143 (31)	− 0.22	< .001
injection	17 (3)	68 (15)	− 0.21	< .001
utrogest	7 (1)	45 (10)	− 0.19	< .001
clomifen	32 (6)	82 (18)	− 0.19	< .001
prescribe	10 (2)	45 (10)	− 0.17	< .001
write	21 (4)	59 (13)	− 0.16	< .001
med	45 (8)	88 (19)	− 0.16	< .001
drug	24 (5)	59 (13)	− 0.15	< .001
pill	20 (4)	53 (12)	− 0.15	< .001
value	48 (9)	88 (19)	− 0.14	< .001
[places 11-50]				
fertile	36 (7)	44 (10)	0.12	< .001

Table 4. Most predictive words for the categories “oviduct” (total requests = 16) and “examination of the oviduct” (total requests = 19)

Category Word	Requests in Which This Word Occurs[a] No. (%)	Cramer’s V	*P*
“Oviduct” (n = 16)			
tube	16 (100)	0.44	< .001
fallopian tube	16 (100)	0.44	< .001
removed	8 (50)	0.40	< .001
exception	2 (13)	0.35	< .001
away	8 (50)	0.29	< .001
link	7 (44)	0.28	< .001
move	1 (6)	0.25	< .001
obliterate	1 (6)	0.25	< .001
inappropriate	1 (6)	0.25	< .001
sterilisation	1 (6)	0.25	< .001
secretion	1 (6)	0.25	< .001
scar	1 (6)	0.25	< .001
opportunity	1 (6)	0.25	< .001
patent	1 (6)	0.25	< .001
open	1 (6)	0.25	< .001
consider	1 (6)	0.25	< .001
extensive	1 (6)	0.25	< .001
attachment	1 (6)	0.25	< .001
abandon	1 (6)	0.25	< .001
cut	1 (6)	0.25	< .001
tubal pregnancy	4 (25)	0.24	< .001
endoscopy	7 (44)	0.21	< .001
level	8 (50)	0.21	< .001
“Examination of the Oviduct” (n = 19)			
tube	15 (79)	0.37	< .001
fallopian tube	15 (79)	0.37	< .001
laparoscopy	11 (58)	0.35	< .001
endoscopy	12 (63)	0.35	< .001
X-ray	3 (16)	0.34	< .001
angiography	2 (11)	0.32	< .001
examination	4 (21)	0.32	< .001
level	12 (63)	0.30	< .001
penetrable	4 (21)	0.27	< .001
stomach	11 (58)	0.27	< .001
hsg	2 (11)	0.26	< .001
structure	12 (63)	0.23	< .001
cycle	1 (5)	0.23	< .001
adhere	1 (5)	0.23	< .001

[a]Some words in this table occur only once or twice (eg, “move”), but not at all in any of the other subject categories. Therefore, they still have predictive power (with a significant Cramer’s V).

Exemplary Comparison Between Automatic and Manual Classification

To give a more vivid picture of the results of our method, we present some of the visitors' requests, including our own manual classification and the automatic classification with scoring values for the probability of falling into a particular category (see Table 5). The first example is a very short request in which the sender wants to know whether a short cycle could be caused by a particular hormone. The automatic classification did not find the central topic of the request, probably because the term "prolactinspiegel" (prolactin level) was not recognized as "hormones." The subject category with the highest probability was "cycle," with a probability of only 2%—meaning that no classification was automatically assigned. In the two other examples, all our manual codes were recognized by the automatic classification. This was also the case in most other requests, representing a high sensitivity of our approach.

In several instances, and also in two of the three examples presented in Table 5, the automatic classification sometimes gave a high score not only for the correct subject category (as determined by the authors) but also for additional subject categories. In the second example, there was a high score for "stimulation" (in addition to the correct "IVF"), and the categories "clomifen" and "stimulation" scored highly in the third example (together with the correct category "multiples"). Consequently, precision, which is a measure of specificity, was not always entirely satisfactory. Some of these additional classifications such as "stimulation" in the second and third examples are provoked by the word "stimulation" or other misleading words in the request. While the additional categories in the automatic classification are not entirely correct, they are also not completely wrong. In all three of these examples, our classification according to the expectation of the sender was confirmed by the automatic classification with different probabilities. Only in the last example did the automatic classification also select "treatment options," which in fact is not entirely incorrect.

Table 5. Sample visitor requests and their classification

After a very long first half of my cycle (14-20 days), the second half of the cycle only takes 8 days. Is this probably because of an elevated prolactin level (I still nurse)? Many thanks for your answer, [Name]
Expert classification: hormones; general information; current treatment
Automated classification: cycle (2%); general information (99%), current treatment (97%)
We are in the middle of our second IVF cycle. During our first follicular puncture (first IVF), only one fallopian tube could be punctured [sic]. The second was hidden behind the uterus. However, at that time, the stimulation regime[n] was quite high. In the current cycle, I was stimulated with consideration. Therefore, only 11 follicles grew. At first, my question regarding sports activities during stimulation was answered with "no problems." After another inquiry I was told that I should stop playing badminton after the 8th day of stimulation. However, swimming would not be a problem. Because of badminton, torsion of the fallopian tube may happen. Today, follicular puncture took place. For the last time, I played badminton on day 8 of stimulation (only half out) but went swimming up to day 11 of stimulation (but not as "hard" as usual) because, supposedly, this should not have any effect.
Expert classification: IVF; general information; current treatment
Automatic classification: IVF (99%); stimulation (68%); general information (99%); current treatment (97%)
Right now, I am in the middle of my second insemination cycle (stimulation with Puregon 50 and Clomifen). Today, on day 12 of the cycle, 4 big follicles are visible. Now I shall decide whether to stop the cycle or to get inseminated. What are your thoughts about the risk of multiples? I would accept twins but not triplets. I am torn…. On one hand I would like to take the chance to get pregnant, but on the other hand I am afraid of multiples. Please tell me your opinion. Because of your experience you might be able to judge this matter much better. I appreciate your answer. Sincerely yours, [Name]
Expert classification: multiples; general information; current treatment
Automatic classification: multiples (98%); clomifen (68%); stimulation (54%); general information (67%); current treatment (98%); treatment options (53%)

Discussion

A combination of different text-mining strategies should classify requests to a medical expert forum into one or several of 38 categories, representing either the subject matter or the sender's expectations. This combined strategy yielded rates of precision and recall above 80% in nearly all categories. Even in the worst classified categories, the rates were at least above 60%.

Meaning of the Study

In order to evaluate these results, the exceptional character of this text-mining process should be considered. The documents to be classified were complex, sometimes rather long, and, most importantly, needed to be classified not only according to content but also to their (sometimes subliminal) expectations. We were able to show that a combination of different text-mining procedures was superior to a single method. Two factors have particularly contributed to this success: (1) an elaborated starting list and (2) a combination of chi-square statistics, PCA, and an SVD method. These factors mirror a recommendation and an experience reported by Balbi and Meglio [29], who built their specific text-mining strategy according to the "nature" of the data.

The creation of good starting (or stopping) lists is necessary to obtain valid and useful results, and comprehensive domain knowledge is essential for creating reasonable lists in the first place. The lists described here contain valuable expert knowledge in the field of involuntary childlessness. It seems reasonable to suppose that creating synonym lists in other medical areas could also be a powerful tool for successful text

mining in other Internet forums. In their extensive paper on predictive data mining, Bellazzi and Zupan [30] stress the importance of additional knowledge that domain experts can make use of for the modeling methods. This starting list demonstrated its full potential when used to generate indicator variables that summed all Cramer's V values for each request and each category over the significant words. This way, we escaped the danger of overestimating the predictive power of single words, especially if words are negated (eg, "I'm not interested in IVF" or "my cycle is not normal").

Nearly all words predictive of the category "general information" were negatively associated in the chi-square statistic. This seems to be a "perfect" finding and evidence for our content-related approach since any treatment with injections, for example, would belong to the categories "treatment options," "interpretation," or "current treatment" rather than the category "general information." It is precisely the lack of technical terms or results from prior investigations that defined this category.

Experts usually classify requests, such as the ones we analyzed, in a dichotomous way (ie, either they do or do not belong to a respective category). In contrast, automatic classification with a scoring system similar to the one presented in this paper gives a probability for any given request to fall into any of the categories. Especially in the case of complex texts, it seems appropriate to classify them into multiple dimensions and multiple categories. We defined a cutoff of 50% for our scoring system (ie, we defined a request to fall into a category if the respective score was over 50%). At the same time, it is possible to change the cutoff according to the purpose of an analysis. For example, if we are interested in recognizing possible health needs, a 50% cutoff may contribute to a high recall (sensitivity) so that we do not miss relevant requests. If we are interested in high precision (ie, specificity of classification) to sort out the requests and thus to support the experts' work, a higher cutoff may be reasonable. Our analysis procedure permits an easy assignment of different cutoff values.

There is another reason why this scoring procedure seems adequate or even superior to a dichotomous expert classification. When we analyze the sender's expectation, we are usually confronted with a mix of different expectations. In many cases, we classified a request into several expectation dimensions. This seems intuitively better represented by a scoring procedure such as the one presented in this paper. And even the subject matter classifications that we employed in our manual procedure as separated (disjunct) categories may not be as clear as they seem in many requests. It is rather likely that a given request may also fall into more than one subject matter, as demonstrated by the examples in Table 5, so that in these cases, a scoring procedure that also permits overlapping categories seems most appropriate [6,20]. In contrast, most studies, even if they have used a multidimensional categorizing scheme such as Shuyler and Knight [20], only permit one category per dimension.

As SVD is a powerful method for automatic classification, it seemed quite logical that this approach proved best to predict categories in about a quarter of instances. However, there is sometimes reluctance to use SVD-based classification strategies because this process can be controlled only to a limited degree [31]. In other words, text mining based on SVD is a procedure that cannot be consciously monitored. As a sort of black box, it automatically runs in the background and we have to rely on the validity of this procedure. In contrast, according to Reincke [31], the data-mining process should be mapped into a continuous IT flow that controls the entire information from the raw data, cleaning aggregation and transformation, analytical modeling, operative scoring, and last but not least, final deployment. In this sense, our analysis is actually far more transparent as demonstrated in the case of the predictive words given in Table 3. That is to say, our analysis not only yields good rates for precision and recall, but it also provides us with a complete view of the analytic process and thus contributes to face validity.

In the last decade, the medical profession has witnessed new developments whereby patients have become their own experts, often through the adoption of strategies to empower themselves [32] and often supported by the Internet [33,34] with email consultation services for electronic patient–caregiver communication [35]. A crucial factor to be able to make use of all of this potential information is time. The Internet is a rapid medium and when questions go unanswered for a few days, users are disappointed and may even resend their queries, as Marco et al [3] experienced in their Internet survey on AIDS and hepatitis. A complex technological solution such as that presented in this paper may effectively help medical experts to process the information needs of requests in advance and to accelerate response times. Once the information needs have been understood, it will also be possible to find similar previous requests, allowing experts to make efficient use of their earlier answers. This technology can therefore be used to both enable experts to answer requests promptly and to lighten their workload.

As a further advantage of our approach, we would like to emphasize our comprehensive list of categories. To date, analyses of email requests [5,6] have tried to categorize these requests in more or less simple categories, especially to learn more about information needs and the possible workload of experts. In contrast, we have been far more specific in the classification of the information needs with 32 categories representing the subject matter dimension. This detailed classification is exactly what experts need if machine-based analysis is to support their work.

Limitations of the Study

The classification of requests according to the senders' expectations could be improved. That this process is not optimal may be due to the somewhat vague definition of what exactly constitutes a certain patient's expectation, and this requires improvement if health experts are to make conclusions about the health needs of a population. However, the overall performance of the subject classification seems to be sufficient, so much so that semiautomated answers to senders' requests, in this medical area, may be a realistic option for the future.

Future Considerations

We consider there to be three relevant applications of our text-mining procedures in the near future:

1. If our scoring procedure proves successful in further tests, it could be integrated into the Rund ums Baby website to facilitate semiautomated answer proposals to be used by the experts and, in cases when classification accuracy is high, direct automated answers to the patients [36]. A multidimensional classification of texts, as in our approach, may be especially appropriate for this purpose since we recognize not only the plain content (ie, subject matter) but also the sender's expectations, something like a hidden subtext.
2. A retrospective application of the scoring procedure to all accumulated requests would allow their mapping into different categories, thus providing an objective historical seismograph and allowing a better understanding of medical and psychological needs that have yet to be met by the current health care system.
3. The scored database forms the basis for a sophisticated FAQ Internet page that does not address those questions and issues considered by experts to be the most important, as is usually the case, but one which is more oriented to the real needs of visitors and patients.

We are not aware of any studies that have tried to analyze similarly complex texts in Internet forums. Further studies are therefore needed to compare and refine our methodology. Then it should also be possible to decide which aspects of our text-mining strategies—the expert-based synonym list or the combination of different strategies—were most important for the success of our automatic classification.

Conclusions

Our analysis suggests a way of classifying and analyzing complex documents to provide a significant as well as a valid information source for politicians, administrators, researchers, and/or counselors. In the case of involuntary childlessness, it will be possible to fulfill not only patients' information and health needs with this Internet expert forum, but also to analyze and follow-up these needs over long periods of time. These techniques also seem promising for the analysis of large samples of documents from other Internet health forums, chat rooms, or email requests to doctors.

Acknowledgments

We are indebted both to Ulrich Schneider, operator of the Rund ums Baby website for his permission to use data from this Internet forum, and Christian Schulz, webmaster, for his technical support. In addition, the authors would like to thank Stephanie Heinemann, who carefully read and discussed the many drafts of this manuscript with the authors and helped them to express their ideas and results as exactly as possible.

Conflicts of Interest

HWM is one of the experts who work for the Rund ums Baby forum on an honorary basis. UR is an employee of SAS Institute Germany and works in the Enterprise Intelligence Competence Centre.

Multimedia Appendix 1

Statistical details of the automatic classification (explanation)

[PDF file (Adobe Acrobat), 824 KB - jmir_v11i3e25_app1.pdf]

Multimedia Appendix 2

Statistical details of the automatic classification (table)

[PDF file (Adobe Acrobat), 1.5 MB - jmir_v11i3e25_app2.pdf]

References

1. Umefjord G, Sandström H, Malker H, Petersson G. Medical text-based consultations on the Internet: a 4-year study. Int J Med Inform 2008 Feb;77(2):114-121. [Medline: 17317292] [doi: 10.1016/j.ijmedinf.2007.01.009]
2. Umefjord G, Hamberg K, Malker H, Petersson G. The use of an Internet-based Ask the Doctor Service involving family physicians: evaluation by a web survey. Fam Pract 2006 Apr;23(2):159-166 [FREE Full text] [Medline: 16464871] [doi: 10.1093/fampra/cmi117]
3. Marco J, Barba R, Losa JE, de la Serna CM, Sainz M, Lantigua IF, et al. Advice from a medical expert through the Internet on queries about AIDS and hepatitis: analysis of a pilot experiment. PLoS Med 2006 Jul;3(7):e256 [FREE Full text] [Medline: 16796404] [doi: 10.1371/journal.pmed.0030256]
4. Widman LE, Tong DA. Requests for medical advice from patients and families to health care providers who publish on the World Wide Web. Arch Intern Med 1997 Jan 27;157(2):209-212. [Medline: 9009978] [doi: 10.1001/archinte.157.2.209]
5. Eysenbach G, Diepgen TL. Patients looking for information on the Internet and seeking teleadvice: motivation, expectations, and misconceptions as expressed in e-mails sent to physicians. Arch Dermatol 1999 Feb;135(2):151-156 [FREE Full text] [Medline: 10052399] [doi: 10.1001/archderm.135.2.151]

6. Huang JY, Al-Fozan H, Tan SL, Tulandi T. Internet use by patients seeking infertility treatment. Int J Gynaecol Obstet 2003 Oct;83(1):75-76. [Medline: 14511879] [doi: 10.1016/S0020-7292(03)00253-4]
7. Himmel W, Meyer J, Kochen MM, Michelmann HW. Information needs and visitors' experience of an Internet expert forum on infertility. J Med Internet Res 2005;7(2):e20 [FREE Full text] [Medline: 15998611] [doi: 10.2196/jmir.7.2.e20]
8. Lasker RD. Strategies for addressing priority information problems in health policy and public health. J Urban Health 1998 Dec;75(4):888-895. [Medline: 9854249] [doi: 10.1007/BF02344517]
9. Weiss SM, Indurkhya N, Zhang T, Damerau FJ. Text Mining: Predictive Methods for Analyzing Unstructured Information. New York: Springer; 2005.
10. Cohen KB, Hunter L. Getting started in text mining. PLoS Comput Biol 2008 Jan;4(1):e20 [FREE Full text] [Medline: 18225946] [doi: 10.1371/journal.pcbi.0040020]
11. Feldman R, Sanger J. The Text Mining Handbook: Advanced Approaches in Analyzing UnstructuredData. Cambridge: Cambridge University Press; 2007.
12. Rund ums Baby website. URL: http://www.rund-ums-baby.de/ [WebCite Cache ID 5hYMJzQ1U]
13. Noorbala AA, Ramezanzadeh F, Abedinia N, Naghizadeh MM. Psychiatric disorders among infertile and fertile women. Soc Psychiatry Psychiatr Epidemiol 2009 Jul;44(7):587-591. [Medline: 19023508]
14. Inhorn MC, Birenbaum-Carmeli D. Assisted Reproductive Technologies and Culture Change. Annu Rev Anthropol 2008;37(1):177-196. [doi: 10.1146/annurev.anthro.37.081407.085230]
15. Himmel W, Michelmann HW. Involuntary childless couples in family practice: recommendations for patient management. In: Allahbadia GN, Merchant R, editors. Gynecological endoscopy and infertility. New Delhi, India: Jaypee Brothers Medical Publishers; 2005:147-153.
16. Sanders A, DeVault C, editors. Using SAS at SAS: The Mining of SAS Technical Support. Proceedings of the Twenty-Ninth Annual SAS Users Group International Conference. Cary, NC: SAS Institute Inc; 2004 Presented at: :Paper 010-29.
17. Albright R. Taming Text with the SVD. URL: ftp://ftp.sas.com/techsup/download/EMiner/TamingTextwiththeSVD.pdf [WebCite Cache ID noarchive]
18. Himmel W, Kroll F. Text mining to analyze the requests to a medical expert forum [in German]. In: Beryer D, Ortseifen C, editors. SAS in Hochschule und Wirtschaft - Proceedings der 8. Konfernez der SAS-Anwender in Forschung und Entwicklung (KSFE). Aachen, Germany: Shaker; 2004 Presented at: p. 69-80.
19. ; Cleveland Clinic Heart Center. The Heart Forum. URL: http://www.medhelp.org/forums/cardio/wwwboard.html [WebCite Cache ID 5a49Nun18]
20. Shuyler KS, Knight KM. What are patients seeking when they turn to the Internet? Qualitative content analysis of questions asked by visitors to an orthopaedics Web site. J Med Internet Res 2003 Oct 10;5(4):e24 [FREE Full text] [Medline: 14713652] [doi: 10.2196/jmir.5.4.e24]
21. Mezzich JE, Kraemer HC, Worthington DR, Coffman GA. Assessment of agreement among several raters formulating multiple diagnoses. J Psychiatr Res 1981;16(1):29-39. [Medline: 7205698] [doi: 10.1016/0022-3956(81)90011-X]
22. Jolliffe IT. Principal Component Analysis. 2nd edition. New York: Springer; 2002.
23. Reincke U, editor. Profiling and classification of scientific documents with SAS Text Miner. Presented at: The third &Knowledge Discovery" Workshop; 2003 Oct 6-8; Karlsruhe, Germany URL: http://km.aifb.uni-karlsruhe.de/ws/LLWA/akkd/8.pdf [WebCite Cache ID 5hPJNHci4]
24. Berry MW, Dumais ST, Letsche TA. Computational methods for intelligent information access. URL: http://www.cs.utk.edu/~berry/sc95/sc95.html [WebCite Cache ID 5a49IiDlA]
25. Evangelopoulos N, editor. Text Mining (SAS Education, Data Mining, lecture 11). Denton, TX: University of North Texas; 2002 Presented at: URL: http://www.coba.unt.edu/itds/courses/dsci4520/slides/DSCI4520_TextMining_11.ppt [WebCite Cache ID 5fXBE0Naa]
26. Beal DJ, editor. Information criteria methods in SAS for multiple linear regression models. SESUG Proceedings. Paper SA05. Raleigh, NC: North Carolina State University; 2007. URL: http://analytics.ncsu.edu/sesug/2007/SA05.pdf [WebCite Cache ID 5hPJLTFj3]
27. Korfhage RR. Information Storage and Retrieval. London: Wiley; 1997.
28. Sebastiani F. Machine learning in automated text categorization. ACM Comput Surv 2002;34:1-47 ISI: 000175267600001.
29. Balbi S, Di Meglio E. Contributions of textual data analysis to text retrieval. In: Banks D, House L, McMorrisi FR, Arabie P, Gaul W, editors. Classification, Clustering, and Data Mining Applications. Berlin, Germany: Springer; 2004 Presented at: p. 511-520.
30. Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. Int J Med Inform 2008 Feb;77(2):81-97. [Medline: 17188928] [doi: 10.1016/j.ijmedinf.2006.11.006]
31. Reincke U. Directions of analytics, data and text mining - a software vendor's view. Presented at: ECML/PKDD 2006 Workshop on Practical Data Mining:Applications, Experiences and Challenges; 2006 Sept 22; Berlin, Germany URL: http://www.ecmlpkdd2006.org/ws-pdmaec.pdf [WebCite Cache ID 5hPJJkCfP]
32. Anderson RM, Funnell MM. Patient empowerment: reflections on the challenge of fostering the adoption of a new paradigm. Patient Educ Couns 2005 May;57(2):153-157. [Medline: 15911187] [doi: 10.1016/j.pec.2004.05.008]

33. Tuil WS, Verhaak CM, Braat DD, de Vries Robbé PF, Kremer JA. Empowering patients undergoing in vitro fertilization by providing Internet access to medical data. Fertil Steril 2007 Aug;88(2):361-368. [Medline: 17416366] [doi: 10.1016/j.fertnstert.2006.11.197]
34. Dumitru RC, Bürkle T, Potapov S, Lausen B, Wiese B, Prokosch HU. Use and perception of internet for health related purposes in Germany: results of a national survey. Int J Public Health 2007;52(5):275-285. [Medline: 18030943] [doi: 10.1007/s00038-007-6067-0]
35. Nijland N, van Gemert-Pijnen J, Boer H, Steehouder MF, Seydel ER. Evaluation of internet-based technology for supporting self-care: problems encountered by patients and caregivers when using self-care applications. J Med Internet Res 2008;10(2):e13 [FREE Full text] [Medline: 18487137] [doi: 10.2196/jmir.957]
36. Himmel W, Reincke U, Michelmann HW. Using text mining to classify lay requests to a medical expert forum and to prepare semiautomatic answers. In: Proceedings of the SAS Global Forum 2008 Conference. Cary, NC: SAS Institute Inc; 2008 Presented at: Paper 210-2008 URL: http://www2.sas.com/proceedings/forum2008/210-2008.pdf [WebCite Cache ID 5hPJHMZaN]

Abbreviations

FAQ: frequently asked question
ICSI: intracytoplasmic sperm injection
IVF: in vitro fertilization
PCA: principle component analysis
SVD: singular value decomposition

Edited by K El Emam; submitted 15.08.08; peer-reviewed by M Sokolova, N Nijland, F Caropreso; comments to author 07.11.08; revised version received 25.03.09; accepted 06.04.09; published 22.07.09.

Please cite as:
Himmel W, Reincke U, Michelmann HW
Text Mining and Natural Language Processing Approaches for Automatic Categorization of Lay Requests to Web-Based Expert Forums
J Med Internet Res 2009;11(3):e25
URL: http://www.jmir.org/2009/3/e25/
doi:10.2196/jmir.1123
PMID:19632978

Original Paper

Nationwide Implementation of Hello World: A Dutch Email-Based Health Promotion Program for Pregnant Women

Mariska Bot[1], MSc; Ivon EJ Milder[1], PhD; Wanda JE Bemelmans[1], PhD

National Institute for Public Health and the Environment, Bilthoven, the Netherlands

Corresponding Author:
Ivon EJ Milder, PhD
Centre for Prevention andHealth Services Research (PZO)
National Institute for Public Health and the Environment
PO Box 1, 3720 BA Bilthoven
The Netherlands
Phone: +31 30 2742871
Fax: +31 30 2744407
Email: ivon.milder@rivm.nl

Abstract

Background: In November 2006, an email-based health promotion program for pregnant women was implemented nationally in the Netherlands. The program consisted of emails containing quizzes with pregnancy-related questions tailored to the number of weeks of pregnancy. Emails were sent out once every 4 weeks, up to a maximum of nine emails.

Objectives: The aims of the study were (1) to assess the recruitment of participants and their representativeness of the Dutch population and (2) to study differences in recruitment, program use, and program appreciation among women with different levels of education.

Methods: Data from 13,946 pregnant women who enrolled during the first year of the program were included. Upon registration, participants were asked how they found out about the program and subsequently received an email questionnaire to assess demographic, lifestyle, and Internet characteristics. Program use was tracked, and participants were classified into five user groups (inactive to very active). Program appreciation (low, intermediate, and high) was assessed twice with an email questionnaire that was sent after the woman had received her third and sixth quiz email. Information about pregnant women and their characteristics was obtained from Dutch registries to assess representativeness of the study population.

Results: About 8% of the pregnant women in the Netherlands enrolled in the program. Immigrants were underrepresented, and women with a low level of education seemed to be slightly underrepresented. Most women knew about the program from a promotional email sent by the organization (32%), followed by the Internet (22%) and midwives (16%). Women with little education were more often inactive users of the program than were highly educated women (15% vs 11%, $P < .001$), whereas highly educated women were more often very active users compared with women with little education (25% vs 20%, $P< .001$). However, women with less education were more likely than women with more education to have a high appreciation of the program after receiving three quiz emails (52% vs 44%, $P = .001$).

Conclusions: In this real-life setting, pregnant women can be reached through an email-based health promotion program. Selective engagement by education level remains a challenge.

(J Med Internet Res 2009;11(3):e24) doi:10.2196/jmir.1183

KEYWORDS

Health promotion; Internet; pregnancy; email

Introduction

Pregnancy is a time when women are more conscious of health issues than during any other period in their life [1]. This may be a time when behavior can be positively changed, to have a long-term effect on the health of the mother, child, and family [2]. Nowadays, most women seek information about their pregnancy on the Internet [3-6], particularly for first-time pregnancies [4]. However, online medical information varies in quality and is often incomplete [7].

In 2006, the Dutch Ministry of Health initiated Hello World, an innovative email-based program for pregnant women in

which reliable information about a healthy pregnancy is brought together into one health promotion program. The program was tested in a pilot study in Amsterdam [8,9] and was then improved and implemented nationwide in the Netherlands. The aim of the program was to provide comprehensive health information for pregnant women. Although knowledge on its own is criticized as being insufficient for behavior change, knowledge is a prerequisite for behavior change [10].

To the best of our knowledge, Hello World is the first eHealth program aimed at pregnant women that was implemented nationally. We are aware of only one previous study of an eHealth program for pregnant women conducted locally in Taiwan [11], which was aimed at promoting breastfeeding.

To be effective as a health promotion program, it is essential to reach the target group. Thus, efficacious recruitment strategies are needed. Several studies have investigated which recruitment strategies for Internet-based eHealth promotion programs result in high participation, but results were mixed [12-18]. In addition to enrollment, engagement is an important condition for efficacy since exposure to the health information is necessary. Unfortunately, the high attrition rate of Internet-based health promotion interventions is still an issue [19,20]. Both enrollment and engagement in health communication programs are often selective, favoring participation of relatively highly educated persons with a healthy lifestyle [19-21]. Indeed, the pilot study of Hello World showed that highly educated women not only enrolled in the program more often, but also used the program more intensively and longer than less-educated women [8].

The present study analyzes data of participants who enrolled during the first year of the nationwide-implemented program. The main purpose of the study was to describe the participants, their representativeness of all pregnant women in the Netherlands, and how they were recruited in a real-life setting. Our second aim was to compare differences in recruitment channels, program use, and program appreciation among women with varying levels of education.

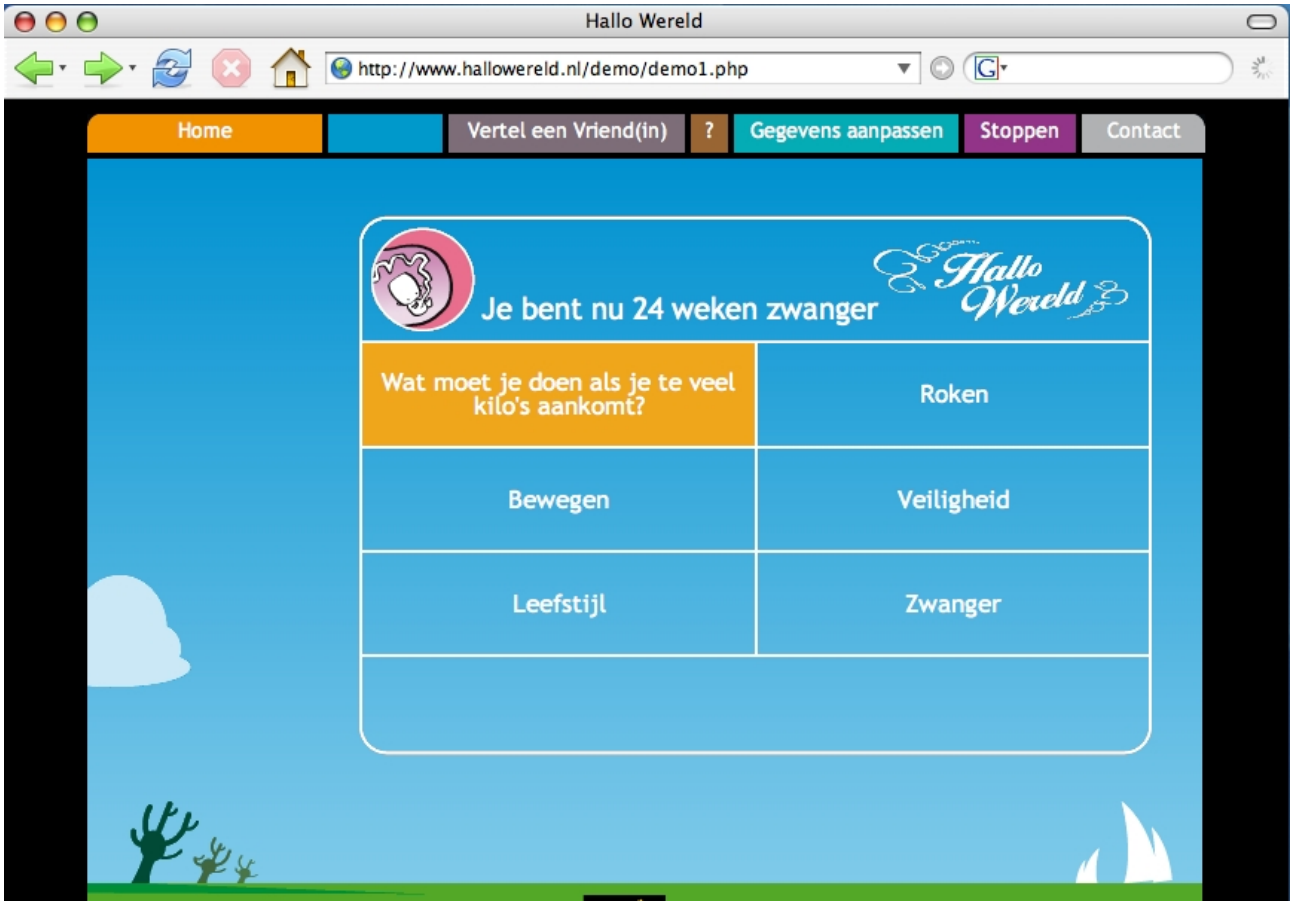
Methods

Intervention

The program started on November 13, 2006, and is still available online. The present study analyzes data of pregnant women who enrolled in the first year. The program consisted of emails containing quizzes with pregnancy-related questions that were tailored to the number of weeks of pregnancy. Emails were sent once every 4 weeks. The first quiz email was sent at 8 weeks of pregnancy and the last one at 40 weeks. Women could enroll in the program at no cost anytime during their pregnancy and only received the quizzes that were applicable to their stage of pregnancy. Each quiz consisted of one question on each of the following six topics: nutrition, smoking, physical activity, safety, lifestyle/care, and pregnancy. Women could complete the quiz after receiving an email that contained one of the six questions as an example and a hyperlink to the entire quiz. Each question had two answer possibilities. Selecting an answer was followed by an automatic message on the website informing the women whether their answer was correct, an explanation of the correct answer, and, in most cases, a practical tip and a hyperlink to a related website providing additional or more detailed information. Several national health promotion institutes and the national platform of midwives were responsible for the content of the topic relevant to their expertise. Furthermore, women could ask online experts pregnancy-related questions.

The program was specifically developed to reach women with basic literacy skills, including the use of short text blocks and plain language. Figure 1 shows an example of a quiz question. For an overview of all quiz questions, translated into English, see the Multimedia Appendix.

Figure 1. Screenshot of the quiz for week 24



Study Participants

Pregnant women signed up for participation in Hello World on the website. Those who subscribed between the start of the program and November 13, 2007, were included in the study (n = 14,154). For the women who registered more than once, the most active account or the most recent inactive account was included; all other double accounts were excluded (n = 127). Additionally, data of nonpregnant persons and test users (n = 19) or those with an unlikely number of weeks of pregnancy (less then 0 or more than 40 weeks; n = 62) were excluded. Data of 13,946 participants were used for the analyses. Women who mentioned miscarriage as the reason for unsubscribing (n = 260) were included for assessment of enrollment, representativeness, and recruitment, but they were excluded from the analysis of program use and appreciation. The study was carried out in accordance with Dutch privacy legislation. Informed consent was obtained for all participants via the user conditions of the website that participants had to agree to when registering for Hello World. It was explained to the participants that answering the research questionnaires was voluntary and not a prerequisite for (further) use of the website.

Recruitment

Several recruitment strategies were used to promote the program. From January 2007 onward, 50,000 Hello World leaflets were included in pregnancy present boxes, which could be requested by pregnant women on another pregnancy-related website. In addition, between January and November 2007, a promotional email was sent by the organization who implemented the intervention to almost 76,000 women who had requested a pregnancy present box. Furthermore, banners were placed on several websites, and advertisements for Hello World appeared when performing pregnancy-related searches in Google. As well, a total of 45,000 Hello World flyers were distributed in the majority of Dutch midwifery practices. In addition, recruitment took place through traditional mass media as advertisements in magazines and through an information stand at a national pregnancy fair.

Measures

Demographic, Lifestyle, and Internet Characteristics

Upon enrollment, the women were asked to enter their first name, date of birth, email address, expected date of delivery, and how they found out about the website. In this paper, the last variable will be referred to as the “recruitment channel,” and it was categorized into the following eight types: midwife, email, Internet, pregnancy present box, pregnancy fair, peers, media (radio, television, magazines, newspapers), and other/unknown. Subsequent to enrollment, women received a baseline questionnaire by email to assess demographic, lifestyle, and Internet characteristics. Age was calculated from the date of birth. Ages < 15 and > 45 years were regarded as errors and were therefore not included in the age description. The number of weeks of pregnancy (0-14 weeks, 15-27 weeks, and 28-40 weeks) was calculated from the date of enrollment and the expected delivery date, assuming a pregnancy duration of 40 weeks.

The education level of the women was assessed as the highest level completed and was categorized as low (lower secondary education or less), intermediate (higher secondary education), or high (college or university) [22]. Women were regarded as immigrants when at least one of their parents was born outside the Netherlands [22]. If the woman herself was also born outside the Netherlands, she was considered a first-generation immigrant. Furthermore, immigrant women were divided into immigrants from Western counties and non-Western countries based on their parents' country of birth. Information on parity, job participation, overweight before pregnancy, smoking status, and alcohol consumption was also obtained with the questionnaire.

Sufficient fruit and vegetable intake was based on meeting the Dutch guidelines for minimum intake levels of fruit and vegetables (ie, two pieces of fruit or one piece of fruit and one glass of fruit juice per day, and 150-200 g of vegetables per day) [23]. Sufficient physical activity was defined as being moderately active for at least 30 minutes on five or more days per week, according to the Dutch recommendations for physical activity [24]. No folic acid intake was defined as not currently using or not having used folic acid tablets. The questionnaire also included questions about Internet availability at home, time spent online per week (< 1 hour, 1-2 hours, and > 2 hours), and the use of the Internet for obtaining information about pregnancy, health, and/or parenting.

Program Use

Program use was registered continuously and included the quiz emails sent to the participant and the quiz questions accessed by the participant. The quiz emails contained a personal identification code to track the use of the quiz questions. All user data up to the end of August 2008 were included, thereby covering a whole pregnancy period after the last moment of registration in this study (November 13, 2007) to ensure that every woman could have received all quiz emails. If at least one of the quiz questions was opened by the participant, that woman was defined as a "user" of the quiz for that specific week. To describe program use, only the women who enrolled before or during week 28 of their pregnancy were included since they could have received a substantial number of quiz emails.

Program users were classified into five categories: (1) inactive (did not use any of the received quizzes), (2) ceased (did not use the last two received quizzes or unsubscribed themselves from the program after using one or more quizzes), (3) moderately active (used at least two of the last quizzes, but less than 2/3 of all quizzes), (4) active (used at least two of the last quizzes, and at least 2/3 of all quizzes), and (5) very active (used all quizzes). The quiz email for week 40 was not taken into account since we expected that not all pregnant women would be able to open quiz emails around the time of delivery. Indeed, the data showed that this quiz was completed less often than other quizzes.

Program Appreciation

Appreciation of the program was assessed by means of an online self-administered questionnaire that was sent in the week after the participants received their third quiz email and their sixth quiz email. As women could enroll themselves anytime throughout their pregnancy, timing of the questionnaire varied among participants. The same questionnaire was used for both evaluations and included eight positively formulated statements about the program that could be rated on a five-point Likert scale ("totally agree" to "totally disagree"). Examples of the statements include "the information on Hello World is reliable" and "the information on Hello World is recent." Based on the number of statements for which the response was "agree" or "totally agree," program appreciation was categorized into low (agreed with 0-4 statements), intermediate (agreed with 5-6 statements), or high appreciation (agreed with 7-8 statements).

Data Analyses

Data analyses were performed in 2008. Response rates of the questionnaire were based on the number of questionnaires sent to the participants and the number of questionnaires completed by the participants. The demographic characteristics of the pregnant women participating in the email-based health promotion program were described and compared with existing data of pregnant women in the Netherlands [22]. Because younger women may not have completed their education, age-stratified education levels were compared. Demographic, lifestyle, and Internet characteristics; recruitment channel; program use; and program appreciation were compared among participants with different levels of education. For program use and education level, additional analyses were conducted after stratification for parity.

For all analyses, differences between education levels were tested with chi-square tests for categorical variables and analysis of variance for continuous variables.

Because education level was associated with several characteristics and lifestyle factors, the associations between education level and program use and program appreciation, after adjustment for these characteristics, were tested using ordinal logistic regression. The final multivariate models were constructed by including all variables that were significantly associated with the independent variable (program use or appreciation) in the bivariate models (including education level and one other characteristic) or the multivariate model (including all characteristics). Statistical significance was set at $P < .05$. All data analyses were carried out using SAS for Windows, version 9.1 (SAS Institute, Cary, NC, USA).

Results

Representativeness of the Participants

Based on the number of enrolled women in the program (n = 13,946) and the number of living newborns in the Netherlands in 2007 (n = 18,1336), we estimated that approximately 8% of all pregnant women in the Netherlands enrolled in the Hello World program (Table 1). Most participants (9210/12,802, 72%) in our study were aged 25-35 years. Although there were fewer low-educated women in the program than in the Netherlands for the 15-25 and 35-45 year age groups, the percentage of low-educated women among the 25-35 year age group was comparable to that of the Netherlands. In our study, the percentage of immigrants (1583/8833, 18%) was smaller than

the overall proportion found in the Netherlands (24%). The majority of participants (5568/8833, 63%) were pregnant with their first child, which was higher than the percentage of first-born babies in the Netherlands.

Table 1. Characteristics of the participants of Hello World

Characteristic		Hello World	The Netherlands[a]
Number of pregnant women		13,946	18,1336[b]
		Mean ± SD	Mean
Age of those pregnant with their first child (years)[c]		28.3 ± 4.4	29.4[d,e]
Age at enrollment (years)[c]		29 ± 4.5	31.1[d,f]
		No. (%)	%
Response rate for baseline questionnaire[g]		8833/13,942 (63)	–
Pregnant with first child		5568 (63)	45[h]
Age group			
15-25 years		2368 (18)	10[d]
25-35 years		9210 (72)	64[d]
35-45 years		1224 (10)	26[d]
Education level			
15-25 years	Low	594 (43)	50[i]
	Intermediate	660 (47)	40[i]
	High	142 (10)	10[i]
25-35 years	Low	1032 (17)	17[i]
	Intermediate	2366 (40)	46[i]
	High	2544 (43)	37[i]
35-45 years	Low	154 (19)	25[i]
	Intermediate	260 (33)	47[i]
	High	386 (48)	28[i]
Employment			
Full-time job (≥ 32 hours/week)		4120 (47)	NA[j]
Part-time job (< 32 hours/week		3058 (35)	NA[j]
No job		1619 (18)	24[k]
Immigrant		1583 (18)	24[k]
First generation		635 (7)	14[k]
Non-Western		753 (9)	14[k]

[a]Data from Statistics Netherlands [22].

[b]Number of living newborns in the Netherlands in 2007.

[c]Only those aged ≥ 15 and ≤ 45 years were included. For Hello World participants, age at enrollment included 12,802 women and age of those pregnant of their first child included 5252 women.

[d]Only pregnant women.

[e] Mean age of those pregnant with their first child in 2007.

[f]Mean age of women at the time of delivery in 2007.

[g]Proportion based on the number of returned questionnaires and the number of questionnaires sent to the women.

[i]Including pregnant and nonpregnant women.

[j] Not available.

[k]Including pregnant and nonpregnant women aged 15-45 years.

Recruitment Channel

As shown in Table 2, the most reported recruitment channel was email (32%), followed by the Internet (22%) and midwives (16%). This order was observed for all education levels. However, relatively more women with lower education mentioned email as the recruitment channel (38%) than did highly educated women (24%; χ^2= 154.5, $P < .001$); the reverse was observed for midwives as the recruitment channel (14% vs 20% for low and high education, respectively; $\chi^2 = 38.1$, $P < .001$).

Table 2. Recruitment channel by education level

Recruitment Channel	Total[a] (n = 13,946)		Low (n = 1940)		Intermediate (n = 3548)		High (n = 3317)	
	No.	%	No.	%	No.	%	No.	%
Email	4399	32	744	38	1280	36	804	24
Internet (surfing, Google, other websites)	3023	22	384	20	767	22	779	23
Midwife	2214	16	279	14	560	16	674	20
Pregnancy present box	1116	8	164	8	285	8	303	9
Peers	1041	7	135	7	229	6	220	7
Media (radio, television, magazine, newspaper)	653	5	76	4	146	4	197	6
Pregnancy fair	585	4	35	2	100	3	115	3
Other/unknown	915	7	123	6	181	5	225	7

[a] The low, intermediate, and high education levels do not add to the total group because not all women completed the baseline questionnaire with the question about education.

Characteristics by Education Level

Table 3 presents the characteristics of the pregnant women by education level. More highly educated women (2257/3312, 68%) were pregnant with their first child than were women with a low level of education (1115/1937, 58%; $\chi^2 = 65.7$, $P < .001$). Low-education women more often smoked during pregnancy, less often had sufficient fruit and vegetable intake, and less often used folic acid than high-education women. Almost all women had access to the Internet at home and used it for obtaining pregnancy-related information.

Table 3. Demographic, lifestyle, and Internet characteristics by education level[a]

	Low		Intermediate		High		χ^{2b}	P [c]
	No.	%[b]	No.	%[b]	No.	%[b]		
Age (years)	1780	*27.5(5.2)*[d]	3286	*28.8 (4.3)*[d]	3072	*30.7 (3.5)*[d]	*351.0*[e]	< .001
Pregnancy								
Pregnant with first child	1115/1937	*58*	2179/3544	*61*	2257/3312	*68*	*65.7*	< .001
Number of weeks of pregnancy at enrollment	1940	*18.4 (9.2)*[d]	3548	*18.0 (9.3)*[d]	3317	*17.4 (9.1)*[d]	*8.1*[e]	< .001
1st trimester	795/1940	*41*	1510/3548	*43*	1481/3317	*45*	*15.2*	.004
2nd trimester	783/1940	*40*	1394/3548	*39*	1326/3317	*40*		
3rd trimester	362/1940	*19*	644/3548	*18*	510/3317	*15*		
Lifestyle characteristics								
Overweight before pregnancy (BMI ≥ 25 kg/m^2) [f]	752/1913	*39*	1297/3495	*37*	845/3280	*26*	*137.8*	< .001
Smoking							*906.1*	< .001
Smoker	465/1938	*24*	329/3543	*9*	57/3317	*2*		
Cessation because of pregnancy	443/1938	*23*	666/3543	*19*	399/3317	*12*		
Nonsmoker	1030/1938	*53*	2548/3543	*72*	2861/3317	*86*		
Alcohol use during pregnancy	41/1934	*2*	87/3538	*2*	142/3308	*4*	*26.9*	< .001
Sufficient vegetable intake[g]	253/1225	*21*	518/2263	*23*	674/2106	*32*	*69.2*	< .001
Sufficient fruit intake[g]	482/1266	*38*	1037/2337	*44*	1133/2159	*52*	*71.0*	< .001
Sufficient physical activity[h]	553/1325	*42*	885/2393	*37*	819/2239	*37*	*10.8*	.005
No folic acid used	304/1932	*16*	287/3537	*8*	141/3308	*4*	*210.4*	< .001
Internet characteristics								
Internet at home	1847/1930	*96*	3428/3544	*97*	3259/3309	*98*	*38.6*	< .001
Internet use							*36.4*	< .001
< 1 hour/week	222/1933	*11*	370/3540	*10*	261/3311	*8*		
1-2 hours/week	610/1933	*32*	1217/3540	*34*	1030/3311	*31*		
> 2 hours/week	1101/1933	*57*	1953/3540	*55*	2020/3311	*61*		
Use of Internet for information about pregnancy	1654/1932	*86*	3176/3542	*90*	2987/3310	*90*	*29.5*	< .001

[a] Due to missing values, the totals in each column vary. Italics represents significant results.

[b] Unless otherwise noted.

[c] The chi-square test was used for categorical variables, and analysis of variance was used for continuous variables.

[d] Mean (SD).

[e] F-value.

[f] Body mass index (BMI) was calculated as the self-reported weight before pregnancy (kg) and height (m^2).

[g] Meeting the current Dutch guidelines for sufficient fruit and vegetable intake. Sufficient fruit intake: at least two pieces of fruit per day or one piece supplemented by fruit juice. Sufficient vegetable intake: at least 150-200 g of vegetables per day.

[h] At least 30 minutes of being moderately active on five or more days per week.

Program Use

For the analyses of program use, we included only the women who enrolled in week 28 of pregnancy or earlier (11,415/13,686; 83%). On average, the participants received a mean of 5.6 ± 1.8 quiz emails and opened 2.8 ± 2.3 of them to complete at least one of the six questions. Women completed, on average, 3.8 ± 1.6 of the six available questions for each opened quiz.

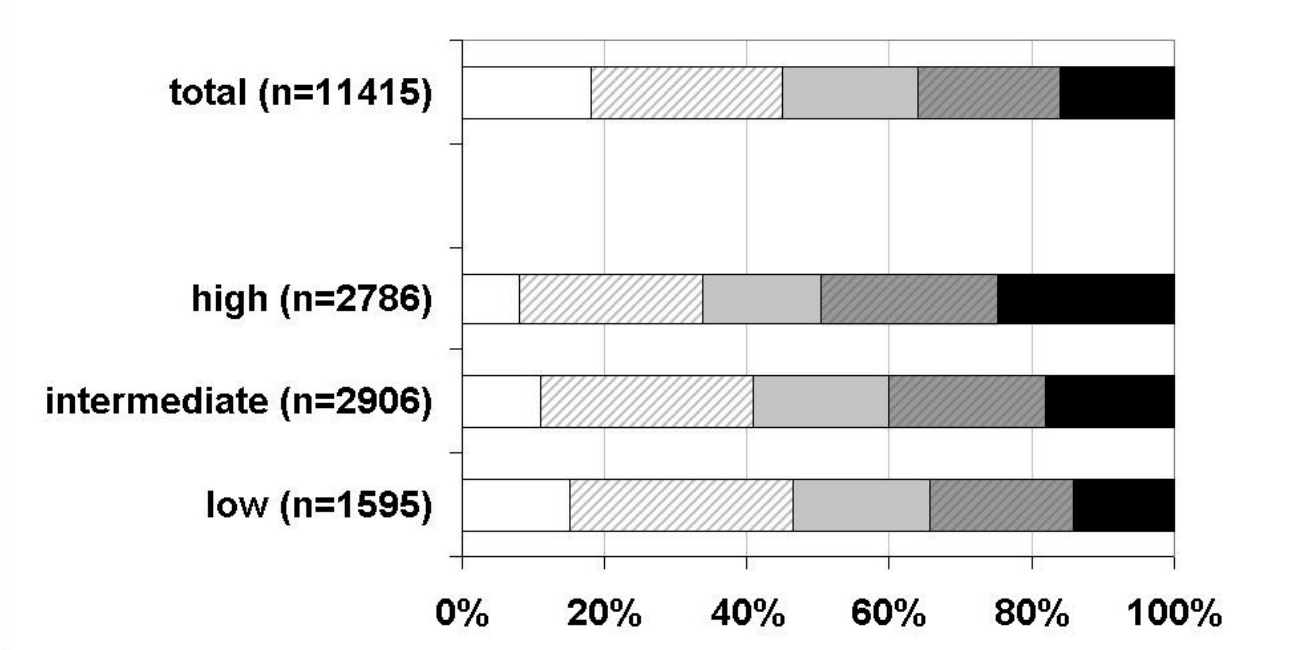
Figure 2 shows the distribution of program users for all women who registered in week 28 of pregnancy or earlier. Program users were classified into five categories: (1) inactive (did not use any of the received quizzes), (2) ceased (did not use the last two received quizzes or unsubscribed themselves from the program after using one or more quizzes), (3) moderately active (used at least two of the last quizzes, but less than 2/3 of all quizzes), (4) active (used at least two of the last quizzes, and at least 2/3 of all quizzes), and (5) very active (used all quizzes). The quiz email of week 40 was not taken into account, and miscarriages were excluded. In Figure 2, the numbers of women with low, intermediate, and high education do not add to the total group because not all women completed the baseline questionnaire with the question about education level.

About 18% (2075/11,415) of the women never opened a quiz email, 27% (n = 5166) ceased participating, and 16% (n = 1849) opened all quiz emails received. Inactivity was observed more among the women with low education (15%, 242/1595) than among the women with high education (8%, 220/2786; χ^2 = 56.7, P < .001). Similarly, ceased participation was more common among low-education women (31%, 497/1595) than among high-education women (26%, 712/2786; χ^2 = 20.5, P < .001).

Very active users were more likely to be women with a high level of education (25%, 710/2786) than with a low level of education (14%, 229/1595; χ^2 = 93.7, P < .001). When repeating these analyses stratified for those pregnant for the first time and those not, comparable results for education level were found (results not shown).

Because education level was associated with program use as well as with several lifestyle factors (see Table 3), we tested the independent association of education level and program use using multivariate ordinal regression. In the multivariate model, higher education, being older, being pregnant with the first child, number of weeks of pregnancy, being a nonsmoker, and not being overweight were independently associated with program use, whereas alcohol use and folic acid use were not.

Figure 2. Type of program user by education level



Program Appreciation

The first appreciation questionnaire was completed by 3763/12,201 women who received it (response rate 31%); the second appreciation questionnaire was completed by 1926/8666 women who received it (response rate 22%). Half (n = 1840) of the 3763 women who answered the first questionnaire had a high appreciation of the program, more than one quarter (n = 1091) showed an intermediate appreciation, and less than one quarter (n = 832) had a low appreciation. Low-education women were more likely to show a high appreciation (52%, 307/593) of the program than high-education women (44% 539/1219; χ^2 = 17.9, P = .001) in the first appreciation questionnaire. A similar pattern was seen for the second appreciation questionnaire (data not shown).

Because education level was associated with program appreciation as well as with several lifestyle factors (see Table 3), we tested the independent association of education level and program appreciation using multivariate ordinal regression. In the multivariate model, being younger, being pregnant with the first child, and not using alcohol during pregnancy were

independently associated with higher program appreciation, while education level and overweight were not.

Discussion

Our study describes the enrollment and engagement rates in a national email-based health education program aimed at pregnant women in a real-life setting. We estimate that about 8% of all pregnant women in the Netherlands enrolled in the Hello World program. Women with less education seemed to be slightly underrepresented, and immigrants seemed to be underrepresented compared to the Dutch female population aged 15-45 years. In particular, women who were pregnant with their first child enrolled in the program. Most women cited email as their recruitment channel, followed by the Internet and midwives. In agreement with results of the pilot study [8], program use differed between women with low and high levels of education. Low-education women were more satisfied with the program than high-education women.

To our knowledge, Hello World is the first national online health education program aimed at pregnant women. The program incorporates several elements that have shown to be effective in online health education programs. These elements include tailored messages, which are known to be appreciated by pregnant women [3]. In addition, emails were sent every 4 weeks to inform the women about the availability of new quizzes, since previous studies showed that email reminders have a positive effect on the number of website revisits and were suggested as a strategy to retain participants [16,25]. Furthermore, a combination of online as well as offline recruitment strategies were used so that women who used the Internet less often could also find out about the program.

In the Netherlands, the majority of pregnant women visit midwives as the first line of care during pregnancy [26]. One of the tasks of midwives is to provide information about a healthy pregnancy. Thus, this is a natural setting for health promotion and interventions. A few studies have evaluated the effectiveness of interventions (eg, breastfeeding and smoking cessation) but with mixed results [27,28]. To the best of our knowledge, none of these interventions have been implemented nationally.

Strengths and Limitations

Strengths of our study were its large sample size and real-life setting. In addition, program use was registered objectively and was not self-reported.

Some limitations need to be stressed. The response rate for the questionnaires dropped from 63% for the baseline questionnaire to 22% for the second appreciation questionnaire. Additional analyses showed that lower response rates on the questionnaires were related to low education level, younger age, and nonactive program use. Further, participation rates could not be calculated for the different recruitment strategies. In addition, it is likely that not all preterm births and miscarriages were reported. Since delivery and miscarriage will probably result in cessation of program use, this may have led to an underestimation of program use in pregnant women.

We estimated that 8% of all pregnant women in the Netherlands enrolled in the Hello World program. Our estimation has a few drawbacks. First, it is based on the number of living newborns in the Netherlands in 2007, thereby not taking miscarriages into account. However, as most miscarriages take place early in pregnancy and the women enrolled at 18 weeks of pregnancy on average, the number of miscarriages probably was small. Second, we cannot exclude that some women registered twice using different email addresses. In addition, nonpregnant persons or Dutch-speaking women living abroad may also register for the program. This may have led to an overestimation of the enrollment rate in our study. Third, our enrollment rate was based on all pregnant women, including those who were not exposed to the intervention. This must be kept in mind when comparing our enrollment rate with other studies. For example, the pilot study showed an enrollment rate of 17% [8], but this rate was based on the number of pregnant women who were invited to participate by midwives. Since only a fraction of all pregnant women probably knew about the program in our study, our enrollment rate appears reasonable.

We had to rely on existing data for describing the representativeness of the participants. In the pilot study, selective enrollment was observed, favoring enrollment of higher-educated women [8]. In contrast, in this study, the percentage of women with a low level of education was comparable to that in the Dutch female population aged 25-35 years. Almost three quarters of the women enrolled in Hello World belong to this age group. However, women with less education have more children than higher-educated women in the Netherlands [29] and may therefore be slightly underrepresented. Pregnant women with a low level of education need more attention since previous studies have shown that these women more often smoked during pregnancy [30], less often breastfed their children up to 6 months after delivery [31], and less often were aware of the benefits of folic acid and used folic acid less often during pregnancy [32,33]. Our study confirmed less favorable health behaviors in lower-educated women compared to higher-educated women, except for alcohol consumption and physical activity. The large number of women pregnant with their first child is desirable since they have the highest need for pregnancy-related information [3].

In our study, more participants were recruited by online strategies than by offline strategies. Similar results were found by Feil et al [14] and Thüring et al [13]. Further, in our study, lower-educated women were more often recruited by email than were highly educated women, whereas the reverse was observed for recruitment by midwives. It is possible that these differences can be partly explained by differences in exposure.

Eysenbach [34] has discussed the importance of studying attrition in Internet-based interventions. High dropout rates in Internet interventions are common, and selective dropout is frequently observed [20,34]. In our study, one sixth of the participants never opened a quiz email, and more than a quarter stopped participating. We found a higher level of program use among highly educated pregnant women compared to lower-educated pregnant women. This is similar to the pilot study [8] but is in contrast to Verheijden et al [20], who found no differences between education level for the usage of an online

health promotion program in a subsample of the general Dutch population. Although the level of engagement was lower among low-education women, these women appreciated the program more than highly educated women. It is possible that the program was too easy and concise for highly educated women and that they therefore appreciated the program less.

Future Work

We have measured both enrollment and engagement, which are important prerequisites for exposure to health messages in the program. However, the effect of these messages on health outcomes is unknown [3] and should be investigated in randomized controlled trials. More research is needed on the differences of processing health information among persons with varying levels of education. Effective methods for increasing enrollment and engagement need to be further studied.

Conclusions

In conclusion, this study showed that about 8% of the pregnant women in the Netherlands could be reached through an Internet-based health promotion program. Lower-educated women seemed to be slightly underrepresented. Lower-educated women were less actively engaged than highly educated women but appreciated the program more. Most women were recruited by online recruitment strategies rather than traditional channels. Reducing selective attrition remains a challenge in Web-delivered health promotion programs.

Acknowledgments

This study was funded by The Dutch Ministry of Health, Welfare and Sports. The authors would like to thank Martin de Jong from OOiP for his assistance in collecting the program usage data and Maarten Schipper for his statistical advice. Also thanks to Moniek van Zutphen and Salome Scholtens for their comments on earlier versions.

Conflicts of Interest

None declared.

Multimedia Appendix

Overview of quiz questions

[PDF file (Adobe PDF), 21 KB - jmir_v11i3e24_app1.pdf]

References

1. Anderson AS. Symposium on 'nutritional adaptation to pregnancy and lactation'. Pregnancy as a time for dietary change? Proc Nutr Soc 2001 Nov;60(4):497-504. [Medline: 12069403] [doi: 10.1079/PNS2001113]
2. Verbeke W, De Bourdeaudhuij I. Dietary behaviour of pregnant versus non-pregnant women. Appetite 2007 Jan;48(1):78-86. [Medline: 17005297] [doi: 10.1016/j.appet.2006.07.078]
3. Bernhardt JM, Felter EM. Online pediatric information seeking among mothers of young children: results from a qualitative study using focus groups. J Med Internet Res 2004 Mar 1;6(1):e7 [FREE Full text] [Medline: 15111273] [doi: 10.2196/jmir.6.1.e7]
4. Szwajcer EM, Hiddink GJ, Koelen MA, van Woerkum CM. Nutrition-related information-seeking behaviours before and throughout the course of pregnancy: consequences for nutrition communication. Eur J Clin Nutr 2005 Aug;59 Suppl 1:S57-S65. [Medline: 16052197] [doi: 10.1038/sj.ejcn.1602175]
5. Larsson M. A descriptive study of the use of the Internet by women seeking pregnancy-related information. Midwifery 2009 Feb;25(1):14-20. [Medline: 17408822] [doi: 10.1016/j.midw.2007.01.010]
6. Shaw E, Howard M, Chan D, Waters H, Kaczorowski J, Price D, et al. Access to web-based personalized antenatal health records for pregnant women: a randomized controlled trial. J Obstet Gynaecol Can 2008;30:38-43. [Medline: 18198066]
7. ; Rand Health (Prepared for the California HealthCare Foundation). Proceed with Caution: A Report on the Quality of Health Information on the Internet. Complete Study. Oakland, CA: California HealthCare Foundation; 2001. URL: http://www.chcf.org/documents/consumer/ProceedWithCautionCompleteStudy.pdf [WebCite Cache ID 5cjqMo9pV]
8. van Zutphen M, Milder IE, Bemelmans WJ. Usage of an online healthy lifestyle program by pregnant women attending midwifery practices in Amsterdam. Prev Med 2008 Jun;46(6):552-557. [Medline: 18302971] [doi: 10.1016/j.ypmed.2008.01.003]
9. van Zutphen M, Milder IE, Bemelmans WJ. Integrating an eHealth program for pregnant women in midwifery care: a feasibility study among midwives and program users. J Med Internet Res 2009;11(1):e7 [FREE Full text] [Medline: 19275983] [doi: 10.2196/jmir.988]
10. Green L, Kreuter MW. Health Promotion Planning: An Educational and Environmental Approach. 2nd edition. Mountain View, CA: Mayfield; 1991.

11. Huang MZ, Kuo SC, Avery MD, Chen W, Lin KC, Gau ML. Evaluating effects of a prenatal web-based breastfeeding education programme in Taiwan. J Clin Nurs 2007 Aug;16(8):1571-1579. [Medline: 17655546] [doi: 10.1111/j.1365-2702.2006.01843.x]
12. Martin-Diener E, Thüring N. Response to a single media announcement for an internet based hepa-intervention. Presented at: Hepa-symposium; September 1-2, 2002; Helsinki, Finland URL: http://www.physicalactivityandhealth.ch/hepa.ch/gf/active-online/publications/02_helsinki_poster.pdf [WebCite Cache ID 5cjpv76fj]
13. Thüring N, Martin-Diener E, Martin B. Effectiveness of an interactive Internet program promoting physical activity: the feasibility of an Internet-based, randomized study design. Presented at: ECSS; July 9-12, 2003; Salzburg, Germany URL: http://www.physicalactivityandhealth.ch/hepa.ch/gf/active-online/publications/03_ECSS_poster.pdf [WebCite Cache ID 5cjq5gPSy]
14. Feil EG, Noell J, Lichtenstein E, Boles SM, McKay HG. Evaluation of an Internet-based smoking cessation program: lessons learned from a pilot study. Nicotine Tob Res 2003 Apr;5(2):189-194. [Medline: 12745491] [doi: 10.1080/1462220031000073694]
15. Koo M, Skinner H. Challenges of Internet Recruitment: A Case Study with Disappointing Results. J Med Internet Res 2005;7(1):e6 [FREE Full text] [Medline: 15829478] [doi: 10.2196/jmir.7.1.e6]
16. Spittaels H, De Bourdeaudhuij I. Implementation of an online tailored physical activity intervention for adults in Belgium. Health Promot Int 2006 Dec;21(4):311-319 [FREE Full text] [Medline: 16963783] [doi: 10.1093/heapro/dal030]
17. Gordon JS, Akers L, Severson HH, Danaher BG, Boles SM. Successful participant recruitment strategies for an online smokeless tobacco cessation program. Nicotine Tob Res 2006 Dec;8 Suppl 1(1):S35-S41. [Medline: 17491169] [doi: 10.1080/14622200601039014]
18. Franklin PD, Rosenbaum PF, Carey MP, Roizen MF. Using sequential e-mail messages to promote health behaviors: evidence of feasibility and reach in a worksite sample. J Med Internet Res 2006;8(1):e3 [FREE Full text] [Medline: 16585028] [doi: 10.2196/jmir.8.1.e3]
19. Glasgow RE, Nelson CC, Kearney KA, Reid R, Ritzwoller DP, Strecher VJ, et al. Reach, engagement, and retention in an Internet-based weight loss program in a multi-site randomized controlled trial. J Med Internet Res 2007;9(2):e11 [FREE Full text] [Medline: 17513282] [doi: 10.2196/jmir.9.2.e11]
20. Verheijden MW, Jans MP, Hildebrandt VH, Hopman-Rock M. Rates and determinants of repeated participation in a web-based behavior change program for healthy body weight and healthy lifestyle. J Med Internet Res 2007;9(1):e1 [FREE Full text] [Medline: 17478410] [doi: 10.2196/jmir.9.1.e1]
21. Thoolen B, de Ridder D, Bensing J, Gorter K, Rutten G. Who participates in diabetes self-management interventions?: Issues of recruitment and retainment. Diabetes Educ 2007;33(3):465-474. [Medline: 17570877] [doi: 10.1177/0145721707301491]
22. Statistics Netherlands. Voorburg/Heerlen. URL: http://www.cbs.nl/ [WebCite Cache ID 5cjqZmMCX]
23. ; Health Council of the Netherlands. Richtlijnen goede voeding 2006 [Guidelines on healthy nutrition 2006]. Den Haag, the Netherlands: Health Council of the Netherlands; 2006. URL: http://www.gr.nl/pdf.php?ID=1479&p=1 [WebCite Cache ID 5hQmck4hM]
24. Kemper HGC, Ooijendijk WTM, Stiggelbout M. Consensus over de Nederlandse Norm voor Gezond Bewegen [Consensus about the Dutch recommendation for physical activity to promote health]. Tijdschr Soc Gezondheidsz 2000;78:180-183.
25. Leslie E, Marshall AL, Owen N, Bauman A. Engagement and retention of participants in a physical activity website. Prev Med 2005 Jan;40(1):54-59. [Medline: 15530581] [doi: 10.1016/j.ypmed.2004.05.002]
26. Althuizen E, van Poppel MN, Seidell JC, van der Wijden C, van Mechelen W. Design of the New Life(style) study: a randomised controlled trial to optimise maternal weight development during pregnancy. [ISRCTN85313483]. BMC Public Health 2006;6(1):168 [FREE Full text] [Medline: 16800869] [doi: 10.1186/1471-2458-6-168]
27. Kools EJ, Thijs C, Kester AD, van den Brandt PA, de Vries H. A breast-feeding promotion and support program a randomized trial in The Netherlands. Prev Med 2005 Jan;40(1):60-70. [Medline: 15530582] [doi: 10.1016/j.ypmed.2004.05.013]
28. de Vries H, Bakker M, Mullen PD, van Breukelen G. The effects of smoking cessation counseling by midwives on Dutch pregnant women and their partners. Patient Educ Couns 2006 Oct;63(1-2):177-187. [Medline: 16406475] [doi: 10.1016/j.pec.2005.10.002]
29. van Agtmaal-Wobma E, van Huis M. De relatie tussen vruchtbaarheid en opleidingsniveau van de vrouw [The relationship between fertility and educational level of the woman]. CBS Bevolkingstrends. 2008. URL: http://www.cbs.nl/NR/rdonlyres/891A0CBE-B05B-48C2-B1CE-7C273CE69152/0/2008k2b15p32art.pdf [WebCite Cache ID 5cjqDcgWJ]
30. Lu Y, Tong S, Oldenburg B. Determinants of smoking and cessation during and after pregnancy. Health Promot Int 2001 Dec;16(4):355-365 [FREE Full text] [Medline: 11733454] [doi: 10.1093/heapro/16.4.355]
31. Flacking R, Nyqvist KH, Ewald U. Effects of socioeconomic status on breastfeeding duration in mothers of preterm and term infants. Eur J Public Health 2007 Dec;17(6):579-584 [FREE Full text] [Medline: 17392294] [doi: 10.1093/eurpub/ckm019]
32. de Walle HE, Cornel MC, de Jong-van den Berg LT. Three years after the dutch folic acid campaign: growing socioeconomic differences. Prev Med 2002 Jul;35(1):65-69. [Medline: 12079442] [doi: 10.1006/pmed.2002.1040]

33. Timmermans S, Jaddoe VW, Mackenbach JP, Hofman A, Steegers-Theunissen RP, Steegers EA. Determinants of folic acid use in early pregnancy in a multi-ethnic urban population in The Netherlands: The Generation R study. Prev Med 2008 Jul 2. [Medline: 18644404] [doi: 10.1016/j.ypmed.2008.06.014]
34. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [Medline: 15829473] [doi: 10.2196/jmir.7.1.e11]

Edited by G Eysenbach; submitted 11.12.08; peer-reviewed by H Spittaels, P Franklin; comments to author 05.02.09; revised version received 25.03.09; accepted 28.04.09; published 30.07.09.

Please cite as:
Bot M, Milder IEJ, Bemelmans WJE
Nationwide Implementation of Hello World: A Dutch Email-Based Health Promotion Program for Pregnant Women
J Med Internet Res 2009;11(3):e24
URL: http://www.jmir.org/2009/3/e24/
doi:10.2196/jmir.1183
PMID:19674957

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org