# Journal of Medical Internet Research

Journal Impact Factor (JIF) (2022): 7.4
Volume 15 (2013), Issue 4 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

## Contents

### Original Papers

## Review



## Letter to the Editor

Original Paper

# Multilevel Growth Curve Analyses of Treatment Effects of a Web-Based Intervention for Stress Reduction: Randomized Controlled Trial

Filip Drozd[1,2], MPhil; Sabine Raeder[2], PhD; Pål Kraft[1,2], PhD; Cato Alexander Bjørkli[2], PhD

[1]Research and Development, Changetech AS, Oslo, Norway

[2]Department of Psychology, University of Oslo, Oslo, Norway

**Corresponding Author:**
Filip Drozd, MPhil
Research and Development
Changetech AS
Changetech AS
Gaustadalléen 21
Oslo, N-0349
Norway
Phone: 47 97516188
Fax: 47 22604427
Email: fd@changetech.no

## Abstract

**Background:** Stress is commonly experienced by many people and it is a contributing factor to many mental and physical health conditions, However, few efforts have been made to develop and test the effects of interventions for stress.

**Objective:** The aim of this study was to examine the effects of a Web-based stress-reduction intervention on stress, investigate mindfulness and procrastination as potential mediators of any treatment effects, and test whether the intervention is equally effective for females as males, all ages, and all levels of education.

**Methods:** We employed a randomized controlled trial in this study. Participants were recruited online via Facebook and randomly assigned to either the stress intervention or a control condition. The Web-based stress intervention was fully automated and consisted of 13 sessions over 1 month. The controls were informed that they would get access to the intervention after the final data collection. Data were collected at baseline and at 1, 2, and 6 months after intervention onset by means of online questionnaires. Outcomes were stress, mindfulness, and procrastination, which were all measured at every measurement occasion.

**Results:** A total of 259 participants were included and were allocated to either the stress intervention (n=126) or the control condition (n=133). Participants in the intervention and control group were comparable at baseline; however, results revealed that participants in the stress intervention followed a statistically different (ie, cubic) developmental trajectory in stress levels over time compared to the controls. A growth curve analysis showed that participants in the stress intervention (unstandardized beta coefficient [B]=–3.45, *P*=.008) recovered more quickly compared to the control group (B=–0.81, *P*=.34) from baseline to 1 month. Although participants in the stress intervention did show increases in stress levels during the study period (B=2.23, *P*=.008), long-term stress levels did decrease again toward study end at 6 months (B=–0.28, *P*=.009). Stress levels in the control group, however, remained largely unchanged after 1 month (B=0.29, *P*=.61) and toward 6 months (B=–0.03, *P*=.67). Mediation analyses showed nonlinear (ie, cubic) specific indirect effects of mindfulness and a linear specific indirect effect of procrastination on stress. In simple terms, the intervention increased mindfulness and decreased procrastination, which was related to lower stress levels. Finally, the effect of the stress intervention was independent of participants' gender, age, or education.

**Conclusions:** The results from this randomized controlled trial suggest that a Web-based intervention can reduce levels of stress in a normal population and that both mindfulness and procrastination may be important components included in future eHealth interventions for stress.

**Trial Registration:** International Standard Randomized Controlled Trial Number (ISRCTN): 25619675; http://controlled-trials.com/ISRCTN25619675 (Archived by Webcite at http://www.webcitation.org/6FxB1gOKY)

***(J Med Internet Res 2013;15(4):e84)*** doi:10.2196/jmir.2570

**KEYWORDS**

stress, multilevel modeling, randomized controlled trial, mindfulness, procrastination, multiple mediation, multiple moderation, Web

## Introduction

Symptoms of stress, such as fatigue, mood changes, and muscle pain, are very common in the general population. More than 90% of the Norwegian population reported having several such symptoms during the past 30 days [1]. According to an international survey, approximately 75% of the general population in developed countries report feeling stressed on a daily basis [2]. Furthermore, the most recent Stress in America survey conducted by the American Psychological Association [3] shows that more than half of the American population report that they recognize when they are feeling stressed; however, less than 1 in 3 report successfully managing their stress levels.

For most people, stress may be perceived as such a minor problem that it does not require any treatment seeking or professional assistance [4]. However, major savings in health care costs can be achieved by reducing stress levels or eliminating some of the subjective health complaints. Although the psychobiological mechanisms remain elusive, stress is a risk factor for a wide range of mental and physical health problems, such as cardiovascular disease [5], diabetes [6], and depression [7]. In fact, of all the health conditions predicted by the World Health Organization as having the greatest disease burden by 2030 [8], many have a common contributing underlying factor of stress as either causing or exacerbating disease. Moreover, as many as 40% to 50% of work-related illnesses are related to stress [9,10]. Consequently, it is important to reduce stress in the general population, but this requires scalable interventions with a potentially high reach.

Systematic reviews demonstrate that face-to-face stress-reduction interventions are effective for reducing stress and various health problems [11-14], but their scalability is limited. On the other hand, eHealth interventions have the scalability potential to reach the general population; however, only a small number of studies have focused on stress reduction. Research has mostly focused on posttraumatic stress [15] or stress as a component in interventions primarily aimed at other health problems, such as diabetes [16] or alcohol use [17].

### Web-Based Stress-Reduction Interventions

Most eHealth interventions for stress in the general population have been evaluated in workplace settings and have shown varying results. The earliest studies have documented intervention effects for anxiety and depression [18], stress responses and job satisfaction [19], and beneficial psychophysiological effects on stress [20]. More recent studies have failed to find any effects on stress [21,22], and a few studies that compared Web-based versus therapist-supported stress management interventions demonstrated only short-term and small effects on stress [23,24].

Studies outside of the workplace setting have shown more unequivocal results. Two studies reported improved outcomes for Web-based family or parental stress interventions [25,26], whereas 2 other studies found reduced health distress among participants with various chronic diseases [27,28]. Another study that recruited participants through the Internet and newspaper articles also observed greater improvements in the treatment group [29].

A few studies have also evaluated the impact of stress management as an add-on component to existing eHealth interventions. These studies, however, showed varying results just like Web-based stress-reduction interventions in workplace settings. Christensen et al [30] did not find any additional contribution of stress reduction for depression, whereas Richards et al [31] found only short-term effects of adding stress management for panic disorder. Prochaska and colleagues [32] demonstrated that the add-on of a tailored component to a brief health risk intervention increased the number of participants that were effectively managing their stress.

### Mechanisms of Change

One reason why findings on Web-based interventions for stress are inconsistent, may be the “black box” phenomenon, or lack of understanding as to how and why some interventions work or do not work. Therefore, it is important to investigate the role of potential mediators and moderators of treatment effects. In this study, the Web-based stress-reduction intervention made use of 2 central intervention components—mindfulness and procrastination—both of which are associated with stress.

Mindfulness involves being in the present moment and accepting thoughts and feelings as they occur in a nonjudgmental way [33]. A meta-analysis has shown that mindfulness can have a broad range of health benefits [34] and that mindfulness-based stress-reduction interventions are generally effective [12]. It appears that mindfulness mediates the effect of interventions on stress [35] and it is associated with higher levels of self-regulation [36] which can facilitate deliberate actions to regulate behavior [37] and lessen avoidant coping [38,39]. The latter is a form of procrastination that is characterized by a voluntary extension of the temporal sequencing between an intended course of action and goal-directed behavior, despite one’s expectation of being worse off than before the delay [40]. Numerous studies have shown the negative effects of procrastination, including its relationship to stress [41]. According to the procrastination-health model, procrastination creates unnecessary stress and delays the onset of health promoting behaviors [42]. It is a strategy that brings immediate, albeit temporary, relief from unpleasant or distressing events [43], but ultimately the event remains unresolved.

Temporal dimensions are clearly important to mindfulness, procrastination, and stress, although few theories explicitly specify changes that occur over time or time as a cause of changes in any of these constructs. For example, the key characteristic of mindfulness is a temporal orientation at the present moment that requires a temporal and attentional shift to a state of awareness. However, the temporal course of changes in mindfulness has not yet been fully explored. Studies on

procrastination, on the other hand, have indicated that procrastinators experience less stress early on, but more stress closer to a deadline as compared to nonprocrastinators [44]. More recent studies have investigated temporal changes in procrastination using growth curve approaches, and most suggest that procrastination is characterized by a hyperbolic or quadratic function [45]. It is also reasonable to assume that the development of stress changes over time. Daily hassles or acute experiences of stress (ie, meeting a deadline, car trouble, or negative affect) that typically affect a person within hours on the same day of occurrence, are highly transient and rarely affect a person the next day as major stressful life events [46]. Thus, modeling time as an independent variable is important and allows one to represent change or the dynamic relationships between variables, although it remains elusive as to what kind of developmental trajectories one can expect over the course of time in intervention settings.

### Moderating Effects on Stress

In addition to identifying mechanisms of change over time, one may expect variations in intervention efficacy among participants (eg, not all participants improve). Thus, it is interesting to identify participants who benefit the most or participants for who the intervention shows contraindications. For example, the effects of the Web-based intervention on stress responses and job satisfaction, as mentioned previously, were shown particularly effective among males and younger employees [19]. In terms of stress, demographic differences between participants, such as gender, age, and education, can be expected to have varying intervention effects. The reason is that there are demographic differences in stress and how people manage their stress. In general, women report higher levels of stress compared with men [47]. This can be due to women's multiple roles [48] or the fact that the roles (eg, caregiving) typically assigned to women are stressful [49]. When it comes to age, people use more problem-focused coping strategies and less avoidance coping strategies as they grow older [50,51]. People also develop the ability to self-regulate emotions when dealing with stress as their age increases [52]. This may explain why researchers have found that adults in their 20s report more perceived stress than those in their 50s [53]. But with education, the picture is less clear. A recent large survey demonstrated that work-related stress is associated with higher education [54], whereas previous studies have shown that lower education is a risk factor for stress [55,56].

### Aims of the Study

This study aimed to test whether treatment was predictive of participants' initial status and different trajectory changes in stress across time. First, it was hypothesized that participants in the Web-based stress-reduction intervention would exhibit lower stress scores at the end of the treatment compared to the beginning, as measured by log server registrations. Second, it was hypothesized that the intervention would reduce levels of stress as measured by online survey data over a period of 6 months as compared to a control group. The control group was expected to remain at approximately the same stress level throughout the study period. Third, the effect of the intervention was expected to be, at least, partially mediated by mindfulness and procrastination over time. Finally, the effect of the intervention was examined with respect to moderating effects of gender, age, and education on the treatment effect on stress over the study period.

## Methods

### Design

The study was a randomized controlled trial consisting of 2 groups to test the effectiveness of a Web-based stress management intervention. Participants were randomized either to the fully automated Web-based Less Stress (LS) intervention or a waitlist control group to test for the natural course of participants' levels of stress. No unexpected events occurred after the commencement of the intervention (eg, bug fixes, downtimes, email delivery service failures, content changes). Participants in the control group received the intervention after the final data collection. The trial received its ethical approval by the Norwegian Social Science Data Services (reference number: 26816).

### Participants and Recruitment

The study was a Web-based trial without any face-to-face components as part of the recruitment procedure, intervention, or follow-up. Participants were recruited online through a master's student's social network on Facebook. In total, 320 first-degree contacts were invited to participate and forward the invitation to their network (ie, viral recruitment).

Potential participants clicked on a link posted on Facebook and were redirected to an external website containing study information and a consent form. Participants had to confirm that they had read the study information and submit the informed consent before they could proceed to the Web-based baseline questionnaire. Eligible participants were implicitly required to (1) read and understand Norwegian, (2) explicitly state that they were 18 years or older, and (3) fill in their email address.

A total of 326 participants were assessed for eligibility. Sixty-five (19.9%) participants did not provide a (valid) email address, and 2 (0.6%) participants reported being younger than 18 years. These 67 (20.6%) potential participants were excluded before randomization. The final sample size that was randomized consisted of 259 participants.

### Randomization

Every participant had an equal probability of being assigned to either the LS or control group. The allocation ratio was set to 1:1 and a series of zeros and ones were generated for each participant using a random integer generator [57]. Because recruitment was carried out through a private and social online network and participants were potentially identifiable through their email addresses, another research member on the team conducted the randomization procedure. This was done to avoid experimenter biases interfering with the randomization. As an extra precaution, email addresses were concealed during randomization.

## Intervention

The LS intervention is a fully automated and Web-based intervention developed for people who feel stressed or experience a lot of negative emotions (for screenshots, see Figure 1). Its objective is to have users learn about stress, build awareness of sources of stress, and prevent or manage prolonged or high levels of stress. LS uses an eclectic approach and includes evidence-based information and exercises that have been documented to be directly or indirectly effective for stress management, such as mindfulness [12] and metacognitive exercises [58]. The LS intervention consists of 13 sessions over a period of 4 weeks. Every Monday, Wednesday, and Friday, users receive an email with a unique hyperlink. By clicking on the hyperlink, users are directed to a sequence of Web pages that are unique for that particular session. Every session is designed to take approximately 10 minutes to complete. It is a prerequisite that the user completes a session successfully before proceeding to the next session. In this way, the user proceeds through a predetermined therapeutic chronology of sessions with restricted degrees of freedom (ie, tunneled design). For a demonstration, see [59].

Each session contains 2 components. The first component is psychoeducational and addresses some stress-related topics (see Table 1). The second section provides users with techniques, exercises, and homework designed to address the particular topic presented in the psychoeducational section. Psychoeducational information is presented by a young male agent, whereas tasks and exercises are presented by a young female, both accompanied by text designed in such a way "as if they were talking". The role of the personal computer agent equals that of a domain expert that guides the user. In this way, knowledge and information are represented in a form that is presumably similar to that of a human therapist or expert. Text is presented in short sentences and with a limited amount of text per Web page (approximately 80 words). Techniques and exercises often include audio files (eg, guided instructions for mindfulness exercises) and are often given in the form of home assignments (eg, keep postponing worries to a scheduled time of the day). See Table 1 for a more thorough overview of the contents in LS.

## Data Collection and Measures

Data were collected at baseline (ie, preintervention), and at 1, 2, and 6 months postintervention by means of Web-based surveys. Participants were given 2 weeks to register their responses at each measurement occasion. A reminder email was sent to all nonresponders after 1 week. Log server registrations were also used to collect data on participants in the LS intervention and extracted at the final data collection at 6 months.

Stress was assessed by the stress subscale of the Depression Anxiety and Stress Scale (DASS-S) [60] at every measurement occasion. The DASS-S is a 7-item measure that assesses the severity of the core symptoms of tension (ie, stress) in the past 7 days developed for use with population samples (eg, "I found it difficult to relax"). In the current study, the Cronbach alpha coefficients were .87, .89, .90, and .89 for baseline, 1, 2, and 6 months, respectively. The DASS-S was the primary outcome for the main analyses.

Stress was also assessed in the LS intervention by means of log server registrations as part of the regular intervention. This scale (constructed by the intervention designers by compiling items from several stress measures) consisted of 20 items, such as "I often feel I have too much to do" and "I often set too high personal goals" measured on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). The Cronbach alphas were .92 and .93 in sessions 1 and 13, respectively.

The Mindful Attention Awareness Scale (MAAS) [36] assesses the frequency of being aware of what is occurring in the present moment at each repeated measurement. The MAAS is a 15-item scale that was reduced to 10 items for this study. One item was dropped ("...do jobs or tasks automatically, without being aware...") because it was very similar in Norwegian language to another item that was retained ("...running on automatic, without much awareness..."). Four other items were dropped because the intervention was not developed to tap into these (ie, breaking things, forgetting a person's name, mindless snacking, and excessive goal focus). The Cronbach alphas in this study were .90, .91, .93, and .92 for each measurement occasion from baseline throughout 6 months, respectively.

Procrastination was measured by the procrastination subscale of the Melbourne Decision Making Questionnaire (MDMQ-P) [61] at each measurement occasion. The MDMQ-P is a 5-item measure of the tendency to avoid decision making (eg, "When I have to make a decision, I wait a long time before starting to think about it"). Its Cronbach alphas were .92, .92, .93, and .94 for baseline, 1, 2, and 6 months, respectively, in this study.

## Statistical Methods

An alpha level of .05 was chosen for all tests and all tests were 2-tailed. To check for baseline differences between groups, *t* tests were used for scales and chi-square ($\chi^2$) tests for categorical data. All $\chi^2$ tests that were based on a 2 × 2 contingency table applied the Yates' continuity correction.

Normality was assessed by means of skewness, kurtosis, and inspection of histograms, with plotted normality curves as visual aids, separately for each treatment group. Skewness was ≤1.43 for the LS group and ≤1.04 for the control group. Kurtosis was ≤4.04 and ≤1.77 for the LS and control group, respectively. This indicates moderate skewness and kurtosis; thus, it was decided not to perform any transformations on data in interest of interpretability.

There were no concerns about violation of homogeneity of variance or variance-covariance matrices with $F_{max}$ ratios ≤1.25. Two participants had excessive *z* scores of ±3.29 ($P$<.001, 2-tailed test) on stress at 1 month in the imputed datasets 1 through 5; however, both participants were retained in the dataset as the influence of outliers on mean scores was less than 1.11% after trimming the means by 5%. There were no multivariate outliers as tested by the Mahalanobis distance ($D$) separately for the LS ($D_6$≤19.83) and control group ($D_4$≤13.72) with $P$<.001.

There were 113 (43.6%) participants that participated on all measurement occasions; hence, many participants had missing data. Thus, a 2-group multiple imputation (MI) procedure was applied to construct 5 complete datasets for the main analyses (ie, data were imputed separately for the LS and control group) [62]. Auxiliary demographic variables, such as gender, age, education, and intervention adherence, were included in the imputation model to avoid suppressed correlations. Intervention adherence was included in the imputation model for the LS group only. Otherwise, the imputation model was identical for both groups.

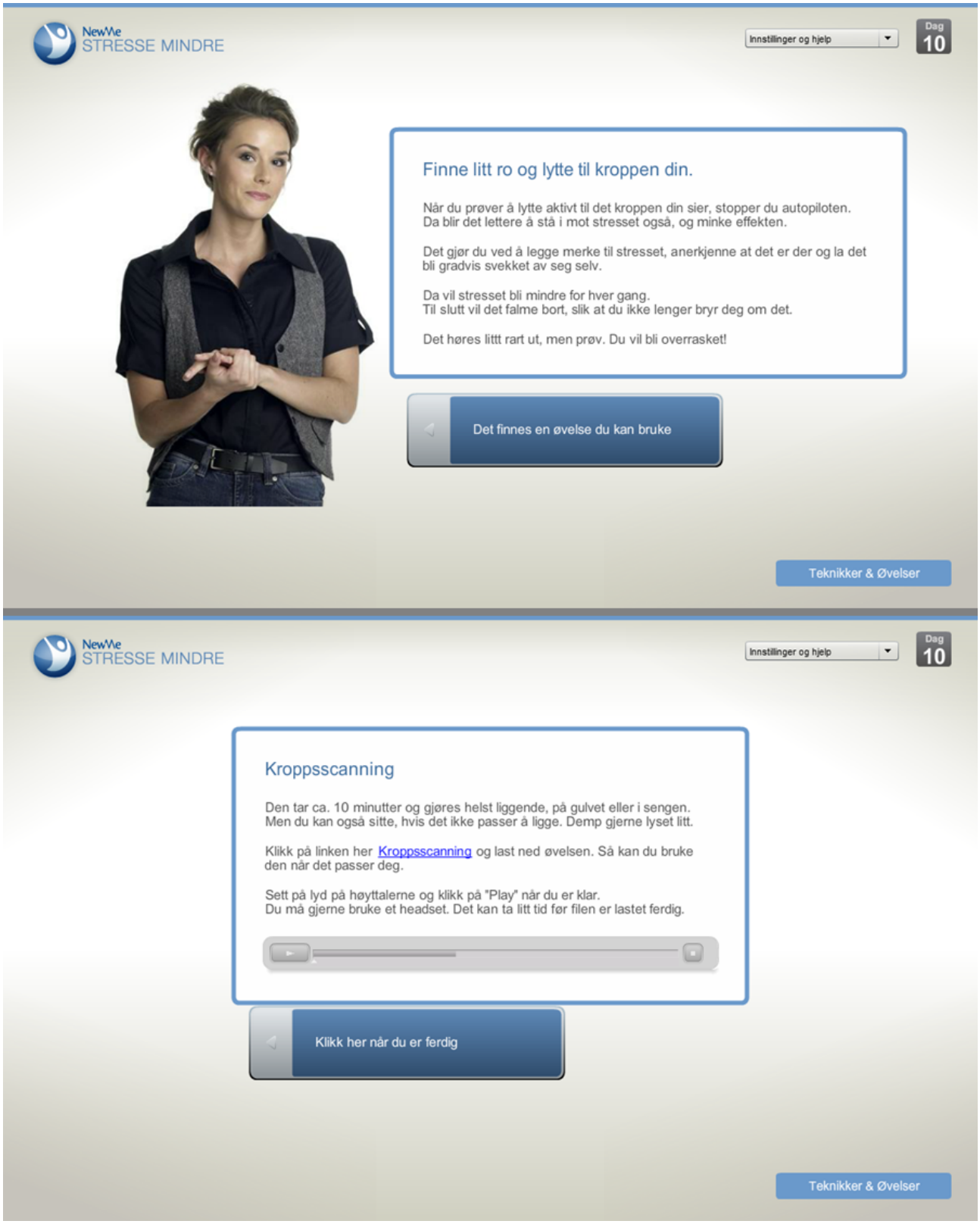
Data were longitudinally nested within 2 hierarchical levels, in which time was nested within participants and defined as level 1, whereas participants were defined as level 2. Level 1 variables included the repeated measures of stress, mindfulness, and procrastination, and were measured at baseline and 1, 2, and 6 months. Level 2 variables included demographics and treatment assignment, and were only measured at baseline. All continuous predictors and covariates were centered on the grand mean before modeling.

A series of multilevel models with maximum likelihood estimation were run to analyze the main treatment effect, multiple mediation, and moderation analyses. The overall fit of the models was evaluated by the Akaike information criterion (AIC) and –2 log likelihood (–2LL) on a smaller-is-better basis. Moreover, comparison of nested models was also evaluated formally by a test of differences in –2LL over the difference in degrees of freedom by using an ordinary $\chi^2$ distribution. A significant difference indicates that the model with the lowest –2LL value fits data better. Analyses were run in SPSS version 20; however, SPSS does not provide pooled model fit indexes in mixed models. Therefore, the median model fit indexes of the 5 imputed datasets are reported. Finally, a pseudo-multivariate coefficient of determination ($R^2$) was calculated to account for the variation between participants in the final main, mediation, and moderation models. Analyses were also conducted separately with available case analysis. Both procedures produced similar results; thus, only data for the imputed sets are reported.

**Table 1.** Overview of program sessions in the Less Stress intervention.

| Session | Description | Psychoeducational content | Techniques and exercises |
|---|---|---|---|
| 1 | Introduction | Information about program, program structure, and a test of stress levels. | |
| 2 | What is stress? | How modern everyday life can affect our stress levels and explanation of stress as a reaction to real or perceived threats to one's physical and mental well-being. | Mindfulness breathing exercise to help become grounded in the present moment. Homework: Practice mindfulness breathing. |
| 3 | What makes you stressed? | Problem-focused vs emotion-focused coping for stress management. | Practical tips on taking care of your economy and worry postponement. |
| 4 | Stress and emotions | Debunking the myth of "positive stress." A bit of stress is not dangerous, but all stress is essentially negative and interlinked with negative emotions. | Mindfulness breathing exercise to help become grounded in the present moment. Homework: practice worry postponement. |
| 5 | Physical activity, stress, and mood | Physical activity has an immediate impact on mood and acts as a buffer against stress. Learning detached mindfulness as a way of relating to negative thoughts and feelings. | Engage in physical activity. Practice detached mindfulness by means of the "Clouds in the sky" exercise. |
| 6 | Causes of stress | The difference of big stressors (eg, wedding, pregnancy, a new job) and everyday stressors or concerns (eg, noisy neighbors, parking ticket, a dreaded phone call). Identifying everyday stressors and finding ways to deal with these. | Strategies for dealing with everyday stressors: (1) avoidance—avoid situations that cause you stress or negative emotions; (2) change—if avoidance fails, adjust the situation such that the cause of stress dissipates; and (3) reattribute—if changing fails, think of the positives that come out of the stressful situation. |
| 7 | Procrastination | Procrastination as a source of stress if delaying all things that are difficult or unpleasant, and how these things can grow into big stressors. | Detached mindfulness: "Tiger task" exercise for relating to difficult or unpleasant situations, thoughts, or feelings. |
| 8 | Why do we procrastinate? | Reasons why we procrastinate (1) lack of commitment, (2) lack of motivation if goals are distant, and (3) anxiousness. Identifying personal reasons for procrastination. | Mindfulness: "Gong gong" exercise to help become grounded in the present moment and relaxing by focusing on an external sound. |
| 9 | Optimism | Optimism leads to less stress, but too much optimism (ie, being overoptimistic) can cause more stress. Beware of situations that require predicting how much time things take and how much things cost; we almost always underestimate. | Passenger train metaphor - alternative to the "Clouds in the sky" exercise. Negative (and positive) thoughts come and go, just stand by and watch your thoughts pass by. |
| 10 | Work-related stress | Explains the "demand, control, and support" model of work-related stress. | Find a balance between demands, control, and support at work. Mindfulness: "Body scan" exercise to increase bodily awareness of how one is doing and become more present in the moment (ie, progressive muscle relaxation). |
| 11 | Time management | Everybody has the same amount of time, but some are better at doing 1 thing at a time. | (1) Make 3 lists of activities you have to, ought to, and can do; (2) prioritize and organize each of your lists; and (3) Start with the "have to" list and fill in with items from the "ought to" list. Still time to spare? Fill in with items from the "can do" list without being overoptimistic. |
| 12 | Goal management | Reassess your personal goals and get rid of unrealistically high goals (eg, weight loss, career ambitions, personal appearance) that cause a constant feeling of guilt or stress. | Mindfulness: "Body scan" exercise to increase bodily awareness of how one is doing and become more present in the moment (ie, progressive muscle relaxation). |
| 13 | Conclusions and summary | Test of your stress level. | |

**Figure 1.** Screenshots from the Less Stress intervention.




## Results

### Subject Characteristics

The flow of participants is depicted in Figure 2. A total of 259 people were eligible for participation and randomized to either the LS intervention or control group. A total of 34 (27.0%) participants in the intervention group discontinued study participation or intervention. Most did not give any reasons for discontinuation, but a few people mentioned reasons such as mail delivery failure (n=2), lack of time (n=2), pregnancy (n=1), too extensive to participate (n=1), and too much to do at work (n=1). Only 3 people provided reasons as to why they discontinued the LS intervention, ie, lack of time (n=2) and too

extensive (n=1). Cumulative losses (ie, loss to follow-up on at least 1 previous follow-up) are shown in curly brackets in Figure 2. Note that participants who discontinued the LS intervention were approached for data collection, although 26 (76.5%) of the 34 intervention dropouts were also lost to follow-up.

There were no significant differences between participants in the LS and control group at baseline (Table 2). However, there were more women (76.0% vs 24.0% males, $P<.01$) and participants with ≤1-3 years of college or university education (59.1% vs 40.9% ≥4-5 years of college or university education, $P<.01$) in the total sample. Most were not acquainted with the recruiter (80.3% vs 19.7% acquainted, $P<.01$) indicating that viral recruitment through Facebook was successful in reaching participants beyond the researcher's first-degree contacts.

**Figure 2.** Flowchart of participants.

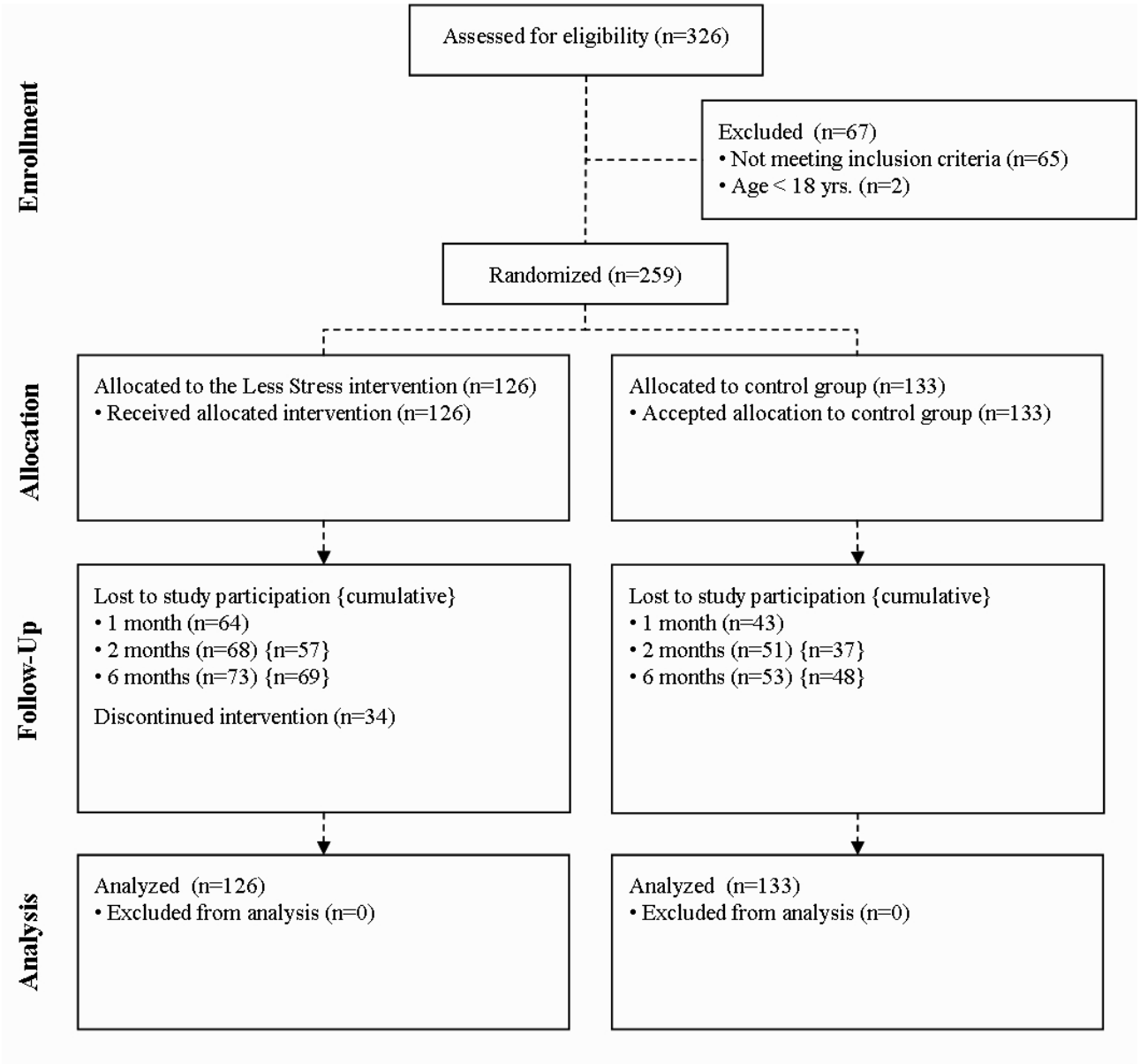


**Table 2.** Baseline sample characteristics by group, Less Stress (LS) intervention and control.

| Characteristic | LS (n=126) | Control (n=133) |
|---|---|---|
| **Gender, n (%)** [a] | | |
| Men | 31 (24.6) | 30 (22.6) |
| Women | 95 (75.4) | 98 (73.7) |
| **Education, n (%)** | | |
| ≤1-3 years in college or university | 77 (61.1) | 76 (57.1) |
| ≥4-5 years in college or university | 49 (38.9) | 57 (42.9) |
| Age (years), mean (SD) | 32.0 (9.6) | 33.2 (9.9) |
| **Acquainted, n (%)** [a] | | |
| No | 97 (78.2) | 107 (80.5) |
| Yes | 27 (21.4) | 24 (18.0) |
| **Psychological variables, mean (SD)** [b] | | |
| Stress | 6.9 (4.7) | 7.5 (4.7) |
| Mindfulness | 34.1 (8.2) | 34.1 (9.0) |
| Procrastination | 18.3 (7.6) | 17.1 (8.1) |

[a]Numbers and percentages adjusted for missing values.

[b]Higher scores indicate higher levels of stress, mindfulness, and more procrastinating behaviors.

## Attrition and Missing Data

The number of respondents to the follow-up surveys in the total sample were 152 (58.7%), 140 (54.1%), and 133 (51.4%) for 1, 2, and 6 months, respectively. The number of respondents within the LS group was 62 (49.2%), 58 (46.0%), and 53 (42.1%) across the 3 follow-up measurements, and 90 (67.7%), 82 (61.7%), and 80 (60.2%) in the control group. Between-group differences in dropout rates at 1-month ($\chi^2_1$=8.4, $P$=.004), 2-month ($\chi^2_1$=5.7, $P$=.017), and 6-month ($\chi^2_1$=7.8, $P$=.005) follow-up, were significant. Hence, selective attrition is a potential problem regarding the interpretation of levels of stress over the study period. Further, it turned out that the proportion of missing data in the total sample ranged from 0% to 3.5% at baseline and from 42.9% to 52.1% at follow-up (ie, missing data due to item or wave nonresponse). Despite this, Little's overall test of randomness indicated that the distribution of missing data was not predictable for the LS group ($\chi^2_{145}$=172.2, $P$=.06) and that data for the control group also could be classified as missing completely at random ($\chi^2_{154}$=158.3, $P$=.39).

Due to these potential problems, the effects of selective attrition on means, variances, and relationships among variables used in subsequent longitudinal analyses of treatment effects were assessed, following the procedure described by Goodman and Blum [63]. Because there was little attrition in the study from 1 to 6 months, selective attrition was assessed on the basis of study dropout from baseline to 1 month. The results showed no mean differences between study dropouts and stayers at baseline (-0.59 ≤ $t$ ≤0.63, all $P$ values ≥.53). There were also no significant differences in variances using the normal approximation to chi-square [64] in the total sample to those who stayed (-1.91 < $z$ <0.36, all $P$ values >.06). In other words, selective attrition did not affect the means or variances. However, testing the relationship among variables with multiple regression analyses on stress in the total sample and stayers separately found gender to be a significant predictor of stress at baseline for the total sample (unstandardized beta coefficient [B]=1.40, $P$=.01), but not among stayers (B=0.63, $P$=.44). An independent samples $t$ test revealed that males (mean 5.4, SD 3.7) had lower stress scores than females at baseline (mean 7.7, SD 4.8; $t_{130}$=–3.96, $P$<.001). Despite that substantial study dropout led to selective attrition, only the relationship between gender and stress was affected. Thus, we can be confident that, other than gender, participant attrition will not affect the results in this study.

## Intervention Use, Acceptance, and Effect

A total of 126 subjects were registered for the LS intervention, of which 92 (73.0%) engaged with LS (ie, initiated use) and 47 (38.5%) completed all 13 sessions. On average, participants completed 6.82 (SD 5.70) sessions and spent 1 hour and 6 minutes (SD 46) on LS. Time spent on LS was below estimated time needed for optimal adherence per session (13 sessions × 10 minutes per session=2 hours and 10 minutes) as tested by a 1-sample $t$ test ($t_{91}$=–13.27, $P$<.001).

Of the 49 (38.9%) participants that reported data on intervention acceptance at 1 month, LS seemed well received among most. Of these, 41 (83.7%) reported that they believed LS to be "useful to me" whereas 8 (16.3%) participants either disagreed or were indifferent with LS being "useful to me." Thirty-five (71.4%) participants would recommend LS to others, 10 (20.4%) would neither recommend LS to others or not, whereas 4 (8.2%) users

reported that they would not recommend LS to others. Moreover, 46 (89.7%) participants agreed that LS was "easy to use."

The first hypothesis was that participants in the LS intervention would have lower stress scores at session 13 compared to session 1, as measured by log server registrations. A paired-samples *t* test showed that participants in the LS group had significantly reduced their stress level from session 1 (mean 65.74, SD 13.71) to session 13 (mean 51.91, SD 13.12; $t_{46}$=7.54, *P*<.001) as measured by the test of stress levels in the intervention. This equals a large effect size (Cohen's *d*=1.10).

## Main Effects Analysis

Tables 3 and 4 present the uncentered correlations among study variables from level 1 (ie, repeated measures) and level 2 (ie, participants), respectively. Separate correlation tables for level 1 and level 2 variables were included to avoid aggregation or disaggregation of data. As can be seen in Table 3, all measures of stress, mindfulness, and procrastination correlated significantly at each measurement occasions (all *P* values ≤.002). Stress correlated negatively with mindfulness and positively with procrastination, and mindfulness correlated negatively with procrastination, as expected.

The main hypothesis concerned the comparison of trajectories in stress levels in the LS group and the control group. It was expected that participants in the LS group would reduce levels of stress over a period of 6 months compared to the control group, which would remain at approximately the same stress level throughout. The main effects from the multilevel regression analysis of stress levels are presented in Tables 5 and 6. Model 1 with the repeated measures only indicates that average levels of stress vary significantly across participants and over time. The intraclass correlation coefficient (ICC) was 0.34. This means that 34% of the variation in stress levels is attributable to interindividual differences. In other words, stress varies (naturally) over time for most participants; however, substantial proportions of the variation in stress levels can be attributed to differences between participants over time.

**Table 3.** Correlations among level 1 variables.

| Measure | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Stress baseline | — | | | | | | | | | | | |
| 2. Stress 1 month | .32 | — | | | | | | | | | | |
| 3. Stress 2 months | .35 | .35 | — | | | | | | | | | |
| 4. Stress 6 months | .40 | .39 | .37 | — | | | | | | | | |
| 5. Mindfulness baseline | –.59 | –.27 | –.23 | –.26 | — | | | | | | | |
| 6. Mindfulness 1 month | –.28 | –.53 | –.31 | –.35 | .41 | — | | | | | | |
| 7. Mindfulness 2 months | –.32 | –.41 | –.52 | –.34 | .42 | .60 | — | | | | | |
| 8. Mindfulness 6 months | –.27 | –.27 | –.26 | –.43 | .35 | .48 | .53 | — | | | | |
| 9. Procrastination baseline | .47 | .21 | .23 | .22 | –.56 | –.26 | –.31 | –.32 | — | | | |
| 10. Procrastination 1 month | .26 | .35 | .20 | .25 | –.28 | –.47 | –.36 | –.45 | .49 | — | | |
| 11. Procrastination 2 months | .29 | .35 | .36 | .33 | –.33 | –.44 | –.51 | –.46 | .48 | .62 | — | |
| 12. Procrastination 6 months | .24 | .35 | .24 | .39 | –.25 | –.43 | –.39 | –.53 | .40 | .57 | .61 | — |

**Table 4.** Correlations among level 2 variables.

| Measure | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1. Treatment[a] | — | | | |
| 2. Gender[b] | –.01 | — | | |
| 3. Age | -06[c] | –.04 | — | |
| 4. Education[d] | –.04 | .11[c] | .07[c] | — |

[a]Treatment: 0 = controls, 1 = LS group.

[b]Gender: 0 = male, 1 = female.

[c]*P*<.05.

[d]Education: 0 = ≤1-3 years of college/university, 1 = ≥4-5 years of college/university.

**Table 5.** Results of functions of time on stress levels over six months.

| Model | 1 | | 2 | | 3 | | 4 | |
|---|---|---|---|---|---|---|---|---|
| Treatment effects | Est | SE | Est | SE | Est | SE | Est | SE |
| **Fixed effects** | | | | | | | | |
| Intercept | 6.31[d] | 0.19 | 6.70[d] | 0.22 | 6.98[d] | 0.26 | 7.20[d] | 0.26 |
| Linear[a] | | | –0.18[d] | 0.05 | –0.60[d] | 0.22 | –2.49[d] | 0.66 |
| Quadratic[b] | | | | | 0.07[d] | 0.03 | 1.37[d] | 0.43 |
| Cubic[c] | | | | | | | –0.17[d] | 0.05 |
| **Random effects** | | | | | | | | |
| $\sigma^2$ within participants | 10.89[d] | 0.60 | 10.75[d] | 0.61 | 10.66[d] | 0.61 | 10.44[d] | 0.58 |
| **$\sigma^2$ between participants** | 5.68[d] | 0.79 | | | | | | |
| Intercept | | | 6.46[d] | 1.08 | 6.92[d] | 1.09 | 7.05[d] | 1.08 |
| Intercept, slope | | | –0.21 | 0.16 | –0.28 | 0.15 | –0.31[d] | 0.14 |
| Slope | | | 0.00 | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 |
| **Model fit indexes** | | | | | | | | |
| AIC | 5708.19 | | 5693.92 | | 5692.41 | | 5686.56 | |
| –2LL ($\chi^2$) | 5702.19 | | 5681.92 | | 5678.41 | | 5670.56 | |
| $R^2$ (level 1) | | | 0.01 | | 0.02 | | 0.04 | |

[a]Linear: 0 = baseline, 1 = 1 month, 2 = 2 months, 6 = 6 months.

[b]Quadratic: 0 = baseline, 1 = 1 month, 4 = 2 months, 36 = 6 months.

[c]Cubic: 0 = baseline, 1 = 1 month, 8 = 2 months, 216 = 6 months.

[d]$P<.05$.

A series of multilevel models with a linear, quadratic, and cubic growth parameter were estimated separately to distinguish the natural or normative development of stress over measurement occasions from the treatment effect (ie, models 2-4 in Table 5). The negative estimate of linear growth in model 2 does indicate that, on average, participants experienced reductions in stress levels over time. A test of differences in model fit between model 2 and model 1 yielded a significant result ($\Delta\chi^2_3$=20.3, $P$<.001). A quadratic function of time was added in model 3 to capture any acceleration or deceleration in the rate of change that might occur over the repeated measurements. Results show that participants' reductions in stress levels tended to accelerate slightly over time. Model 3 showed an improvement in model fit from model 2 at the 10% level ($\Delta\chi^2_1$=3.5, $P$=.06). The variation within groups over time decreased, the AIC value decreased, and the explained variance increased slightly (ie, $R^2$) for model 3. Thus, it was decided to retain the quadratic function in further analyses. By including a cubic term in model 4, the rate of change in stress levels decelerated again and provided significant improvements in model fit ($\Delta\chi^2_1$=7.85, $P$=.005). Consequently, treatment effects had to be modeled with a linear, quadratic, and cubic growth parameter.

Model 5 with treatment as level 2 predictor shows that, after controlling for the natural development of stress over time, participants in the LS group had significantly lower stress levels as compared to the control group (see Table 6). A formal test of differences comparing models 5 and 4, demonstrated that model 5 fits data better than model 4 ($\Delta\chi^2_1$=9.4, $P$=.002). However, a test of differences in slopes or developmental trajectories between groups was first examined in model 6 that included cross-level interactions between treatment and normative growth over time. The results show that there was a significant interaction effect between treatment and the linear, quadratic, and cubic growth parameters which suggests that, on average, participants experienced reductions in their stress levels, but that participants in the LS group experienced a different developmental trajectory over and above the natural variation in stress over time (model 6 vs model 5: $\Delta\chi^2_3$=8.0, $P$=.046).

**Table 6.** Results of treatment effects on stress levels over six months.

| Model | 5 | | 6 | |
|---|---|---|---|---|
| Treatment effects | Est | SE | Est | SE |
| **Fixed effects** | | | | |
| Intercept | 7.74[e] | 0.31 | 7.40[e] | 0.37 |
| Linear[a] | –2.49[e] | 0.66 | –0.81 | 0.85 |
| Quadratic[b] | 1.37[e] | 0.43 | 0.29 | 0.56 |
| Cubic[c] | –0.17[e] | 0.05 | –0.03 | 0.07 |
| **Predictors** | | | | |
| Treatment[d] | –1.10[e] | 0.35 | –0.40 | 0.52 |
| Treatment × linear | | | –3.45[e] | 1.28 |
| Treatment × quadratic | | | 2.23[e] | 0.84 |
| Treatment × cubic | | | –0.28[e] | 0.11 |
| **Random effects** | | | | |
| $\sigma^2$ within participants | 10.48[e] | 0.59 | 10.30[e] | 0.57 |
| **$\sigma^2$ between participants** | | | | |
| Intercept | 6.83[e] | 1.07 | 6.90[e] | 1.06 |
| Intercept, slope | –0.32[e] | 0.15 | –0.34 | 0.14 |
| Slope | 0.02 | 0.00 | 0.01 | 0.00 |
| **Model fit indexes** | | | | |
| AIC | 5679.12 | | 5677.14 | |
| –2LL ($\chi^2$) | 5661.12 | | 5653.14 | |
| $R^2$ (level 1) | 0.04 | | 0.05 | |

[a]Linear: 0 = baseline, 1 = 1 month, 2 = 2 months, 6 = 6 months.

[b]Quadratic: 0 = baseline, 1 = 1 month, 4 = 2 months, 36 = 6 months.

[c]Cubic: 0 = baseline, 1 = 1 month, 8 = 2 months, 216 = 6 months.

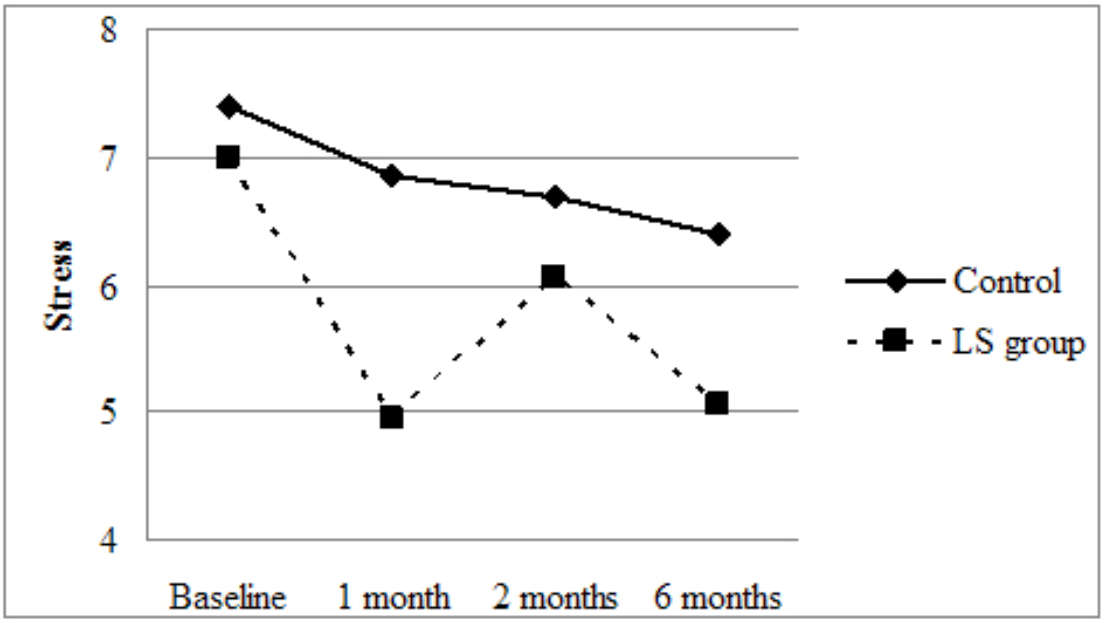
[d]Treatment: 0 = control group, 1 = LS group.

[e]$P<.05$.

To demonstrate the effect of treatment on initial status and the rate of change over time, Singer and Willett [65] proposed creating prototypical plots. Prototypical plots can be obtained by substituting the values of the treatment variable and time (ie, linear, quadratic, and cubic) variables in the final estimated model (ie, model 6): $Y_{it}$=7.40 + (–0.81)(Time) + (0.29)(Time$^2$) + (–0.03)(Time$^3$) + (–0.40)(Treatment) + (–3.45)(Time × Treatment) + (2.23)(Time$^2$ × Treatment) + (–0.28)(Time$^3$ × Treatment), where $Y_{it}$ is the repeated measure of stress for person *i* at time *t*. The trajectories in Figure 3 demonstrate that the LS group experienced a more immediate and rapid reduction in stress levels as compared to the control group. Although stress levels increased from 1 to 2 months, stress levels returned to the 1-month level by 6 months. In comparison, the control group followed a more modest, steady, and nonsignificant linear decline in stress levels (ie, normative development).

Finally, 3 types of covariance structures relevant and common in studies with repeated measurements were tested. The purpose was to examine how errors were distributed and whether the properties imposed on the covariance structure fit well to the data. The unstructured covariance structure fit the data best as determined by having the lowest information criterion (AIC = 5663.27; compound symmetry AIC = 5678.61 and first-order autoregressive AIC = 5735.48). The assumption of the unstructured error covariance is that all parameters are estimated and allowed to vary freely. In other words, it is the least restrictive covariance structure. Estimates remained largely unchanged with an unstructured covariance structure and are thus not reported.

**Figure 3.** Fitted developmental trajectories for the control and LS group, respectively.




## Multiple Mediation Analysis

The third hypothesis was that changes in stress levels over time attributed to the LS intervention could be accounted for by changes in mindfulness and procrastination over time. The model-building approach suggested by Bliese and Ployhart [66] was used and a product-of-coefficients strategy was employed to test for multiple mediation effects of treatment on stress via mindfulness and procrastination [67-69]. In other words, the unstandardized beta coefficients for each mediator were multiplied to represent the different indirect effects. Furthermore, the Monte Carlo Method for Assessing Mediation (MCMAM), as described by MacKinnon et al [70], was applied to generate 95% confidence intervals for indirect effects with 20,000 bootstraps [71]. The MCMAM performs reasonably well and can be implemented on the pooled estimates of multiply imputed datasets.

The total effect of treatment on stress was demonstrated previously in model 6 in Table 6 and found to be significant. So, the first step in the mediation analysis was to estimate an unconditional model (ie, model 7a) to test for variation between participants in mindfulness as a mediator over time (see Table 7). The ICC was 0.44, which indicated that average levels of mindfulness vary significantly across participants and over time; hence, a multilevel mediation analysis would be adequate. The next step was to determine the effect of treatment on mindfulness and the fixed functions for time, controlling for procrastination (ie, the second mediator), in a sequence of steps similar to models 2-5 in Tables 5 and 6, although only the final model is reported. Model 7b in Table 7 shows that there was a significant interaction effect between treatment and the linear, quadratic, and cubic growth parameters. This suggests that, on average, participants in the LS group experienced a different developmental trajectory over and above the natural variation in mindfulness over time, after controlling for procrastination.

The effect of treatment on procrastination was tested similarly. The ICC in the unconditional model 8a was 0.51 and suggested that average levels of procrastination vary significantly across participants and over time. Then the effect of treatment on procrastination and the fixed functions for time, controlling for mindfulness was determined. The final model providing the best-fit indexes suggested that treatment had a quadratic effect on procrastination. On average, participants in the LS group did experience a significant decrease in procrastination, although levels of procrastination seemed to increase slightly over time.

Finally, model 9 examined whether the direct effect of treatment on stress was reduced. The results show that mindfulness and procrastination as a set do mediate the effect of treatment on stress levels (see Figure 4). A comparison of the final 2-mediator model (ie, model 9) indicates that this model fit data better than model 6 of the treatment effect ($\Delta\chi^2_2$=279.1, $P$<.001). However, in contrast to model 6 of the treatment effect, an examination of the distribution of errors showed that the best fitting covariance structure was that of compound symmetry (AIC = 5399.26; unstructured AIC = 5403.44 and first-order autoregressive AIC = 5434.60). The assumption of compound symmetry suggests that variances and covariances across the repeated measures were equal.

In simplified terms, the directions of the paths in Table 7 can be interpreted such that the LS intervention leads to greater mindfulness and, in turn, leads to lower stress over time. It also appears that the LS intervention leads to less procrastination, which appears to lead to slightly higher stress over time. An examination of the specific indirect effects in Table 8 indicates that all the specific indirect effects of mindfulness are significant; however, it appears that only the linear specific indirect effect of procrastination is significant. The 95% confidence interval for the quadratic specific indirect effect of procrastination ranged from 0.00 to 0.04 and does not seem to contribute to the indirect effect.

### Multiple Moderation Analysis

The final hypothesis was concerned with examining any moderating effects of gender, age, and education on the treatment effect of the LS intervention on stress scores over time. A reference model of the normative growth and treatment effect was modeled previously (see model 6 in Table 6). Consequently, the first step in the multiple moderation analysis was to model the main effects of gender, age, and education conditional on the normative growth and treatment effects (see model 9 in Table 9). The results show that gender and age had a significant contribution above and beyond normative development and treatment effect on stress. Female participants had higher stress levels than males, whereas stress levels decreased slightly with age. There were, however, no main effects of education. A comparison of model 9 with the addition of the moderators to model 6 of the treatment effects (see Table 6), demonstrated that model 9 performed better ($\Delta\chi^2_3$=17.7, $P$<.001).

The next step was to model the interaction effects between treatment and gender, age, and education, respectively (see model 10 in Table 9). All 3 interaction terms were added simultaneously to adjust for multiple statistical tests and to estimate conditional interaction effects. Results show no interaction effects between treatment and the moderators and no overall improvement in model fit over model 9 ($\Delta\chi^2_3$=6.4, $P$=.09). Finally, a model with interactions between linear growth and all other variables were included to determine whether development was consistent across levels of the other variables (ie, determine time-specific interaction effects; see model 11). Three-way interactions were also included between linear development and other 2-way interactions to test for parallel slopes. Again, there was a significant effect of gender and age on stress over and above the treatment effect, and education also had a significant main effect on stress this time (ie, higher education was related to lower stress). However, although model 11 had improved model fit indexes compared to model 10 ($\Delta\chi^2_6$=24.7, $P$<.001), none of the interactions contributed substantially to the model.

In conclusion, model 9 appears to be the most parsimonious model in which the unstructured covariance structure fit the data best (AIC = 5609.71; compound symmetry AIC = 5678.61 and first-order autoregressive AIC = 5735.48). Estimates remained largely unchanged with an unstructured covariance structure and are not reported.

**Figure 4.** The estimated dynamic multiple mediation model with unstandardized beta coefficients.

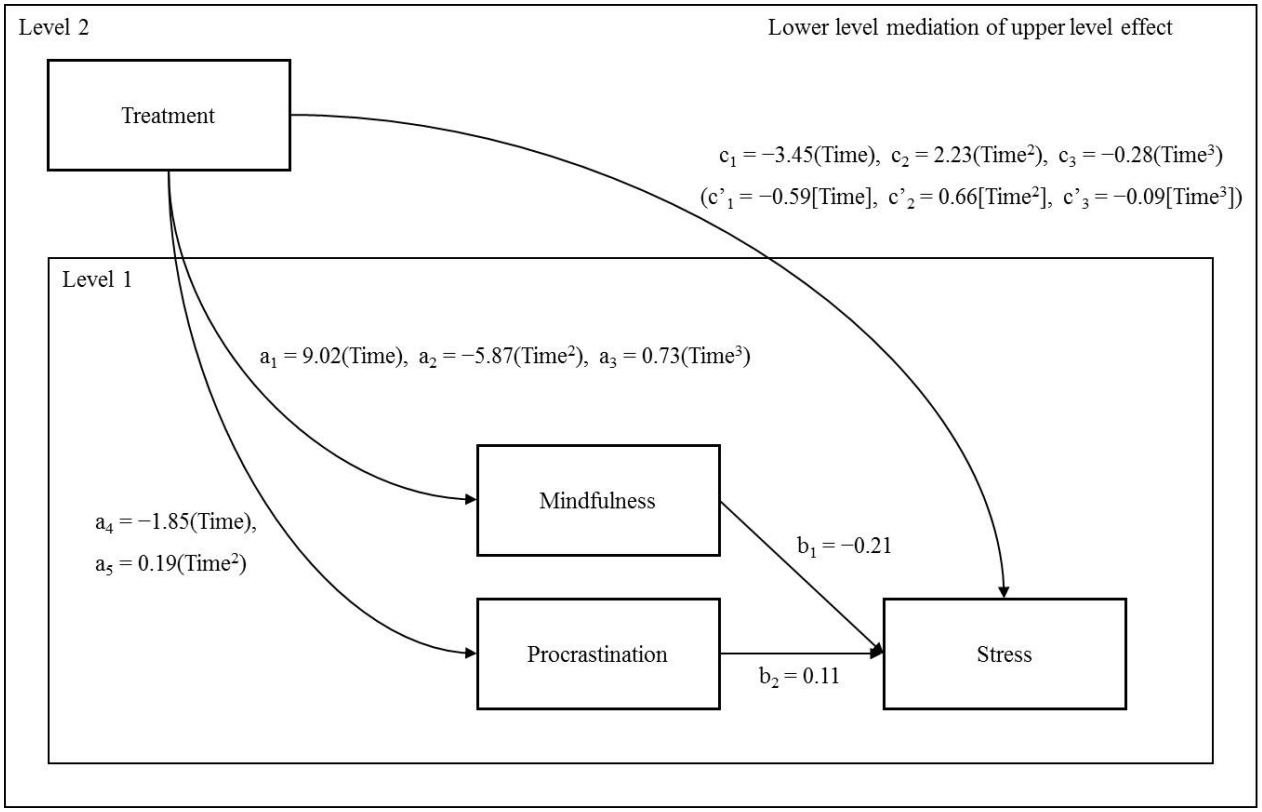


**Table 7.** The effects of treatment and mindfulness and procrastination on stress levels over time.

| Model (dependent variable) | 7a (mindfulness) | | 7b (mindfulness) | | 8a (procrastination) | | 8b (procrastination) | | Model 9 (stress) | |
|---|---|---|---|---|---|---|---|---|---|---|
| Treatment effects | Est | SE | Est | SE | Est | SE | Est | SE | Est | SE |
| **Fixed effects** | | | | | | | | | | |
| Intercept | 36.04[e] | 0.37 | 34.90[e] | 0.60 | 16.81[e] | 0.34 | 16.15[e] | 0.53 | 7.20[e] | 0.30 |
| Linear[a] | | | –2.57 | 1.41 | | | 0.42 | 1.01 | –1.51 | 0.78 |
| Quadratic[b] | | | 1.97[e] | 0.93 | | | 0.16 | 0.63 | 0.70 | 0.52 |
| Cubic[c] | | | –0.24[e] | 0.12 | | | –0.03 | 0.08 | –0.08 | 0.07 |
| Treatment[d] | | | –0.10 | 0.85 | | | 1.48[e] | 0.74 | –0.79 | 0.43 |
| Treatment × linear | | | 9.02[e] | 1.95 | | | –1.85[e] | 0.54 | –0.59 | 1.25 |
| Treatment × quadratic | | | –5.87[e] | 1.28 | | | 0.19[e] | 0.08 | 0.66 | 0.84 |
| Treatment × cubic | | | 0.73[e] | 0.16 | | | | | –0.09 | 0.11 |
| Mindfulness | | | | | | | –0.35[e] | 0.02 | –0.21[e] | 0.02 |
| Procrastination | | | –0.53[e] | 0.04 | | | | | 0.11[e] | 0.02 |
| **Random effects** | | | | | | | | | | |
| $\sigma^2$ within participants | 33.77[e] | 1.81 | 25.39[e] | 1.74 | 21.09[e] | 1.10 | 16.66[e] | 1.22 | 8.57[e] | 0.47 |
| **$\sigma^2$ between participants** | 26.44[e] | 3.18 | | | 21.79[e] | 2.46 | | | | |
| Intercept | | | 20.71[e] | 3.26 | | | 18.03[e] | 2.54 | 3.00[f] | 0.67 |
| Intercept, slope | | | –1.54[e] | 0.57 | | | –1.20[e] | 0.41 | –0.08 | 0.10 |
| Slope | | | 0.30 | 0.17 | | | 0.16 | 0.12 | 0.00 | 0.00 |
| **Model fit indexes** | | | | | | | | | | |
| AIC | 6959.34 | | 6687.13 | | 6528.11 | | 6284.20 | | 5402.07 | |
| –2LL ($\chi^2$) | 6953.34 | | 6661.13 | | 6522.11 | | 6260.20 | | 5374.07 | |
| $R^2$ (level 1) | | | 0.25 | | | | 0.21 | | 0.21 | |

[a]Linear: 0 = baseline, 1 = 1 month, 2 = 2 months, 6 = 6 months.

[b]Quadratic: 0 = baseline, 1 = 1 month, 4 = 2 months, 36 = 6 months.

[c]Cubic: 0 = baseline, 1 = 1 month, 8 = 2 months, 216 = 6 months.

[d]Treatment: 0 = control group, 1 = LS group.

[e]$P<.05$.

**Table 8.** Specific indirect effects of mindfulness and procrastination on the effect of treatment on stress over time.

| Specific indirect effects | | Est | 95% CI | |
|---|---|---|---|---|
| | | | Low | High |
| **Mindfulness** | | | | |
| | Linear | –1.89 | –2.83 | –1.06 |
| | Quadratic | 1.23 | 0.68 | 1.84 |
| | Cubic | –0.15 | –0.23 | –0.08 |
| **Procrastination** | | | | |
| | Linear | –0.20 | –0.35 | –0.08 |
| | Quadratic | 0.02 | 0.00 | 0.04 |

**Table 9.** Results of multilevel analysis: multiple moderation.

| Model | 9 | | 10 | | 11 | |
|---|---|---|---|---|---|---|
| Treatment effects | Est | SE | Est | SE | Est | SE |
| **Fixed effects** | | | | | | |
| Intercept | 6.80[g] | 0.49 | 6.59[g] | 0.58 | 6.65[g] | 0.70 |
| Linear[a] | –0.81 | 0.85 | –0.81 | 0.85 | –0.83 | 0.85 |
| Quadratic[b] | 0.29 | 0.55 | 0.29 | 0.55 | 0.29 | 0.55 |
| Cubic[c] | –0.03 | 0.07 | –0.03 | 0.07 | –0.03 | 0.07 |
| **Predictors** | | | | | | |
| Treatment[d] | –0.45 | 0.51 | –0.09 | 0.83 | –0.58 | 0.95 |
| Treatment×linear | –3.45[g] | 1.28 | –3.45[g] | 1.28 | –3.27[g] | 1.28 |
| Treatment×quadratic | 2.23[g] | 0.84 | 2.23[g] | 0.84 | 2.23[g] | 0.83 |
| Treatment×cubic | –0.28[g] | 0.11 | –0.28[g] | 0.11 | –0.28[g] | 0.11 |
| Gender[e] | 1.07[g] | 0.42 | 1.37[g] | 0.57 | 1.81[g] | 0.71 |
| Age | –0.04[g] | 0.02 | –0.08[g] | 0.03 | –0.08[g] | 0.03 |
| Education[f] | –0.47 | 0.35 | –0.47 | 0.49 | –1.39[g] | 0.60 |
| Treatment×gender | | | –0.48 | 0.81 | –0.23 | 1.01 |
| Treatment×age | | | 0.07 | 0.04 | 0.05 | 0.04 |
| Treatment×education | | | 0.02 | 0.72 | 0.68 | 0.87 |
| Linear×gender | | | | | –0.16 | 0.15 |
| Linear×age | | | | | 0.00 | 0.01 |
| Linear×education | | | | | 0.34 | 0.14 |
| Linear×treatment×gender | | | | | –0.09 | 0.22 |
| Linear×treatment×age | | | | | 0.01 | 0.01 |
| Linear×treatment×education | | | | | –0.24 | 0.21 |
| **Random effects** | | | | | | |
| $\sigma^2$ within participants | 10.29[g] | 0.57 | 10.28[g] | 0.57 | 10.10[g] | 0.59 |
| **$\sigma^2$ between participants** | | | | | | |
| Intercept | 6.02[g] | 1.00 | 5.93[g] | 0.98 | 5.81[g] | 1.00 |
| Intercept, slope | –0.25 | 0.14 | –0.27 | 0.14 | –0.19 | 0.16 |
| Slope | 0.01 | 0.00 | 0.01 | 0.00 | 0.01 | 0.01 |
| **Model fit indexes** | | | | | | |
| AIC | 5665.48 | | 5665.09 | | 5676.43 | |
| –2LL ($\chi^2$) | 5635.48 | | 5629.09 | | 5604.43 | |
| $R^2$ (level 1) | 0.06 | | 0.06 | | 0.07 | |

[a]Linear: 0 = baseline, 1 = 1 month, 2 = 2 months, 6 = 6 months.

[b]Quadratic: 0 = baseline, 1 = 1 month, 4 = 2 months, 36 = 6 months.

[c]Cubic: 0 = baseline, 1 = 1 month, 8 = 2 months, 216 = 6 months.

[d]Treatment: 0 = control group, 1 = LS group.

[e]Gender: 0 = male, 1 = female.

[f]Education: 0 = ≤1-3 years of college/university, 1 = ≥4-5 years of college/university.

[g]$P<.05$.

## Discussion

### Principal Findings

Despite the fact that stress is experienced by many people and that stress is a contributing factor to many mental and physical health conditions, few efforts are made to develop and test the effects of interventions for stress. Findings from this study suggest that Web-based interventions can potentially reduce levels of stress. First of all, analysis of log server registrations found large reductions of the LS intervention on levels of stress among intervention completers. Second, treatment was a significant predictor of linear, quadratic, and cubic changes in stress, but not associated with initial status (see model 6 in Table 6). For the linear slope of stress, participants in the LS group showed a faster recovery from stress, although they also had a faster rate of change in stress (ie, increase, quadratic growth) when compared to the control group. Lastly, the LS group had a slower rate of cubic change in stress levels (ie, decrease) than the control group. In other words, despite variations in stress levels, long-term (ie, 6 months) stress levels returned to the level of the immediate short-term effect at 1 month in the LS group. This implies that participants learned ways of managing their stress levels during the course of the intervention that they carried on with them and used to lower their stress levels over time.

There are a limited number of Web-based stress interventions although there is great variability in terms of intervention content and the methods used to evaluate these [18-32]. However, to the extent that there are similarities between some of the studies, it seems that interventions outside of workplace settings (eg, general or family setting) have shown more unequivocally positive findings. It also appears that the single-target interventions are more likely to have an effect on stress [29] than multitarget interventions (eg, dietary behaviors and stress) [23,24]. In some studies, it is not unreasonable to assume that, for example, a small sample may have affected the results [22] or that a high attrition rate was not sufficiently addressed [23]. In contrast to studies which have shown no or only short-term effects, this study has a reasonably high number of participants and data points, addressed attrition and missing data, and examined a single-target intervention in a general setting whose only aim is to reduce levels of stress.

Treatment accounted for approximately 5% of the variation in stress levels across time within participants (ie, 5% of the overall variability in stress is explained by the LS intervention). However, relatively modest treatment effects need not be a problem for eHealth interventions. The distribution of many psychological treatments is concentrated on a large effect for relatively few patients. However, eHealth technologies have the potential to shift this balance. Online consumer behavior suggests that by creating a longer tail in the distribution of eHealth interventions (ie, reaching more users), the market has the potential to substantially increase the collective effect of eHealth technologies [72]. As such, even small and modest changes can be meaningful at the population level. It should, however, be noted that eHealth technologies have yet to reach a large number of users, in particular, the computer illiterate, those with lower incomes, and those without access to the Internet. Even in Norway where the access and use of the Internet is very high in the population, there are digital divides [73]. For example, almost everyone with incomes above NOK 600,000 have Internet access at home, whereas 18% of those with incomes below NOK 200,000 are without Internet access.

This study has not only documented the effect of a Web-based intervention for stress reduction, but also identified its mechanisms of change. As expected and in-line with previous research [35], it turned out that mindfulness mediated the effect of the LS intervention. This is the first study that has examined the relationship between temporal changes in mindfulness and outcomes in an intervention setting. Overall, the results show that mindfulness can be successfully enhanced in Web-based interventions, but that momentary variations in mindfulness can be expected. The LS intervention also led to less procrastination that, in turn, reduced levels of stress as expected based on previous research [41]. More specifically, the results indicate that the LS intervention successfully managed to interrupt the U-shaped (ie, quadratic) pattern of procrastination that can be expected to occur naturally over time [45]. This means the LS intervention led to reduced procrastination that was maintained over time and participants did not, on average, experience the expected increases in procrastination.

Since there often are differences in stress by gender, age, and education, an important finding in this study is that the LS intervention seems to work equally well regardless of these demographic characteristics. In general, female participants reported higher levels of stress and participants that were older reported somewhat lower stress levels, but no demographic characteristics moderated the effect of the LS intervention. This does not mean that there are no psychological moderators of the effects of the LS intervention, for instance, but it may be that the LS intervention can provide a cost-efficient one-size-fits-all approach in terms of demographic characteristics. However, these findings (or lack thereof) should be interpreted with some caution, at least in regards to the result of the analyses of gender.

### Limitations

This study has several limitations. First, the sample in this study was based on viral recruitment on a social networking site (ie, Facebook). Earlier reviews have shown that Internet-based recruitment procedures have faced challenges in recruiting diverse samples [74]. This may be a part of the reason 3 out of 4 participants in this study were women, albeit 80% of those recruited were not acquainted with the female recruiter. This may indicate other explanations of why more women were recruited, such as that more women generally participate in research or that more women are attracted to Web-based self-help interventions [75]. In fact, a recent study did show success in recruiting a diverse sample using Facebook for a randomized controlled trial [76], which further supports the argument for alternative explanations for the gender bias in the recruitment procedure rather than viral online recruitment per se.

There were no reports of negative side effects of using Facebook for participant recruitment in this study; however, the use of

social networking sites is an area in need of research and guidelines. Although most were not acquainted with the recruiter, they were acquainted with the person who told them about the study. Thus, a recommendation or study invitation from a friend would have more impact than from a researcher. This also raises ethical issues concerning confidentiality and security in research with peer-to-peer recruitment, but also because websites, such as Facebook, frequently change or update their privacy policies, many of which have been highly controversial. Therefore, it is of utmost importance to carefully consider the recruitment and communication strategies employed via social media, especially for sensitive topics (eg, sexually transmitted diseases), and ensure that participants are redirected to an external website so that the amount of information exchanged on Facebook or similar sites is minimized as in this study or the study by Fenner and colleagues [77] by using advertisements.

The second limitation has to do with selective attrition and missing data. In the LS group, more participants dropped out during follow-up than in the control group. However, the only substantial explanation for study attrition was that more males dropped out most likely because they, in general, had lower stress scores than females. The moderation analyses further confirmed this assumption that inadvertently may have had implications for the power to detect potential interaction effects which is considerably reduced with categorical variables whose categories differ in sample size [78]. However, other than gender, there were no indications that selective attrition or missing data affected the means, variances, or the relationships among variables between those who remained in the study and those who dropped out. Hence, we can be confident about the validity of the results in this study.

The third limitation of this study concerns the mediation analysis. It is becoming more common to investigate complex models in intervention research by using multilevel mediation models, testing for multiple mediators or testing for nonlinear mediation effects [79,80]. In many cases, researchers will assume that there is more than 1 mediator that can potentially affect the outcome of an intervention. Most often, researchers examine mediation with only 1 mediator at a time. Consequently, the effects of multiple mediators cannot be simply examined or compared against each other if researchers examine mediators singly. However, several complications arise when testing for multiple mediators in multilevel models and, unfortunately, there is currently a lack of established procedures or methods for testing indirect effects in multilevel models with multiple mediators where the constituent paths are nonlinear. So, although we may have used the best available methods to date, such as bootstrapping, it is obvious that there is a need to develop a set of recommendations or procedures in this area.

### Conclusion and Future Research

The results from this randomized controlled trial suggest that a Web-based intervention can reduce levels of stress over time and that both mindfulness and procrastination could be important components for inclusion in future eHealth interventions for stress. Future research should make sure to examine the effects of the LS or similar interventions for stress reduction among more male participants and investigate the role of psychological moderators of treatment effects.

### Acknowledgments

This research was funded by the Norwegian Research Council (Project No 187979). The authors would like to thank Silje Henriksen for her role as participating investigator during recruitment.

### Conflicts of Interest

FD was employed by Changetech AS, which developed the Less Stress intervention, at the time of investigation. PK has a financial interest in the intervention as a shareholder in Changetech AS.

### Multimedia Appendix 1

Less Stress intervention screenshots.

[PDF File (Adobe PDF File), 962KB - jmir_v15i4e84_app1.pdf ]

### Multimedia Appendix 2

CONSORT EHEALTH checklist V1.6.2 [81].

[PDF File (Adobe PDF File), 990KB - jmir_v15i4e84_app2.pdf ]

### References

1. Ihlebaek C, Eriksen HR, Ursin H. Prevalence of subjective health complaints (SHC) in Norway. Scand J Public Health 2002;30(1):20-29. [Medline: 11928829]
2. Ipsos Public Affairs. The Associated Press International Affairs Poll. Washington, DC: Ipsos Public Affairs; 2006. URL: http://surveys.ap.org/data/Ipsos/international/2006-11%20AP%20Globus%20topline_112606.pdf [accessed 2013-02-06] [WebCite Cache ID 6EEi0IQOG]

3. American Psychological Association. Stress in America: Our Health at Risk. 2012 Jan 11. URL: http://www.apa.org/news/press/releases/stress/2011/final-2011.pdf [accessed 2013-02-06] [WebCite Cache ID 6EEhmDYsY]
4. Oliver MI, Pearson N, Coe N, Gunnell D. Help-seeking behaviour in men and women with common mental health problems: cross-sectional study. Br J Psychiatry 2005 Apr;186:297-301 [FREE Full text] [doi: 10.1192/bjp.186.4.297] [Medline: 15802685]
5. Kivimäki M, Virtanen M, Elovainio M, Kouvonen A, Väänänen A, Vahtera J. Work stress in the etiology of coronary heart disease--a meta-analysis. Scand J Work Environ Health 2006 Dec;32(6):431-442 [FREE Full text] [Medline: 17173200]
6. Golden SH. A review of the evidence for a neuroendocrine link between stress, depression and diabetes mellitus. Curr Diabetes Rev 2007 Nov;3(4):252-259. [Medline: 18220683]
7. Burke HM, Davis MC, Otte C, Mohr DC. Depression and cortisol responses to psychological stress: a meta-analysis. Psychoneuroendocrinology 2005 Oct;30(9):846-856. [doi: 10.1016/j.psyneuen.2005.02.010] [Medline: 15961250]
8. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 2006 Nov;3(11):e442 [FREE Full text] [doi: 10.1371/journal.pmed.0030442] [Medline: 17132052]
9. Health and Safety Executive. 2012. Stress-related and psychological disorders in Great Britain (GB) URL: http://www.hse.gov.uk/statistics/causdis/stress/index.htm [accessed 2013-02-06] [WebCite Cache ID 6EEi7wKuc]
10. National Institute of Occupational Health. Faktabok om arbeidsmiljø og helse 2011: Status og utviklingstrekk Facts about working environment and health 2011: Status and trends. Oslo, Norway: National Institute of Occupational Health; 2011.
11. Bohlmeijer E, Prenger R, Taal E, Cuijpers P. The effects of mindfulness-based stress reduction therapy on mental health of adults with a chronic medical disease: a meta-analysis. J Psychosom Res 2010 Jun;68(6):539-544. [doi: 10.1016/j.jpsychores.2009.10.005] [Medline: 20488270]
12. Chiesa A, Serretti A. Mindfulness-based stress reduction for stress management in healthy people: a review and meta-analysis. J Altern Complement Med 2009 May;15(5):593-600. [doi: 10.1089/acm.2008.0495] [Medline: 19432513]
13. Rainforth MV, Schneider RH, Nidich SI, Gaylord-King C, Salerno JW, Anderson JW. Stress reduction programs in patients with elevated blood pressure: a systematic review and meta-analysis. Curr Hypertens Rep 2007 Dec;9(6):520-528 [FREE Full text] [Medline: 18350109]
14. Richardson KM, Rothstein HR. Effects of occupational stress management intervention programs: a meta-analysis. J Occup Health Psychol 2008 Jan;13(1):69-93. [doi: 10.1037/1076-8998.13.1.69] [Medline: 18211170]
15. Amstadter AB, Broman-Fulks J, Zinzow H, Ruggiero KJ, Cercone J. Internet-based interventions for traumatic stress-related mental health problems: a review and suggestion for future research. Clin Psychol Rev 2009 Jul;29(5):410-420 [FREE Full text] [doi: 10.1016/j.cpr.2009.04.001] [Medline: 19403215]
16. van Bastelaar KM, Pouwer F, Cuijpers P, Riper H, Snoek FJ. Web-based depression treatment for type 1 and type 2 diabetic patients: a randomized, controlled trial. Diabetes Care 2011 Feb;34(2):320-325 [FREE Full text] [doi: 10.2337/dc10-1248] [Medline: 21216855]
17. Matano RA, Futa KT, Wanat SF, Mussman LM, Leung CW. The Employee Stress and Alcohol Project: the development of a computer-based alcohol abuse prevention program for employees. J Behav Health Serv Res 2000 May;27(2):152-165. [Medline: 10795126]
18. Dolezal-Wood S, Belar CD, Snibbe J. A comparison of computer-assisted psychotherapy cognitive-behavioral therapy in groups. J Clin Psychol Med S 1998;5(1):103-115. [doi: 10.1023/A:1026210020906]
19. Shimazu A, Kawakami N, Irimajiri H, Sakamoto M, Amano S. Effects of web-based psychoeducation on self-efficacy, problem solving behavior, stress responses and job satisfaction among workers: a controlled clinical trial. J Occup Health 2005 Sep;47(5):405-413 [FREE Full text] [Medline: 16230834]
20. Hasson D, Anderberg UM, Theorell T, Arnetz BB. Psychophysiological effects of a web-based stress management system: a prospective, randomized controlled intervention study of IT and media workers [ISRCTN54254861]. BMC Public Health 2005;5:78 [FREE Full text] [doi: 10.1186/1471-2458-5-78] [Medline: 16042796]
21. Cook RR, 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. [doi: 10.2196/jmir.9.2.e17]
22. Abbott JA, Klein B, Hamilton C, Rosenthal A. The impact of online resilience training for sales managers on wellbeing and work performance. EJAP 2009;5(1):89-95. [doi: http://dx.doi.org/10.7790/ejap.v5i1.145]
23. Eisen K, Allen G, Bollash M, Pescatello L. Stress management in the workplace: A comparison of a computer-based and an in-person stress-management intervention. Comput Hum Behav 2008 Mar;24(2):486-496. [doi: doi:10.1016/j.chb.2007.02.003]
24. van Straten A, Cuijpers P, Smits N. Effectiveness of a web-based self-help intervention for symptoms of depression, anxiety, and stress: randomized controlled trial. J Med Internet Res 2008;10(1):e7 [FREE Full text] [doi: 10.2196/jmir.954] [Medline: 18364344]
25. Hänggi Y. Kann web-basierte Stressprävention erfolgreichsein? Zeitschrift für Klinische Psychologie und Psychotherapie 2006 Jul 2006;35(3):169-177. [doi: 10.1026/1616-3443.35.3.169]
26. Wade SL, Carey J, Wolfe CR. An online family intervention to reduce parental distress following pediatric brain injury. J Consult Clin Psychol 2006 Jun;74(3):445-454. [doi: 10.1037/0022-006X.74.3.445] [Medline: 16822102]

27. 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. [doi: 10.1097/01.mlr.0000233678.80203.c1] [Medline: 17063127]
28. 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] [doi: 10.1002/art.23817] [Medline: 18576310]
29. Zetterqvist K, Maanmies J, Ström L, Andersson G. Randomized controlled trial of internet-based stress management. Cogn Behav Ther 2003;32(3):151-160. [doi: 10.1080/16506070302316] [Medline: 16291546]
30. Christensen H, Griffiths KM, Mackinnon AJ, Brittliffe K. Online randomized controlled trial of brief and full cognitive behaviour therapy for depression. Psychol Med 2006 Dec;36(12):1737-1746. [doi: 10.1017/S0033291706008695] [Medline: 16938144]
31. Richards JC, Klein B, Austin DW. Internet cognitive behavioural therapy for panic disorder: Does the inclusion of stress management information improve end-state functioning? Clin Psychol 2006 Mar;10(1):2-15. [doi: 10.1080/13284200500378795]
32. Prochaska JO, Butterworth S, Redding CA, Burden V, Perrin N, Leo M, et al. Initial efficacy of MI, TTM tailoring and HRI's with multiple behaviors for employee health promotion. Prev Med 2008 Mar;46(3):226-231 [FREE Full text] [doi: 10.1016/j.ypmed.2007.11.007] [Medline: 18155287]
33. Baer RA. Mindfulness training as a clinical intervention: A conceptual and empirical review. Clin Psychol-Sci Pr 2003;10(2):125-143. [doi: 10.1093/clipsy.bpg015]
34. Grossman P, Niemann L, Schmidt S, Walach H. Mindfulness-based stress reduction and health benefits: A meta-analysis. J Psychosom Res 2004:35-43. [doi: doi:10.1016/S0022-3999(03)00573-7]
35. Dobkin PL, Zhao Q. Increased mindfulness--the active component of the mindfulness-based stress reduction program? Complement Ther Clin Pract 2011 Feb;17(1):22-27. [doi: 10.1016/j.ctcp.2010.03.002] [Medline: 21168110]
36. Brown KW, Ryan RM. The benefits of being present: mindfulness and its role in psychological well-being. J Pers Soc Psychol 2003 Apr;84(4):822-848. [Medline: 12703651]
37. Evans DR, Baer RA, Segerstrom SC. The effects of mindfulness and self-consciousness on persistence. Pers Indiv Differ 2009 Sep 2009;47(4):379-382. [doi: 10.1016/j.paid.2009.03.026]
38. Weinstein N, Brown KW, Ryan RM. A multi-method examination of the effects of mindfulness on stress attribution, coping, and emotional well-being. J Res Pers 2009 Jun 2009;43(3):374-385. [doi: 10.1016/j.jrp.2008.12.008]
39. Palmer A, Rodger S. Mindfulness, stress, coping among university students. Canadian Journal of Counselling and Psychotherapy/Revue canadienne de counseling et de psychothérapie 2009;43(3):198-212.
40. Steel P. The nature of procrastination: a meta-analytic and theoretical review of quintessential self-regulatory failure. Psychol Bull 2007 Jan;133(1):65-94. [doi: 10.1037/0033-2909.133.1.65] [Medline: 17201571]
41. Stead R, Shanahan MJ, Neufeld RWJ. 'I'll go to therapy, eventually': Procrastination, stress and mental health. Pers Indiv Differ 2010 Aug;49(3):175-180. [doi: 10.1016/j.paid.2010.03.028]
42. Sirois FM. “I’ll look after my health, later”: A replication and extension of the procrastination–health model with community-dwelling adults. Pers Indiv Differ 2007 Jul 2007;43(1):15-26. [doi: 10.1016/j.paid.2006.11.003]
43. Fernie BA, Spada MM. Metacognitions about procrastination: A preliminary investigation. Behav Cogn Psychoth 2008 Feb;36(3):359-364. [doi: 10.1017/S135246580800413X]
44. Tice DM, Baumeister RF. Longitudinal study of procrastination, performance, stress, and health: The costs and benefits of dawdling. Psychol Sci 1997;8(6):454-458.
45. Moon SM, Illingworth AJ. Exploring the dynamic nature of procrastination: A latent growth curve analysis of academic procrastination. Pers Indiv Differ 2005 Jan;38(2):297-309. [doi: 10.1016/j.paid.2004.04.009]
46. Affleck G, Tennen H, Urrows S, Higgins P. Person and contextual features of daily stress reactivity: individual differences in relations of undesirable daily events with mood disturbance and chronic pain intensity. J Pers Soc Psychol 1994 Feb;66(2):329-340. [Medline: 8195989]
47. Matud MP. Gender differences in stress and coping styles. Pers Indiv Differ 2004 Nov 2004;37(7):1401-1415. [doi: 10.1016/j.paid.2004.01.010]
48. Terrill AL, Garofalo JP, Soliday E, Craft R. Multiple roles and stress burden in women: A conceptual model of heart disease risk. J Appl Biobehav Res 2012 Mar;17(1):4-22. [doi: 10.1111/j.1751-9861.2011.00071.x]
49. Fredman L, Cauley JA, Hochberg M, Ensrud KE, Doros G, Study of Osteoporotic Fractures. Mortality associated with caregiving, general stress, and caregiving-related stress in elderly women: results of caregiver-study of osteoporotic fractures. J Am Geriatr Soc 2010 May;58(5):937-943 [FREE Full text] [doi: 10.1111/j.1532-5415.2010.02808.x] [Medline: 20374395]
50. Amirkhan J, Auyeung B. Coping with stress across the lifespan: Absolute vs. relative changes in strategy. J Appl Dev Psychol 2007 Jul 2007;28(4):298-317. [doi: 10.1016/j.appdev.2007.04.002]
51. Steel P, Ferrari J. Sex, education and procrastination: an epidemiological study of procrastinators' characteristics from a global sample. Eur J Pers 2012 Apr 2012;27(1):51-58. [doi: 10.1002/per.1851]
52. Compas BE. An agenda for coping research and theory: basic and applied developmental issues. Int J Behav Dev 1998 Jun 1998;22(2):231-237. [doi: 10.1080/016502598384351]

53. Stone AA, Schwartz JE, Broderick JE, Deaton A. A snapshot of the age distribution of psychological well-being in the United States. Proc Natl Acad Sci U S A 2010 Jun 1;107(22):9985-9990 [FREE Full text] [doi: 10.1073/pnas.1003744107] [Medline: 20479218]
54. Bergene AC, Mamelund SE, Steen AH. Norsk arbeidsliv 2012 - Svekket motstand i gode tider Norwegian worklife 2012 - Impaired resistance in good times. Oslo, Norway: Work Research Institute; 2012.
55. Gallo LC, Matthews KA. Understanding the association between socioeconomic status and physical health: do negative emotions play a role? Psychol Bull 2003 Jan;129(1):10-51. [Medline: 12555793]
56. Galanakis M, Stalikas A, Kallia H, Karagianni C, Karela C. Gender differences in experiencing occupational stress: The role of age, education and marital status. Stress Health 2009;25:397-404. [doi: 10.1002/smi]
57. Random integer generator. URL: http://www.random.org/integers/ [accessed 2013-04-05] [WebCite Cache ID 6Fe86JCxr]
58. Wells A. Detached mindfulness in cognitive therapy: a metacognitive analysis and ten techniques. J Rat-Emo Cognitive-Behav Ther 2006 Mar 2006;23(4):337-355. [doi: 10.1007/s10942-005-0018-6]
59. Demonstration of Less Stress intervention. 2013. URL: http://bit.ly/less_stress [WebCite Cache ID 6Fe965pAO]
60. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [Medline: 7726811]
61. Mann L, Burnett P, Radford M, Ford S. The Melbourne decision making questionnaire: an instrument for measuring patterns for coping with decisional conflict. J. Behav. Decis. Making 1997 Mar 1997;10(1):1-19. [doi: 10.1002/(SICI)1099-0771(199703)10:1<1::AID-BDM242>3.0.CO;2-X]
62. Graham JW. Missing data analysis: making it work in the real world. Annu Rev Psychol 2009 Jan;60:549-576. [doi: 10.1146/annurev.psych.58.110405.085530] [Medline: 18652544]
63. Goodman JS. Assessing the non-random sampling effects of subject attrition in longitudinal research. J Manage 1996 Aug 1996;22(4):627-652. [doi: 10.1177/014920639602200405]
64. Hays WL. Statistics. New York: Holt, Rinehart and Winston; 1988.
65. Singer JD, Willett JB. Applied longitudinal data analysis: modeling change and event occurrence. Oxford: Oxford University Press; 2003.
66. Bliese PD, Ployhart RE. Growth modeling using random coefficient models: Model building, testing, and illustrations. Organ Res Methods 2002 Oct;5(4):362-387. [doi: 10.1177/1094428102237116]
67. MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. Psychol Methods 2002 Mar;7(1):83-104 [FREE Full text] [Medline: 11928892]
68. Preacher KJ, Hayes AF. SPSS and SAS procedures for estimating indirect effects in simple mediation models. Behav Res Methods Instrum Comput 2004 Nov;36(4):717-731. [Medline: 15641418]
69. Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav Res Methods 2008 Aug;40(3):879-891. [Medline: 18697684]
70. Mackinnon DP, Lockwood CM, Williams J. Confidence limits for the indirect effect: distribution of the product and resampling methods. Multivariate Behav Res 2004 Jan 1;39(1):99 [FREE Full text] [doi: 10.1207/s15327906mbr3901_4] [Medline: 20157642]
71. Selig JP, Preacher KJ. Monte Carlo method for assessing mediation: An interactive tool for creating confidence intervals for indirect effects. 2008. URL: http://www.quantpsy.org/medmc/medmc.htm [accessed 2013-02-07] [WebCite Cache ID 6EFQJpXEV]
72. Brynjolfsson E, Hu YJ, Simester D. Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. Manage Sci 2011 Jun;57(8):1373-1386. [doi: 10.1287/mnsc.1110.1371]
73. Statistics Norway. 2012 Oct 12. ICT usage in households, 2nd quarter URL: http://ssb.no/en/teknologi-og-innovasjon/statistikker/ikthus [accessed 2013-03-10] [WebCite Cache ID 6F0jjsxN4]
74. Mathy RM, Kerr DL, Haydin BM. Methodological rigor and ethical considerations in internet-mediated research. Psychother Theor Res Pract Train 2003;40(1):77-85. [doi: 10.1037/0033-3204.40.1/2.77]
75. Stopponi MA, Alexander GL, McClure JB, Carroll NM, Divine GW, Calvi JH, et al. Recruitment to a randomized web-based nutritional intervention trial: characteristics of participants compared to non-participants. J Med Internet Res 2009;11(3):e38 [FREE Full text] [doi: 10.2196/jmir.1086] [Medline: 19709990]
76. Bull SS, Levine D, Schmiege S, Santelli J. Recruitment and retention of youth for research using social media: Experiences from the Just/Us study. Vulnerable Children and Youth Studies 2012 Dec 2012:1-11. [doi: 10.1080/17450128.2012.748238]
77. Fenner Y, Garland SM, Moore EE, Jayasinghe Y, Fletcher A, Tabrizi SN, et al. Web-based recruiting for health research using a social networking site: an exploratory study. J Med Internet Res 2012 Feb;14(1):e20 [FREE Full text] [doi: 10.2196/jmir.1978] [Medline: 22297093]
78. Aguinis H. Statistical power with moderated mutliple regression in management research. J Manage 1995 Dec;21(6):1141-1158. [doi: 10.1177/014920639502100607]
79. Krull JL, MacKinnon DP. Multilevel mediation modeling in group-based intervention studies. Eval Rev 1999 Aug;23(4):418-444. [Medline: 10558394]
80. Hayes AF, Preacher KJ. Quantifying and testing indirect effects in simple mediation models when the constituent paths are nonlinear. Multivar Behav Res 2010 Jul;45(4):627-660. [doi: 10.1080/00273171.2010.498290]

81. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

## Abbreviations

**–2LL:** –2 log likelihood
**AIC:** Akaike information criterion
**DASS-S:** Depression Anxiety Stress Scale–stress subscale
**ICC:** intraclass correlation coefficient
**LS:** Less Stress
**MAAS:** Mindfulness Attention Awareness Scale
**MCMAM:** Monte Carlo Method for Assessing Mediation
**MDMQ-P:** Melbourne Decision Making Questionnaire–procrastination subscale
**MI:** multiple imputation

*Edited by G Eysenbach; submitted 07.02.13; peer-reviewed by J Abbott, L Kilpatrick; comments to author 02.03.13; revised version received 10.03.13; accepted 26.03.13; published 22.04.13.*

*Please cite as:*
*Drozd F, Raeder S, Kraft P, Bjørkli CA*
*Multilevel Growth Curve Analyses of Treatment Effects of a Web-Based Intervention for Stress Reduction: Randomized Controlled Trial*
*J Med Internet Res 2013;15(4):e84*
*URL: http://www.jmir.org/2013/4/e84/*
*doi:10.2196/jmir.2570*
*PMID:23607962*

Original Paper

# Use of a Text Message-Based Pharmacovigilance Tool in Cambodia: Pilot Study

Sophie Baron[1], MSc Epi, DVM; Flavie Goutard[1,2], MSc Epi, DVM; Kunthy Nguon[1], BA; Arnaud Tarantola[1], MSc Epi, MD

[1]Epidemiology and Public Health Unit (epi@ipc), Institut Pasteur du Cambodge, Phnom Penh, Cambodia

[2]AGIRs unit, Centre de coopération internationale en recherche agronomique pour le développement (CIRAD), Montpellier, France

**Corresponding Author:**
Arnaud Tarantola, MSc Epi, MD
Epidemiology and Public Health Unit (epi@ipc)
Institut Pasteur du Cambodge
5, Bvd Monivong
Phnom Penh, BP983
Cambodia
Phone: 855 12 333 650
Fax: 855 23 706 606
Email: atarantola@pasteur-kh.org

## Abstract

**Background:** There is no functional pharmacovigilance system in Cambodia to our knowledge. Mobile phone–based tools, such as short message service (SMS) text messages, are increasingly used for surveillance purposes.

**Objective:** To pilot-test the FrontlineSMS mobile phone–based tool for notification of adverse events, using Cambodia's only International Vaccination Center at the Institut Pasteur du Cambodge as a field site.

**Methods:** People receiving vaccinations, aged over 18 years, and who owned a cell phone were recruited in the study following informed consent. The names and mobile phone numbers of the participants interviewed were entered each day into the FrontlineSMS software. Two days after being vaccinated, participants received an automatically generated SMS text message asking whether any adverse events had occurred. Their SMS reply was number-coded and exported from the software daily to an Excel spreadsheet and examined before being saved. If the participant replied with a code for a severe adverse event (8 or 9), they were automatically advised to consult the nearest doctor.

**Results:** The active surveillance study was conducted over 72 days in the spring of 2012. Patients agreed to be asked by SMS text message whether unwanted events had occurred after vaccination. Of 1331 persons aged over 18 years referred to the vaccination unit, 184 (13.8%) were asked and agreed to participate. When texted for clinical status 48 hours after vaccination, 52 (28.3%) participants did not reply, 101 (54.9%) sent an immediate SMS reply, and 31 (16.8%) sent an SMS reply after additional prompting. Of the initial 184 participants, 132 (71.7%) replied. These 132 participants received 135 vaccine doses and 109 (82.6%) reported no adverse events, whereas 23 (17.4%) reported adverse events, all benign.

**Conclusions:** Notification using an SMS-based text message system is already used in Cambodia for syndromic surveillance in health centers and reporting by health care workers. Our results show that such tools can also be useful for notification by patients or health users in Cambodia, especially in an urban setting.

***(J Med Internet Res 2013;15(4):e68)*** doi:10.2196/jmir.2477

**KEYWORDS**

cellular phone; text messages; texting; short message service; vaccines; adverse events; surveillance; adverse drug reaction reporting systems; pharmacovigilance

## Introduction

The burden of disease can be threefold in developing countries faced with communicable diseases, noncommunicable diseases, and sociobehavioral illnesses [1]. Most of these may require treatment or prevention through medication or vaccines. In addition to the expectable adverse events in the normal usage of registered drugs, developing countries also face a plague of

counterfeit or substandard drugs [2-5]. Pharmacovigilance—a form of epidemiological surveillance that monitors the occurrence of adverse events of drugs [6,7] or vaccines [8-10] to guide timely corrective action and mitigate risk—is an essential tool for patient safety in developed countries and in an increasing number of developing countries [11]. Cambodia, a developing country in Southeast Asia, is not among the 109 countries participating in the World Health Organization (WHO) Programme for International Drug Monitoring maintained in collaboration with the Uppsala Monitoring Center in Uppsala, Sweden [12]. To our knowledge, there is no functional pharmacovigilance program in Cambodia.

Pharmacovigilance has been conducted using complex and rigorous notification procedures in developed countries, but it faces challenges in developing countries because of issues with clinician awareness, clinical expertise, or nonfunctional reporting systems [11,13,14]. Therefore, some programs or countries have circumvented these challenges by implementing mobile phone-based and Web-based tools using short message service (SMS) text messaging for immediate notification of adverse events [15,16].

FrontlineSMS is one such reporting tool [17]. It has been used in Cambodia for various health-related projects, but always for information exchange among health providers and stakeholders [18-20]. An Epidemiology Master student helped pilot its use for surveillance in direct link with health users in a vaccination center. More than 25,000 vaccine doses were administered to 16,630 health users in 2012 (Figure 1) referred to the International Vaccination Center at the Institut Pasteur du Cambodge (IPC), the only such vaccination center in Cambodia (Figure 2). The aim of this study was to field-test the FrontlineSMS software to see whether it could provide effective and timely notification of vaccine adverse events.

**Figure 1.** Vaccination at the International Vaccination Center, Institut Pasteur du Cambodge.

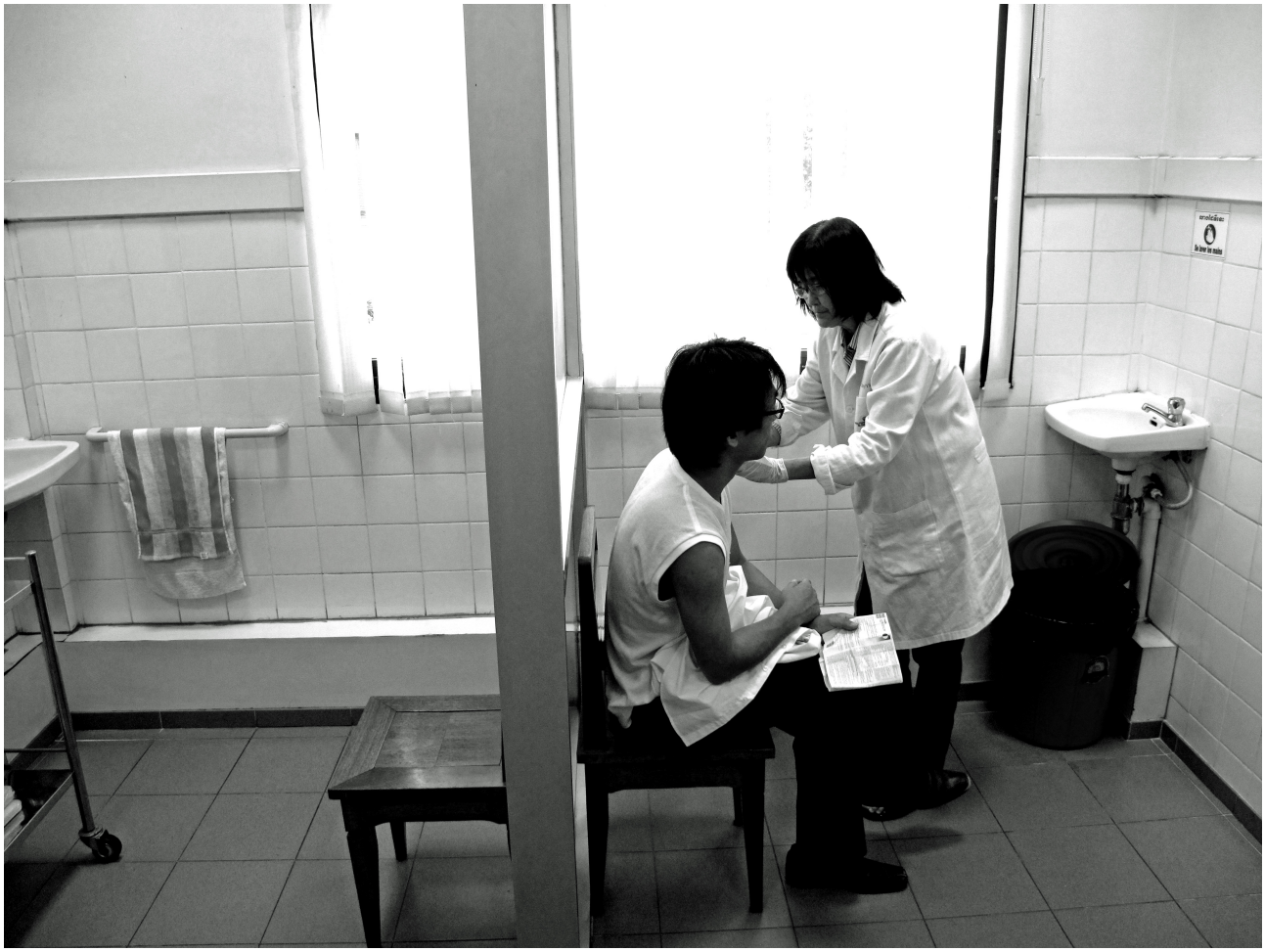


**Figure 2.** The waiting room of the international vaccination center at the Institut Pasteur du Cambodge.

## Method

The research project received ethical approval from the national ethics committee on health research on February 17, 2012. Data collection began March 12, 2012, and ended May 31, 2012.

During that period, a research assistant from the epidemiology and public health unit at IPC (a native Khmer speaker) spent several hours each day at the International Vaccination Center to recruit participants. Participants were eligible to be included in the study if they were aged over 18 years, came to the center to be vaccinated themselves (ie, not for their children), agreed to participate, owned a cell phone, and knew how to send SMS text messages. If so, the research assistant informed them about the study, read through the protocol, explained the objectives of the pharmacovigilance project in Khmer, and asked the participants to complete an informed consent form. Information about their name, age, place of residence, type of vaccine, and mobile phone service provider were collected. The names and mobile phone numbers of the participants interviewed were entered each day in the FrontlineSMS software.

Two days after being vaccinated, participants received an automatically generated SMS text message. This message thanked them for participating and asked whether any adverse events had occurred. Their SMS reply was number-coded as follows: 0=no adverse event, 1=mainly redness and/or pain at the injection site, 2=mainly fatigue and/or weakness, 3=mainly headaches, 4=mainly fever, 5=mainly runny or congested nose, 6=mainly muscle pains, 7=mainly abdominal pain, 8=seizures or neurological problems, and 9=severe allergic reaction. Only 1 code was allowed per reply. Messages received were exported from the software daily to an Excel spreadsheet and examined before being saved.

Upon receiving the text reply, a software-generated message was sent. If the codes corresponded to no or a moderate adverse event (codes 0-7), the reply was: “We have received your message. Thank you for having participated.” If the participant sent back a code for a severe adverse event (8 or 9), the reply read: “You have reported a severe adverse event. Please consult the nearest doctor as soon as possible. If in Phnom Penh, we recommend that you refer to Hôpital Calmette.” (The Hôpital Calmette is a national reference hospital where the emergency department team had been informed of the study.) In cases when a severe adverse event was reported, the research assistant called the participant’s cell phone number directly to follow-up on the participant. If the case was referred to Hôpital Calmette, transportation costs were covered by the study. There was no other financial compensation to participants, including for costs associated with sending SMS text messages.

## Results

The study took place between March 13 and May 25, 2012, for a total of 72 days: 53 days of which the International Vaccination Center was open and 19 days of which it was closed (weekend or national holidays, Figure 3). During some of those days, the research assistant did not recruit participants because she was involved in another study.

During the 53 days that the International Vaccination Center was open in that period, 1331 persons aged over 18 years (684 women and 647 men) came for vaccinations at the center (mean 25.1 persons aged ≥18 years per day, Figure 4).

Of the 1331 vaccinees, 184 were asked and agreed to participate to this pilot study (97 female, 87 male). Of these 184 participants, 165 (90.2%) resided in Phnom Penh; 6 (3.3%) in Kampong Speu; 3 (1.6%) each in Kandal and Kompong Cham, respectively; 2 (1.1%) in Siem Reap; and 1 (0.5%) each in Takeo, Battambang, Mondulkiri, and Rattanakiri, respectively. Mean age of all 1331 health users during that period was 34.8 years (SD 12.7, range 18-95), although the subgroup of study participants was younger (mean 26.9 years, SD 7.8, range 18-65). The 184 participants subscribed to 6 cell phone companies, the largest of which accounted for 83 (45.1%) of the participants.

The 184 health users who initially agreed to participate in the study received a total of 192 vaccine doses for 17 different diseases (Table 1).

**Table 1.** Vaccines administered to study participants and reported adverse events for the pharmacovigilance pilot study at the Institut Pasteur du Cambodge from March to May 2012 (N=184).

| Vaccine[a] | No adverse event | Mainly redness and/or pain at the injection site | Mainly fatigue and/or weakness | Mainly fever | Mainly runny or congested nose | No reply | Total |
|---|---|---|---|---|---|---|---|
| Hepatitis A | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| Hepatitis B | 39 | 4 | 3 | 1 | 0 | 30 | 77 |
| Haemophilus influenza | 2 | 0 | 0 | 1 | 0 | 0 | 3 |
| Japanese encephalitis | 9 | 0 | 0 | 0 | 0 | 1 | 10 |
| MMR | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| Meningitis | 2 | 0 | 0 | 0 | 0 | 0 | 2 |
| Chickenpox | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| Pneumococcus | 2 | 0 | 0 | 0 | 0 | 2 | 4 |
| Rubella | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| Tetanus | 13 | 3 | 1 | 0 | 0 | 3 | 20 |
| Tetanus & rabies | 2 | 0 | 0 | 0 | 0 | 2 | 4 |
| DTCP | 2 | 1 | 1 | 0 | 1 | 0 | 5 |
| Typhoid | 2 | 0 | 0 | 0 | 0 | 0 | 2 |
| HPV | 10 | 3 | 0 | 1 | 2 | 4 | 20 |
| Influenza | 17 | 0 | 0 | 0 | 0 | 5 | 22 |
| Influenza & Hib | 2 | 0 | 0 | 0 | 0 | 0 | 2 |
| Influenza & tetanus | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| Rabies | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| Yellow fever | 4 | 0 | 0 | 0 | 0 | 2 | 6 |
| Yellow fever & meningitis | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| Total | 109 | 12 | 5 | 3 | 3 | 52 | 184 |

[a] MMR: measles, mumps, rubella; DTCP: diphtheria, tetanus, pertussis, polio; HPV: human papillomavirus; Hib: *Haemophilus influenzae* serotype B.

Most of the vaccinations given were against hepatitis B (41.8%), influenza (11.9%), tetanus (10.9%), and human papillomavirus (HPV) (10.9%). Fifty-two participants (28.2%) did not send a reply (including 30 of 77 who received vaccination against hepatitis B), and 132 (71.7%) did send a reply, of which 101 (76.5%) completed the study as per the study protocol and sent a correct SMS text reply, whereas 31 (23.5%) participants had to be contacted twice because they replied incorrectly or did not reply at all (Figure 2). These 132 respondents received a total of 137 doses of vaccine. In all, 109 (82.6%) respondents reported no occurrences of adverse events, whereas 23 (17.4%) did report adverse events, none of which were severe (Table 1). Twelve (52.2%) of these adverse events pertained to redness at the injection site, and 3 (13.0%) pertained to fever. The time between the initial sending of the SMS text and the reply from participants was documented in 120 messages (mean 0.4 days, SD 0.82, range 0-5). The reply was received within the same day in 91 (75.8%) of documented answers.

**Figure 3.** Number of health users at the international vaccination center and inclusions in the SMS pilot study, Institut Pasteur du Cambodge, March-May 2012.

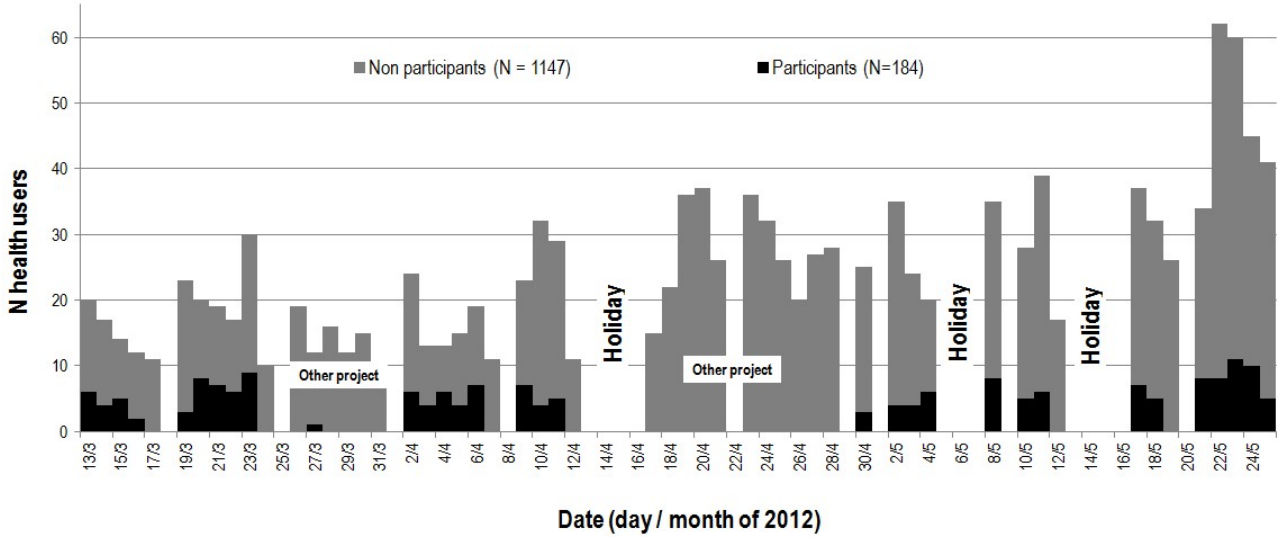
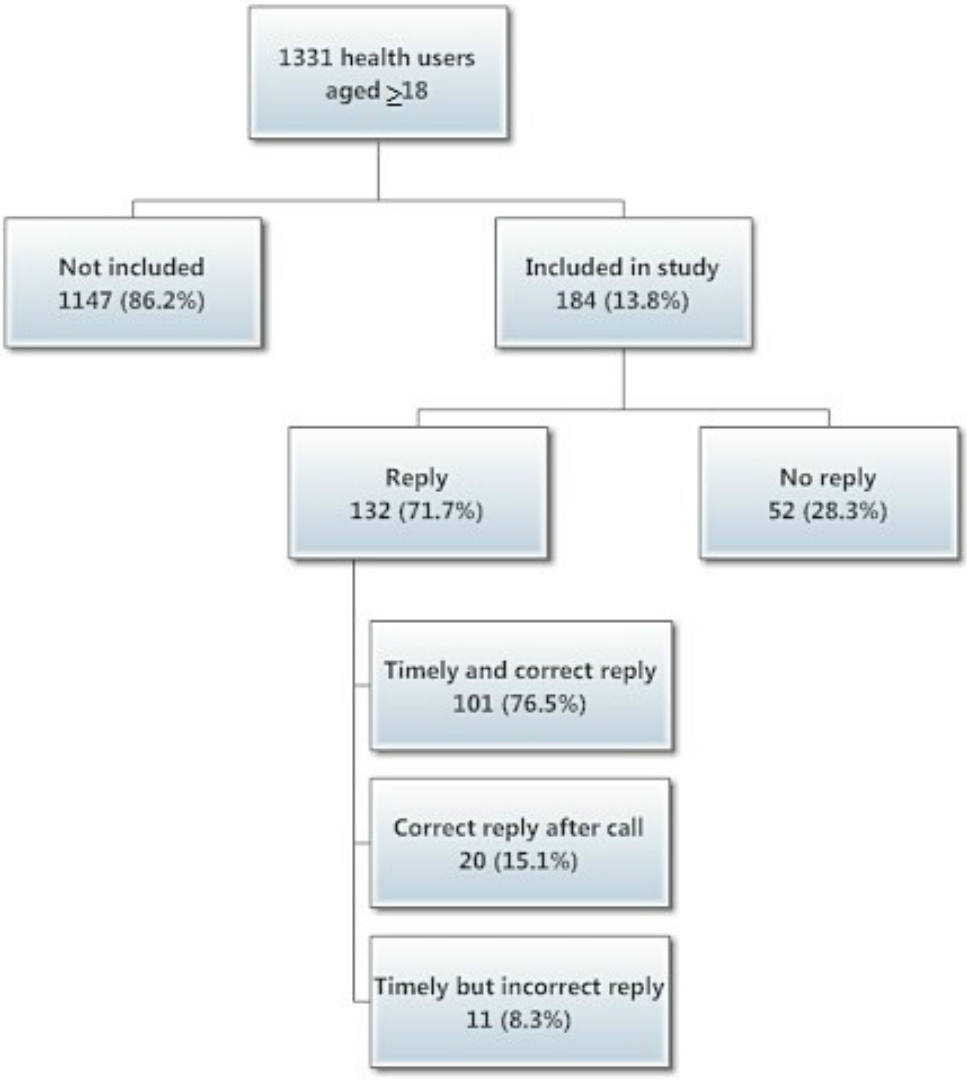



**Figure 4.** Flowchart of patient recruitment in the pilot SMS text-based pharmacovigilance study at International Vaccination Center, Institut Pasteur du Cambodge, March-May 2012 (percentage total may not equal 100% due to rounding).



## Discussion

The objective of this pilot study was to assess the feasibility of an SMS text-based reporting tool for pharmacovigilance in Cambodia. According to the International Communications Union database, there were 69.9 mobile-cellular telephone subscriptions per 100 (%) inhabitants in Cambodia in 2011 (slightly less than India), up from 7.95% in 2005 [21]. This translates to approximately 10 million cellular telephone subscriptions in Cambodia in 2011, a country with an estimated population of approximately 14 million [22].

To our knowledge, this pilot SMS text-based system for active detection of adverse events following vaccination is the first functional pharmacovigilance system in Cambodia, the first SMS text-based surveillance system relying on health users' participation in Cambodia, and the first to use SMS text-based surveillance for adverse event detection anywhere [20]. In this project, operating costs were modest, with only part-time activity dedicated to entering cell phone numbers, downloading SMS replies, and checking that no SMS texts received alerted to severe adverse events.

During the 2.5-month pilot phase, the participation rate was high with a 71.7% response rate. Of 77 persons who were vaccinated against hepatitis B, 30 (39%) sent no reply. Although this percentage may seem high, the hepatitis B vaccines were the most frequently administered, and the small study group makes nonreply percentages vary greatly.

With 132 participants replying and 24 reporting adverse events (none severe), the prevalence of adverse events was slightly higher than expected. The expected rate of adverse events following immunization is estimated to be in the order of 10%, except for diphtheria, tetanus, pertussis (DTP) or tetanus boosters in which fever occurs in nearly half of recipients [23]. In 2008, in Australia, 1542 adverse events (7.2 per 100,000 population) following immunization were notified by manufacturers, health professionals, or the public [24]. This passive surveillance system also includes children, who are the main recipients of vaccines. Of these adverse events, 41% were site reactions and 16% were fevers. Our data, albeit on a far more modest scale and not including children, showed comparable percentages (52.2% and 13.0%, respectively). A study conducted in the United Kingdom showed that patients tend to report more benign adverse drug reactions than health care providers, and concluded that patient-based pharmacovigilance may usefully complement health care provider reporting of adverse events [25].

The participants' response rate was unexpectedly high, probably enhanced by the use of short and simple SMS text-based reply codes. This research project also had a research assistant to explain the protocol at length in the national language, which would not be the case in a daily operational setting where a simple leaflet or poster would provide information to the health users.

Our study suffers from biases and limitations. Firstly, the number of participants was very limited. The epidemiological findings of such a small and short study on adverse events are difficult to extrapolate. Secondly, the patients recruited all attended the IPC's International Vaccination Center in Phnom Penh. Costs there may be higher because of state-of-the-art quality control carried out on vaccines imported primarily from Europe. Therefore, recruitment may have been biased toward younger, more affluent, well-informed, urban residents more accustomed to sending SMS text messages. This limits the possibility of extending similar systems outside of large urban centers within Cambodia, at least in the short term. Experience with SMS text-based surveillance in the rural setting found that many farmers did not know how to send text messages [18]. Thirdly, only the last-generation cellular telephones support Khmer fonts or pictures of Khmer-language text, whereas the vast majority of cellular phones do not. Lastly, some operators mainly offer voice-based communications and no data or SMS text transfers because telephone communications are relatively cheap and there is no advantage to sending a text message.

Bearing these limitations in mind, this small pilot study serves as a proof of concept that health user–sent, SMS text-based surveillance strategies can be used in an urban Cambodian setting. This is an important step in Cambodia where health surveillance systems facing numerous challenges may often be dysfunctional and where pharmacovigilance is absent. Our secondary objective was also met, which was to become proficient in the use of FrontlineSMS for other potential applications. Technology for mobile telephone–based active surveillance appears to be cheap, easy to implement, simple, and quick to be mastered by field staff. This approach will be used by the Epidemiology and Public Health Unit at the Institut Pasteur du Cambodge to implement follow-up programs, such as monitoring outcomes in rabies postexposure prophylaxis at IPC's Rabies Prevention Clinic (over 20,300 referrals in 2011) or in prospective studies on dengue in urban settings.

### Acknowledgments

We wish to gratefully acknowledge the health users at IPC's international vaccination center for agreeing to participate in this pilot study. This study was conducted thanks to a grant from the Pierre Ledoux Foundation and additional support from the Centre de coopération internationale en recherche agronomique pour le développement (Cirad). We also want to acknowledge the Social Impact Lab Foundation for developing and allowing free access to the open-source FrontlineSMS software package.

### References

1. World Health Organization. Health transition URL: http://www.who.int/trade/glossary/story050/en/index.html [accessed 2013-03-08] [WebCite Cache ID 6ExdJ5yyr]

2. Attaran A, Barry D, Basheer S, Bate R, Benton D, Chauvin J, et al. How to achieve international action on falsified and substandard medicines. BMJ 2012;345:e7381. [Medline: 23149211]
3. World Health Organization Working Group of Member States on Substandard/Spurious/Falsely-Labelled/Falsified/Counterfeit Medical Products. WHO's role in the prevention and control of medical products of compromised quality, safety and efficacy such as substandard/spurious/falsely-labelled/falsified/counterfeit medical products. 2011 Feb 17. URL: http://apps.who.int/gb/ssffc/pdf_files/A_SSFFC_WG3Rev1-en.pdf [accessed 2013-03-06] [WebCite Cache ID 6EvtcgElp]
4. Delepierre A, Gayot A, Carpentier A. Update on counterfeit antibiotics worldwide; public health risks. Med Mal Infect 2012 Jun;42(6):247-255. [doi: 10.1016/j.medmal.2012.04.007] [Medline: 22621827]
5. World Health Organization. Counterfeit and substandard drugs in Myanmar and Viet Nam: Report of a study carried out in cooperation with the Governments of Myanmar and Viet Nam. 1999. URL: http://apps.who.int/medicinedocs/pdf/s2276e/s2276e.pdf [accessed 2013-03-06] [WebCite Cache ID 6EvthjKGT]
6. Stergachis A, Bartlein RJ, Dodoo A, Nwokike J, Kachur SP. A situational analysis of pharmacovigilance plans in the Global Fund Malaria and US President's Malaria Initiative proposals. Malar J 2010;9:148 [FREE Full text] [doi: 10.1186/1475-2875-9-148] [Medline: 20509971]
7. Huff-Rousselle M, Simooya O, Kabwe V, Hollander I, Handema R, Mwango A, et al. Pharmacovigilance and new essential drugs in Africa: Zambia draws lessons from its own experiences and beyond. Glob Public Health 2007;2(2):184-203. [doi: 10.1080/17441690601063299] [Medline: 19280399]
8. Letourneau M, Wells G, Walop W, Duclos P. Improving global monitoring of vaccine safety: a survey of national centres participating in the WHO Programme for International Drug Monitoring. Drug Saf 2008;31(5):389-398. [Medline: 18422379]
9. Bentsi-Enchill AD, Zongo I, Khamassi S, Pless R, Thombiano R, Tiéndrebéogo S, et al. Monitoring of adverse events during the 2003 mass vaccination campaign with a trivalent meningococcal A/C/W135 polysaccharide vaccine in Burkina Faso. Vaccine 2007 Sep 3;25 Suppl 1:A72-A78. [doi: 10.1016/j.vaccine.2007.04.045] [Medline: 17544552]
10. Kraigher A, Ucakar V. Surveillance of adverse events following immunization against pandemic influenza in Slovenia in season 2009/10. Vaccine 2010 Jul 26;28(33):5467-5472. [doi: 10.1016/j.vaccine.2010.05.066] [Medline: 20598408]
11. Pirmohamed M, Atuah KN, Dodoo AN, Winstanley P. Pharmacovigilance in developing countries. BMJ 2007 Sep 8;335(7618):462 [FREE Full text] [doi: 10.1136/bmj.39323.586123.BE] [Medline: 17823149]
12. The Uppsala Monitoring Centre. WHO programme members URL: http://www.who-umc.org/DynPage.aspx?id=100653&mn1=7347&mn2=7252&mn3=7322&mn4=7442 [accessed 2012-10-18]
13. Edwards IR. The WHO World Alliance for Patient Safety: a new challenge or an old one neglected? Drug Saf 2005;28(5):379-386. [Medline: 15853440]
14. Fernandopulle RB, Weerasuriya K. What can consumer adverse drug reaction reporting add to existing health professional-based systems? Focus on the developing world. Drug Saf 2003;26(4):219-225. [Medline: 12608886]
15. Gaur RK. Cellular network based drug monitoring. Asia Pac J Public Health 2011 Nov;23(6):1025-1026. [doi: 10.1177/1010539511422741] [Medline: 22144716]
16. Curioso WH, Karras BT, Campos PE, Buendia C, Holmes KK, Kimball AM. Design and implementation of Cell-PREVEN: a real-time surveillance system for adverse events using cell phones in Peru. AMIA Annu Symp Proc 2005:176-180 [FREE Full text] [Medline: 16779025]
17. FrontlineSMS. Software URL: http://www.frontlinesms.com/the-software/ [accessed 2013-03-06] [WebCite Cache ID 6Evu3q4F6]
18. GlobalHealthHub.org. 2012 Apr 23. Data Collected via FrontlineSMS is used by Ministry of Agriculture in Cambodia URL: http://www.globalhealthhub.org/2012/04/23/data-collected-via-frontlinesms-is-used-by-ministry-of-agriculture-in/ [accessed 2013-03-06] [WebCite Cache ID 6EvuEn55B]
19. FrontlineSMS. 2011 Apr 20. Malaria Diagnosis in Real-Time via SMS URL: http://www.frontlinesms.com/2011/04/20/malaria-diagnosis-in-real-time-via-sms/ [accessed 2013-03-07] [WebCite Cache ID 6EvuN9LHa]
20. World Health Organization. Global Observatory for eHealth. 2011. mHealth: New horizons for health through mobile technologies URL: http://www.who.int/goe/publications/ehealth_series_vol3/en/index.html [accessed 2013-03-07] [WebCite Cache ID 6Evum9HK3]
21. International Telecommunications Union. Google Public Data. ICT Statistics Database: mobile-cellular subscriptions URL: https://www.google.com/publicdata/explore?ds=emi9ik86jcuic_ [accessed 2012-10-19] [WebCite Cache ID 6EzkX633X]
22. National Institute of Statistics, Ministry of Planning Phnom Penh, Cambodia. General Population Census of Cambodia 2008. 2008 Aug. URL: http://www.stat.go.jp/english/info/meetings/cambodia/pdf/pre_rep1.pdf [accessed 2013-03-07] [WebCite Cache ID 6EvuwRVmQ]
23. Working Group on Immunization Safety Division of Vaccines and Immunization. Pan American Health Organization Regional Office of the World Health Organization Division of Vaccines and Immunization. 2002. Immunization safety: How to address events allegedly attributable to vaccination or immunization? URL: http://www.who.int/vaccine_safety/publications/en/immunization_safety_E.pdf [accessed 2013-03-07] [WebCite Cache ID 6EvuzgNzd]
24. Menzies R, Mahajan D, Gold MS, Roomiani I, McIntyre P, Lawrence G. Annual report: surveillance of adverse events following immunisation in Australia, 2008. Commun Dis Intell Q Rep 2009 Dec;33(4):365-381. [Medline: 20301966]

25. Avery AJ, Anderson C, Bond CM, Fortnum H, Gifford A, Hannaford PC, et al. Evaluation of patient reporting of adverse drug reactions to the UK 'Yellow Card Scheme': literature review, descriptive and qualitative analyses, and questionnaire surveys. Health Technol Assess 2011 May;15(20):1-234, iii [FREE Full text] [doi: 10.3310/hta15200] [Medline: 21545758]

## Abbreviations

**DTP:** diphtheria, tetanus, polio
**DTCP:** diphtheria, tetanus, pertussis, polio
**HPV:** human papillomavirus
**IPC:** Institut Pasteur du Cambodge
**SMS:** short message service
**WHO:** World Health Organization

*Edited by G Eysenbach; submitted 13.12.12; peer-reviewed by JL Oliveira, B Edwards; comments to author 20.01.13; revised version received 07.02.13; accepted 13.02.13; published 16.04.13.*

*Please cite as:*
*Baron S, Goutard F, Nguon K, Tarantola A*
*Use of a Text Message-Based Pharmacovigilance Tool in Cambodia: Pilot Study*
*J Med Internet Res 2013;15(4):e68*
*URL: http://www.jmir.org/2013/4/e68/*
*doi:10.2196/jmir.2477*
*PMID:23591700*

Original Paper

# A Comparison of Two Delivery Modalities of a Mobile Phone-Based Assessment for Serious Mental Illness: Native Smartphone Application vs Text-Messaging Only Implementations

John Ainsworth[1], BSC(Hons), MSc; Jasper E Palmier-Claus[2], BSC, MSc, PhD; Matthew Machin[1], BEng, CEng, MIET; Christine Barrowclough[2], BA, MSc, PhD; Graham Dunn[3], BA, MSc, MA, PhD; Anne Rogers[4], BA, MSc, PhD; Iain Buchan[5], MD, FFPH; Emma Barkus[6], BA, MSc, PhD; Shitij Kapur[7], MBBS, PhD, FRCPC; Til Wykes[7], PhD; Richard S Hopkins[8], MB ChB; Shôn Lewis[9], MD, FMedSci

[1]NIBHI Manchester Health e-Research Centre, Institue of Population Health, University of Manchester, Manchester, United Kingdom

[2]School of Psychological Science, Division of Clinicial Psychology, University of Manchester, Manchester, United Kingdom

[3]Centre for Biostatistics, Institute of Population Health, University of Manchester, Manchester, United Kingdom

[4]Health Sciences, University of Southampton, Southampton, United Kingdom

[5]NIBHI Manchester Health e-Research Centre, Institute of Population Health, University of Manchester, Manchester, United Kingdom

[6]School of Psychology, University of Wollongong, Wollongong, Australia

[7]Institute of Psychiatry, Kings's College London, London, United Kingdom

[8]Manchester Mental Health and Social Care Trust, Manchester, United Kingdom

[9]Institute of Brain, Behaviour and Mental Health, University of Manchester, Manchester, United Kingdom

**Corresponding Author:**
John Ainsworth, BSC(Hons), MSc
NIBHI Manchester Health e-Research Centre
Institue of Population Health
University of Manchester
Oxford Road
Manchester, M13 9PL
United Kingdom
Phone: 44 1612751129
Fax: 44 1613067945
Email: John.Ainsworth@manchester.ac.uk

## Abstract

**Background:** Mobile phone–based assessment may represent a cost-effective and clinically effective method of monitoring psychotic symptoms in real-time. There are several software options, including the use of native smartphone applications and text messages (short message service, SMS). Little is known about the strengths and limitations of these two approaches in monitoring symptoms in individuals with serious mental illness.

**Objective:** The objective of this study was to compare two different delivery modalities of the same diagnostic assessment for individuals with non-affective psychosis—a native smartphone application employing a graphical, touch user interface against an SMS text-only implementation. The overall hypothesis of the study was that patient participants with sewrious mental illness would find both delivery modalities feasible and acceptable to use, measured by the quantitative post-assessment feedback questionnaire scores, the number of data points completed, and the time taken to complete the assessment. It was also predicted that a native smartphone application would (1) yield a greater number of data points, (2) take less time, and (3) be more positively appraised by patient participant users than the text-based system.

**Methods:** A randomized repeated measures crossover design was employed. Participants with currently treated Diagnostic and Statistical Manual (Fourth Edition) schizophrenia or related disorders (n=24) were randomly allocated to completing 6 days of assessment (four sets of questions per day) with a native smartphone application or the SMS text-only implementation. There was then a 1-week break before completing a further 6 days with the alternative delivery modality. Quantitative feedback questionnaires were administered at the end of each period of sampling.

**Results:** A greater proportion of data points were completed with the native smartphone application in comparison to the SMS text-only implementation (β = -.25, SE=.11, *P*=.02), which also took significantly less time to complete (β =.78, SE= .09, *P*<.001). Although there were no significant differences in participants' quantitative feedback for the two delivery modalities, most participants reported preferring the native smartphone application (67%; n=16) and found it easier to use (71%; n=16). 33% of participants reported that they would be willing to complete mobile phone assessment for 5 weeks or longer.

**Conclusions:** Native smartphone applications and SMS text are both valuable methods of delivering real-time assessment in individuals with schizophrenia. However, a more streamlined graphical user interface may lead to better compliance and shorter entry times. Further research is needed to test the efficacy of this technology within clinical services, to assess validity over longer periods of time and when delivered on patients' own phones.

***(J Med Internet Res 2013;15(4):e60)*** doi:10.2196/jmir.2328

**KEYWORDS**

mobile phone; psychosis; assessment; schizophrenia; text-messages

## Introduction

Schizophrenia is a major public health problem affecting approximately 1% of the general population. It represents both a substantive emotional burden for those involved and a socioeconomic burden in the United Kingdom [1]. There is now increasing emphasis on treating patients within the community [2]. However, symptoms, mood, and functioning can change suddenly, potentially leading to relapse, unscheduled acute care, or self-injury. Monitoring of psychotic symptoms usually relies on interview assessments infrequently conducted by clinical staff, with limited resources, reducing the capacity to detect sudden change. Additionally, interviews relying on retrospective accounts introduces bias and averaging, thereby losing clinical information [3]. A more sensitive approach would be to closely monitor patients during the course of their everyday lives. Real-time assessment of psychotic symptoms is attractive in that it can lead to early and immediate intervention, potentially preventing the deterioration of an individual's mental state.

Over the past decade, assessment technologies have been increasingly employed in the monitoring of psychosis [4]. Typically, patients complete self-report questions at several different times of the day in real-world settings [5]. Early research in this field focused on the use of Personal Digital Assistants (PDAs) [6,7], but these technologies are becoming increasingly obsolescent and now occupy a very small share of the commercial market. Mobile phones have the advantage that they allow the automatic and wireless uploading of information to a central computer. As mobile phone technology becomes increasingly widespread, software applications can operate from users' own phones, omitting the need for them to carry an additional device. In a recent study, we observed that 83% (n=30/36) of a sample with psychosis currently owned a mobile phone, suggesting that usage in individuals with psychosis is comparable to the general population in the United Kingdom. Other studies have also reported high levels of familiarity with and access to mobile phone technology in individuals with serious mental illness [8].

There are multiple ways in which mobile phone-based assessment can be employed in health care. Text messages have the advantage that they are relatively easy and inexpensive to deliver and are not limited by the model or make of an individual's phone [9]. Individuals will likely be familiar with sending and receiving text messages. To date, two studies have used text messaging in the clinical care of psychotic illness. Španiel and colleagues [10] used weekly text-based assessments in order to alert consulting psychiatrists to early signs of relapse in patients with psychosis, which significantly reduced the number of inpatient admissions over one year. Additionally, Granholm and colleagues [11] have used text messages to promote cognitive behavioral therapy techniques and health-promoting behaviors. Pilot data showed a significant increase in social interaction and a reduction in hallucinations, but no significant effect on medication use.

Native software applications, with graphical user interfaces, uploaded onto smartphone devices pose another viable option for conducting real-time assessment. For the purpose of this research, a smartphone will be defined as having (1) computing power, (2) a touch screen, (3) third party application development and distribution, and (4) the option of 3G connectivity and high-speed data transfer. Native smartphone applications can be purpose built allowing for greater flexibility and ease of use [9] and are now increasingly being used in the treatment of physical health problems (eg, diabetes) [12]. Our research group has recently developed one such software application for use on Android smartphones. This software enables users to respond to self-report questions relating to their symptoms on touch-screen analogue scales. Initial piloting of this technology showed little dropout and high levels of data points completed in 44 individuals experiencing psychosis, sampled for 1 week. 13 self-assessment scales representing different dimensions of psychotic illness were validated against commonly employed, gold-standard clinical interviews. However, further information is required to assess the benefits and limitations of native smartphone applications in comparison to other methods of mobile phone-based data collection.

The objective of this study was to compare two different delivery modalities of the same diagnostic assessment for individuals with nonaffective psychosis—a native smartphone application employing a graphical, touch user interface against an short message service (SMS) text-only implementation. It was thought that both approaches would have certain advantages in the real-time assessment of psychosis but that the native smartphone application would display greater usability and functionality. Understanding the appropriate technology, user interface, procedures, and methods for administering mobile

phone–based assessment is an important step in making it both cost- and clinically effective, with the eventual aim of integrating it into long-term illness management.

The overall hypothesis of the study was that patient participants with serious mental illness would find both the smartphone application and SMS text-only implementation feasible and acceptable to use, measured by quantitative postassessment feedback questionnaire scores, the number of data points completed, and the time taken to complete the assessment. We also had three more specific predictions: (1) that the participants would complete a significantly greater number of entries on a native smartphone application when compared to a SMS text-only implementation, (2) that entries would take significantly longer to complete on a SMS text-only implementation, and (3) that the smartphone application would be appraised more favorably in quantitative feedback scores. We also asked participants about the maximum length of time that individuals thought they would be willing to complete assessment over for each delivery modality.

## Methods

### Participants

24 community-based patients meeting or having met the criteria for a Diagnostic and Statistical Manual (Fourth Edition) diagnosis of schizophrenia (n=22) or schizoaffective disorder (n=2) as made by the clinical service and checked against DSM-IV criteria by the researcher were admitted to the study. All participants were aged 18-50 and provided informed consent to take part. Participants were required to own and have access to a mobile phone with which to use the SMS text-only implementation. Organic or substance induced psychosis were exclusion criteria.

The sample was predominantly male (79%; n=19) and white British (71%; n=17) with a mean age of 33.0 (SD 9.5, min=18, max=49). Patients were recruited through Community Mental Health Teams (63%; n=15), Early Intervention Services (33%; n=8) and through a supported-living organization (4%; n=1). The mean number of past hospital admissions in the sample was 2.3 (SD 2.5, min = 0, max = 9) and the majority of the sample were currently taking antipsychotic medication (n=20). The Positive and Negative Syndrome Scale (PANSS [13]) is taken as a gold standard-clinical measure commonly used to assess psychosis. PANSS scores in this sample (at baseline) varied considerably between individuals, suggesting a range of symptom severities (positive subscale: mean 15.0 (SD 4.5, min = 7, max = 28); negative subscale: mean 12.3 (SD 3.5, min = 7, max = 21); general subscale: mean 28.8 (SD 6.8, min = 18, max = 50).

### Equipment

The two different modalities for delivering real-time assessment, namely a smartphone application and an SMS text-only implementation, were designed such that protocol for the monitoring procedure was common to each. The core functionality for the protocol that is common to both modalities is shown in Table 1.

From the end user's perspective, the two modalities differ only in the way the participant interacts with the device. These differences are summarized in Table 2.

The native smartphone application was specifically developed for Android mobile phones (Figures 1 and 2). Android is an operating system created by Google, which runs on mobile phones from different manufacturers. Although this software was developed to run on any Android phone, for this study we used the Orange San Francisco device. For the purpose of this study, it was not wirelessly enabled and all answers were stored on the mobile phone handset for downloading at the end of the sampling procedure. However, this made no difference to the user's perspective of the software.

The SMS text-only implementation was driven by openCDMS [14], an open source, secure online platform, which facilitated both the sending of questions and the storing of responses (Figure 3). openCDMS is a Web-enabled server-side application written in Java developed for electronic data collection in clinical studies and trials. It had pre-existing capabilities for SMS text messaging, individual participant records, and study questionnaires. Where new functionality was needed for the text message system, it was specifically developed.

### Semistructured Interviews

The PANSS [13] was conducted by an experienced administrator before and after each period of sampling. The PANSS is a semistructured interview assessing positive (7 items), negative (7 items), and global (16 items) symptoms of psychosis and has been extensively validated [13,15]. In this study, the PANSS was employed to provide convergent validity to the adapted mobile phone assessment depression items and to determine which form of delusions should be assessed.

### Quantitative Feedback Questionnaire

A purpose-designed quantitative feedback questionnaire was developed to assess the acceptability and feasibility of the native smartphone application and SMS text-only implementations (see Tables 3 and 4). This included reactivity to the methodology and whether it had been successfully integrated into an individual's everyday routine. Three items were taken from a previous study assessing the use of PDAs in individuals with psychosis [16]. These were "Overall, this was stressful", "Overall, this was challenging", and "Overall, this was pleasing". All items were rated from 1 (Not at all) to 7 (Very much so). In order to gauge the feasibility of the two methods of mobile phone assessment in the long-term management of patient's symptoms, participants were also asked to the maximum length of time within which they hypothetically would be willing to complete questions with each delivery modality. At the end of the study, participants were also asked which delivery modality they found easier to use and which they preferred (smartphone application, SMS text-only implementation, or no preference).

**Table 1.** Core functionality common to both the native smartphone application and text-message systems.

| | |
|---|---|
| Configurable number and times of question sets each day | At the start of the study, this can be configured for the desired number of questions sets and the times of these questions. During this study, 4 question sets per day were used. |
| Configurable questions | The wording of the questions is set up at the start of the study. It is easily configurable to support multiple studies with different questions. In addition, delusion questions are configured at the point when the researcher meets with the participant, through an administrative dialogue. |
| Multiple question sets | Multiple sets of questions are supported, and the software will switch between these at each consecutive alarm. So, if there are 2 question sets, it will ask set 1 at the first alarm, set 2 at the second, set 1 at the third, etc. The only exception to this is if the participant fails to answer some of the questions. In that case, the same question set is presented again at the next alarm. |
| Question branching | The next question displayed to the user can depend on the answer to one or more previous questions. This allows the questions asked to match the participant's symptoms or situation. For example, if a participant does not endorse the first question about a particular psychotic symptom, all remaining questions about it will be skipped. This means that the participant does not waste time answering unnecessary questions. |
| Questionnaire timeout | A time window is enforced, within which the questionnaire has to be completed. No further answers will be accepted outside of this time window. |
| Logging | The time at which each question was answered is recorded to the nearest second. This can be used to analyze the time taken to answer each question as well as the time to complete the whole questionnaire. |
| Branching logic | A range of different branching logic types is available. In the following list, the first three items apply to branching based on an answer to a single question. All other items are applicable when branching is based on answers to multiple questions.<br>• Less than<br>• Greater than<br>• Equal to<br>• Greater than sum<br>• Less than sum<br>• One is less<br>• One is greater<br>• All are less<br>• All are greater |

**Table 2.** Human-machine interface difference between native smartphone application and SMS text-only implementations of the common diagnostic assessment.

| | Native smartphone application | SMS text message |
|---|---|---|
| Alerts | Reuses Android's Alarm Manager so user definable alerts are available; delivered at semirandom intervals during the data collection period; users can snooze the alert to be reminded 5 minutes later. Only a single alert for each question set. | Alert is the phone's SMS alert, triggered when a SMS question is received; delivered at semi-random intervals during the data collection period. As each question is delivered as an SMS text message, an alert is triggered for each question. A reminder SMS is sent after 5 minutes if no response is received. |
| Questions | Presented as one question per page in the application. User is able to navigate through the pages of questions. | Delivered as SMS text messages one question per SMS. User must respond with an SMS text message and wait for the next question in the set. |
| Data input | Continuous slider bar, user slides with finger touching screen. Position of slider mapped to 7-point Likert scale. | User enters number between 1 and 7 as a response. |
| Saving data | Fully automatic, no user input required | User must send SMS text message containing the response value in reply to the question to record their response. |

**Figure 1.** A typical question from the Android app implementation, showing the full screen with the question and the slider for data entry.

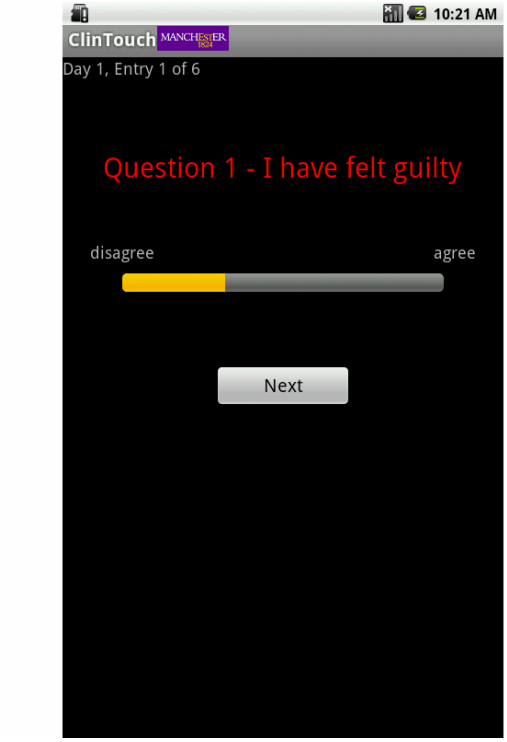



**Figure 2.** The start screen shown to the user in the Android app at the start of each set of questions (from which the users may proceed or delay ["snooze"] for a further 10 minutes they wish).

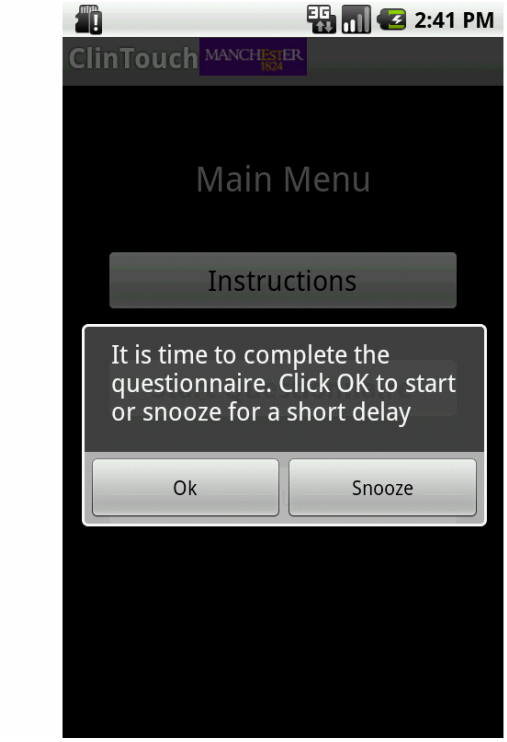
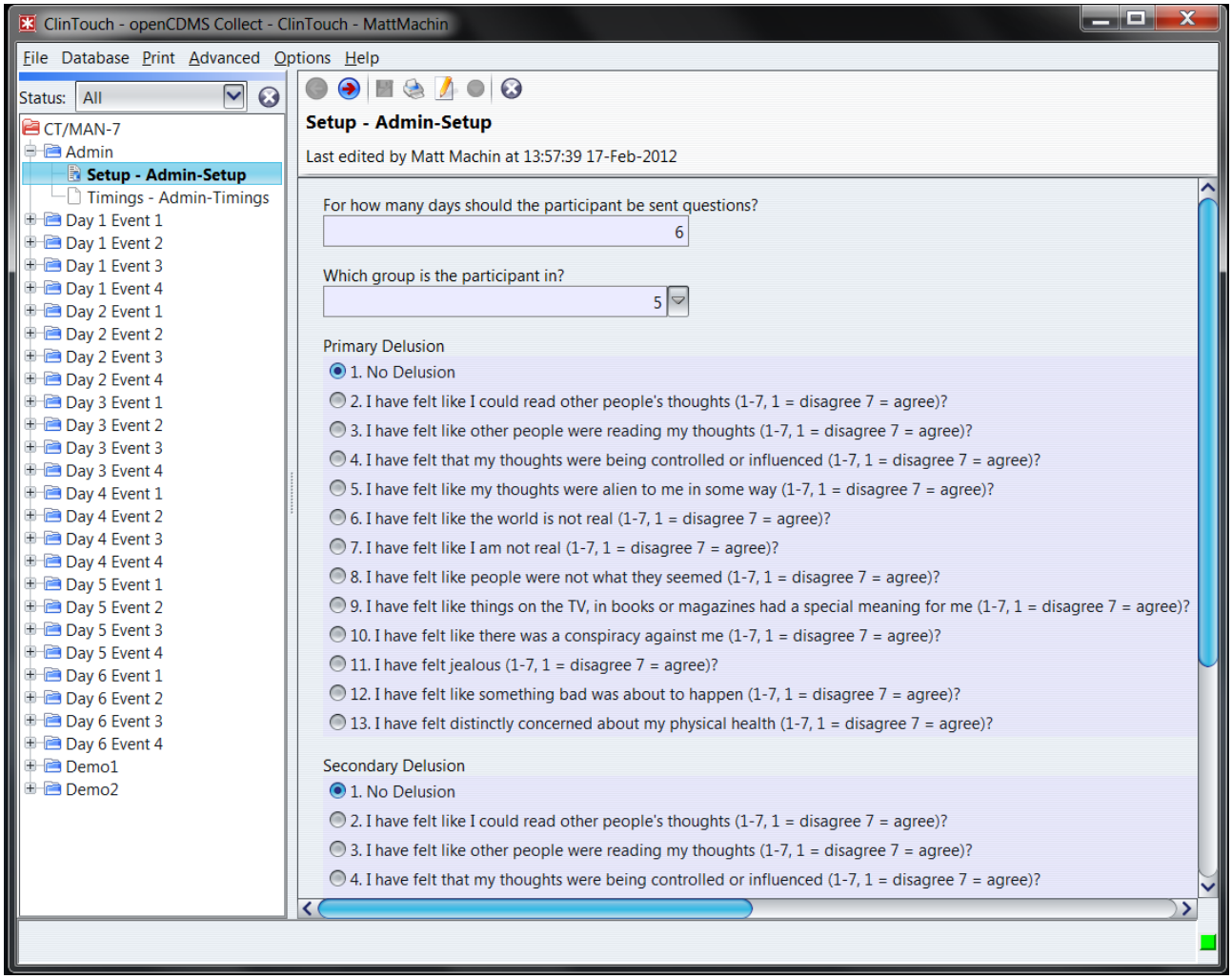


**Figure 3.** SMS configuration screen implemented in openCDMS.




## Diagnostic Assessment Items

Seven different symptom dimensions were assessed using previously validated mobile phone assessment scales. In order to minimize the burden on participants, these were split into two alternating sets. In set one, hopelessness (2-4 items), depression (2-6 items), and hallucinations (2-8 items) were assessed; and in set two, anxiety (1-4 items), grandiosity (2-3 items), paranoia (3-6 items), and delusions (0-8 items) were examined. Therefore, although there were four assessment points, each set of questions was assessed only twice per day. All questions related to the period of time that had elapsed since the last entry. Items were branched so that the questions changed depending on an individual's previous responses.

The depression subscale was adapted from the original validation study in an effort to increase its association with the PANSS depression scale, since the correlation was moderately low. The new subscale consisted of the items: "I have felt miserable" (new), "I have had no interest in seeing other people" (new), "I have felt worthless" (new), "I have felt sad", "My mood has affected my appetite or sleep", and "I have had thoughts about harming myself". As a result, the correlation between the mobile phone assessment scale and the PANSS was increased from rho = 0.45 ($P$=.01) in the original research to rho = 0.56 ($P$=.01) in the current study (calculated for the week of sampling with the native smartphone application). A strong correlation indicates that the interview and mobile phone–based assessments are tapping into similar concepts. The adapted depression assessment scale also showed considerable variability across time (mean squared successive difference: 2.2 (SD 2.9); within participant standard deviation: 0.8 (SD 0.6), suggesting that it was sensitive to subtle fluctuations in mood.

A wide range of delusions have been reported in the psychosis literature [17,18]. Therefore, during the initial briefing session, the researcher selected which delusion items were presented to the participant through the administrative dialogue. The choice of items was based on consultation with clinical staff and an initial PANSS interview to assess symptoms. For those individuals experiencing three or greater delusions, those two with the greatest conviction and distress were entered. The frequency of the different delusions were: "I have felt like other people were reading my thoughts" (n=5), "I have felt that my thoughts were being controlled or influenced" (n=4), "I have felt like I could read other people's thoughts" (n=3), "I have felt like things on the TV, in books or magazines had a special meaning for me" (n=3), "I have felt like something bad was about to happen" (n=2), "I have felt distinctly concerned about my physical health" (n=2), and "I have felt like my thoughts were alien to me in some way" (n=1). The delusion items were

kept constant for each participant across the two conditions of the trial.

### Procedure

A randomized repeated measures crossover design was employed. Participants were randomly allocated to either completing 6 days of sampling using the native smartphone application implementation on a smartphone provided to them for the purpose of the study, or via SMS text-only implementation using their own phone. There was then a 7-day rest period in order to reduce carryover effects before the individual completed a further 6 days of sampling with the alternate delivery modality. Each participant, therefore, completed two periods of sampling: 1 week with the native smartphone application implementation and 1 week with the SMS text-only implementation. Randomization was achieved through the openCDMS software. openCDMS uses a permuted block randomizer, and in this case we used a minimum block size of 4 and maximum of 6.

The researcher initially met with participants to obtain written consent and demographic information, complete clinical interviews, provide training in the first delivery modality, and administer practice questions. The participant number and delusion items were set by the researcher on either a password-protected administrators' page (on the native smartphone application) or through the openCMDS website (on the SMS text-only implementation). The volume of the alarm prompts was also set according to the preference of the participant when using the native smartphone application implementation.

On each day of sampling, participants could complete a maximum of four sets of questions, available at pseudorandom times (selected by a random number generator at least 1 hour apart) when prompted by the mobile phone device between 09:00 and 21:00 hours. All participants had 15 minutes from the first alarm within which to complete the questions. A forced entry time was thought to prevent a self-selection bias (ie, answering the questions only when asymptomatic) [19]. The researcher telephoned once or twice (as per the participant's preference) during the week in order to encourage compliance, answer questions, and to ascertain any problems with the software. The researcher attempted to keep the number of calls made to participants balanced over the two conditions: smartphone: mean, 1.7 (SD 0.5); text-messages: mean, 1.7 (SD 0.6).

Upon completion of the first week of sampling, the researcher met with the participant to re-administer the clinical interviews and to gather quantitative feedback on the first device. This procedure was then repeated the following week with the alternate device. At the end of the study, participants rated which device they preferred and found easier to use. Qualitative interviews were also conducted (the results of which are provided in a separate manuscript).

Participants on pay-as-you-go tariffs received £50 worth of phone credit, which they topped up prior to the week of text-message questions. If the participant was on a contract tariff (ie, direct debit) but did not have unlimited free texts, then they were reimbursed £50 at the end of the study. All participants were reimbursed an additional £30 upon completion of both sampling procedures.

### Statistics

All analysis was conducted in Stata 10.0 [20]. First, it was important to determine whether there was an interaction between period and delivery modality allocation. Delivery modality order (smartphone application then SMS text-only implementation or SMS text-only implementation then smartphone application) was entered into regression analysis as a predictor of the total score for each of the three outcome measures (mean time taken for each entry, number of completed data points, or quantitative feedback score) summed across the two conditions. A Spearman correlation was used to assess the similarity between scores across the two conditions.

Subsequent analysis was performed in a nested (long) form of the data. This data structure violates the assumption of independent observations meaning that additional constraints need to be placed on the statistical models. Multilevel modelling (xtreg) was therefore used to investigate whether delivery modality (smartphone application or SMS text-only implementation) or time-point (week 1 or week 2) significantly predicted any of the outcome variables. The outcomes were the mean time taken to complete each entry, the number of completed data points, and quantitative feedback scores. As the highest level of clustering, participant number was entered as the random effect.

Bootstrapping was used to account for the use of non-normal distributed variables in all analysis. This has been suggested as an appropriate alternative when parametric assumptions are not met [21]. The variables were manually standardized in order to aid the interpretation of the results.

## Results

We were not able to keep a systematic record of how many individuals were approached by their care coordinators for participation and declined. However, it was the impression of clinical teams that the refusal rate was around 30%. Of the 38 individuals who were referred to the research team, 8 changed their mind, 3 were unable to be contacted, and 3 were deemed ineligible prior to consent. This provided a final sample of 24 individuals, all of whom completed the feedback assessments (ie, no participants withdrew from the study). One individual, however, did ask for the SMS text-only assessment to be ended 2 days early because she found it was making her ruminative.

### Testing for an Interaction Between Sampling Period and Method of Assessment

First, we assessed whether there was an interaction between the sampling period and device type. Regression analysis showed that the order of the two conditions did not significantly predict the total number of entries an individual completed ($\beta$ =.06, SE.22, $P$=0.78), nor the length of time it took to complete each entry ($\beta$ = -.06, SE.20, $P$=.78). However, starting with the SMS text-only implementation did show nonsignificantly increased negative appraisals of the devices ($\beta$ = -.35, SE.20, $P$=.08).

### Comparing Native Smartphone Application and SMS Text-only Implementation

The number of entries completed during sampling was strongly correlated between the two delivery modalities (rho = 0.61, *P*=.002). When controlling for order effect, participants completed significantly greater numbers of entries on the native smartphone application when compared to SMS text-only implementation (Tables 3 and 4). Past studies have defined compliance as completing at least one-third of all possible data -points [22]. Using this criterion, 88% (n=21) of individuals were compliant when using the native smartphone application, whereas 71% (n=19) were compliant when using SMS text-only implementation. Participants completed a significantly greater number of entries in week one (mean = 16.4, SD 16.4) when compared to week two of sampling (mean = 13.7, SD 6.5; β = -.22, SE=.11, *P*=.04). When broken down by day, participants completed a mean of 3.4 (SD 0.8) entries on day 1, 3.8 (SD 0.5) on day 2, 3.0 (SD 1.2) on day 3, 3.4 (SD 1.1) on day 4, 3.0 (SD 1.3) on day 5, and 3.1 (SD 1.2) entries on day 6 when using the smartphone application. When using the SMS system this was lower at 2.3 (SD 1.4) entries on day 1, 2.4 (SD 1.3) entries on day 2, 2.6 (SD 1.5) entries on day 3, 2.6 (SD 1.3) entries on day 4, 2.0 (SD 1.4) entries on day 5, and 1.9 (SD 1.7) entries on day 6.

The length of time taken to complete each entry was highly correlated between the delivery modalities (rho = 0.62, *P*=.02), suggesting that those individuals who took longer on the native smartphone application also took longer to complete the SMS text-only implementation. However, as can be seen in Tables 3 and 4, individuals took an average 68.4 seconds (SD 39.5) to complete a full set of questions on the native smartphone application and 325.5 seconds (SD 145.6) on the SMS text-only implementation (β = 0.78, SE 0.09, *P*<.001). Thus, questions sets took 4.8 times longer on the SMS text-only implementation, than on native smartphone application. There was no significant difference in the length of time (seconds) it took participants to complete entries in week 1 (mean = 210.4, SD 171.3), when compared to week 2 (mean = 183.5, SD 166.3; β = -.08, SE=.09, *P*=.36) of sampling.

The total quantitative feedback scores for each device were also strongly correlated (rho = 0.48, *P*=0.02), suggesting that appraisals of the procedure were similar across both delivery modalities. No significant difference was observed for the total quantitative feedback score when comparing the two delivery modalities. Although none of the individual quantitative feedback items were significantly different, nonsignificantly higher scores (suggesting greater agreement) were observed for SMS text-only implementation on the items “Were there times where you had to stop doing something in order to answer the questions?”, ”Were there times when you felt like not answering?”, and “Was filling in the questions inconvenient?” (Table 3).

When considering an order effect, participants scored higher on the items “Were there times when you felt like not answering?” (week 1 mean 2.2, SD 1.1), week two mean 3.1, SD 2.2; β = 0.26, SE=.12, *P*=.03) and “Was it difficult to keep the device with you or carry it around?” (week 1 mean 1.8, SD 1.2; week 2 mean 2.6, SD 1.9; β = 0.26, SE=.12, *P*=.03) in the second, when compared to the first week of sampling. No other significant order effects were observed on the quantitative feedback scores.

Participants also reported the maximum length of time within which they would hypothetically be willing to complete questions on each delivery modality, which is displayed in Table 5. At the end of the final period of sampling, 67% (n=16) of individuals preferred the native smartphone application, 13% (n=3) of people preferred SMS text-only implementation, and 21% (n=5) of people had no preference. Additionally, 71% (n=17) of individuals found the native smartphone application easier to use, 17% (n=4) of people found SMS text-only implementation easier to use, and 13% (n=3) of people had no preference. Worthy of note is that 2 out of the 3 individuals who preferred their own phone currently owned a smartphone, which they used to complete the SMS text-only implementation.

**Table 3.** Quantitative feedback scores for the native smartphone application and SMS text-only implementation, and summary statistics.

| | Native smartphone application | | | | SMS text-only implementation | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Min | Max | Mean | SD | Min | Max | $\beta^P$ | SE |
| Time taken to complete questions (seconds) | 68.4 | 39.5 | 18.8 | 179.7 | 325.5 | 145.6 | 118.8 | 686.9 | 0.78$^{<.001}$ | 0.09 |
| Number of entries completed | 16.5 | 5.5 | 4.0 | 24.0 | 13.5 | 6.6 | 0.0 | 24.0 | -0.25$^{.02}$ | 0.11 |
| Did answering the questions take a lot of work? | 1.8 | 1.1 | 1 | 5 | 2.3 | 1.6 | 1 | 6 | 0.17 | 0.13 |
| Were there times when you felt like not answering? | 2.3 | 1.3 | 1 | 5 | 3.0 | 2.1 | 1 | 7 | 0.21$^{.07}$ | 0.12 |
| Did answering the questions take up a lot of time? | 1.7 | 0.9 | 1 | 4 | 2.3 | 1.6 | 1 | 7 | 0.24 | 0.14 |
| Were there times where you had to stop doing something in order to answer the questions? | 3.4 | 1.7 | 1 | 7 | 4.1 | 1.7 | 1 | 7 | 0.20$^{.10}$ | 0.12 |
| Was it difficult to keep track of what the questions were asking you? | 1.6 | 1.2 | 1 | 7 | 1.9 | 1.7 | 1 | 7 | 0.11 | 0.15 |
| Were you familiar with using this type of technology? | 4.7 | 2.3 | 1 | 7 | 5.3 | 2.2 | 1 | 7 | 0.14 | 0.14 |
| Was it difficult to keep the device with you or carry it around? | 1.9 | 1.4 | 1 | 6 | 2.4 | 1.8 | 1 | 6 | 0.16 | 0.12 |
| Did you ever lose or forget the device? | 1.7 | 0.9 | 1 | 4 | 1.8 | 1.4 | 1 | 6 | 0.06 | 0.13 |
| Was using the key pad/touch screen difficult to use? | 2.0 | 1.3 | 1 | 5 | 1.8 | 1.4 | 1 | 6 | -0.08 | 0.15 |
| Do you think other people would find the software easy to use? | 5.3 | 1.8 | 2 | 7 | 5.9 | 1.4 | 3 | 7 | 0.19 | 0.16 |
| Do you think you could make use of this approach in your everyday life? | 4.0 | 1.8 | 1 | 7 | 3.9 | 2.2 | 1 | 7 | -0.02 | 0.13 |
| Do you think that this approach could help you or other service users? | 5.3 | 1.9 | 1 | 7 | 5.6 | 1.2 | 3 | 7 | 0.11 | 0.15 |
| Overall, this experience was stressful. | 1.8 | 1.1 | 1 | 5 | 1.8 | 1.3 | 1 | 6 | -0.04 | 0.18 |
| Overall, this experience was challenging. | 2.2 | 1.6 | 1 | 7 | 2.7 | 1.7 | 1 | 6 | 0.16 | 0.13 |
| Overall, this experience was pleasing. | 3.7 | 2.0 | 1 | 7 | 3.7 | 1.7 | 1 | 7 | 0.01 | 0.11 |
| Did filling in the questions make you feel worse? | 1.8 | 1.1 | 1 | 5 | 2.1 | 1.4 | 1 | 5 | 0.14 | 0.15 |
| Did filling in the questions make you feel better? | 2.8 | 1.5 | 1 | 6 | 3.0 | 1.6 | 1 | 7 | 0.08 | 0.13 |
| Did you find the questions intrusive? | 2.2 | 1.2 | 1 | 4 | 2.6 | 1.8 | 1 | 7 | 0.15 | 0.14 |
| Was filling in the questions inconvenient? | 2.0 | 1.0 | 1 | 4 | 2.5 | 1.4 | 1 | 5 | 0.23$^{.09}$ | 0.14 |
| Did you enjoy filling in the questions? | 3.6 | 2.0 | 1 | 7 | 3.7 | 1.6 | 1 | 7 | 0.01 | 0.12 |
| Total quantitative feedback score (positive items reversed): | 53.0 | 11.2 | 33 | 76 | 56.2 | 14.2 | 27 | 88 | 0.13 | 0.10 |

**Table 4.** Quantitative feedback scores for the native smartphone application and SMS text-only implementation–momentary assessment symptom scores.

| Momentary assessment symptom scores | Native smartphone application | | | | SMS text-only implementation | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Min | Max | Mean | SD | Min | Max |
| Hallucinations | 2.7 | 1.8 | 1.0 | 6.5 | 2.5 | 1.8 | 1.0 | 6.0 |
| Anxiety | 2.8 | 2.0 | 1.0 | 6.3 | 2.1 | 1.4 | 1.0 | 6.4 |
| Grandiosity | 2.3 | 1.5 | 1.0 | 6.0 | 2.3 | 1.4 | 1.0 | 6.0 |
| Delusions | 2.0 | 1.3 | 1.0 | 5.1 | 1.9 | 1.1 | 1.0 | 4.7 |
| Paranoia | 2.9 | 1.8 | 1.1 | 6.5 | 2.6 | 1.7 | 1.0 | 7.0 |
| Hopelessness | 3.2 | 1.4 | 1.0 | 5.7 | 3.0 | 1.3 | 1.0 | 5.1 |

**Table 5.** Maximum length of time willing to complete questions on the two implementations.

| Maximum length of time willing to complete questions | Smartphone | Text messages |
|---|---|---|
| | n (%) | n (%) |
| U<2 weeks | 2 (8%) | 3 (13%) |
| 2-3 weeks | 10 (42%) | 10 (42%) |
| 3-4 weeks | 1 (4%) | 5 (21%) |
| 4-5 weeks | 3 (13%) | 1 (4%) |
| 5+ weeks | 8 (33%) | 5 (21%) |

## Discussion

The objective of this study was to compare two different delivery modalities of the same diagnostic assessment for individuals with nonaffective psychosis—a native smartphone application employing a graphical, touch user interface against an SMS text-only implementation. The overall hypothesis of the study was that participants with serious mental illness would find both systems feasible and acceptable to use, measured by the quantitative postassessment feedback questionnaire scores, the number of data points completed, and the time taken to complete the assessment. We also predicted that participants would complete a significantly greater number of data points, in less time, when using a native smartphone application when compared to the text messages and that the former software would be more positively appraised. The length of time that participants would hypothetically be willing to complete questions was also assessed in order to gauge the feasibility of longer-term assessment.

In line with the hypotheses, participants completed a significantly greater number of entries on the native smartphone application (69%, mean = 16.5), when compared to the text message interface (56%, mean = 13.5). It is desirable to keep missing data to a minimum in order to generate a representative picture of an individual's symptoms over time. Other studies have observed similarly high rates of compliance to the native smartphone application when using PDAs (ie, 69-72%) [7,23]. The increased usability and streamlined, graphical interface seen in purpose built software applications may encourage greater levels of compliance than those of SMS text-only implementations.

On average, sets of questions on the SMS text-only implementation took 4.8 times longer to complete when compared to the native smartphone application, which may have contributed to the reduced rates of compliance seen with this method of sampling. Although just failing to reach statistical significance, the quantitative feedback highlighted greater disruption to activities and inconvenience and less inclination to complete the questions when participants used the SMS text-only implementation. This is perhaps understandable given the greater time investment incurred by this method. Past research has suggested that the movement of mobile technology from the foreground to the background of an individual's life represents an important step in the accommodation of these technologies [24], which may be less likely if it perceived as burdensome.

Somewhat surprisingly, the two delivery modalities did not differ in any of the other quantitative feedback items or the total quantitative feedback score. The mean scores suggest that these appraisals were generally positive across both conditions. Thus, both forms of technology were deemed acceptable and well integrated and may represent suitable methods for facilitating real-time assessment. However, in the forced choice questions, the majority of the sample stated that they preferred the native smartphone application and that they found it easier to use, suggesting that this may be a more attractive delivery modality. Of the three individuals who preferred the SMS text-only implementation, two currently owned a smartphone, which they used for completing the text messages. These mobile phones may have allowed for a more streamlined text system (eg, threaded messages) reducing the differences between the two types of devices.

While the mean quantitative feedback scores suggested generally positive appraisals of both delivery modalities, these varied considerably between individuals. For example, some participants stated that they felt this technology could help them, whereas others were more skeptical about its advantages. Some participants reported mild negative reactivity to the method. It is possible that mobile phone assessment is less suitable in certain subgroups of patients, and in its future application, reactivity should be carefully monitored.

As this technology makes the transition from research to real-world clinical application, it will be vital to assess the feasibility and uptake of this software over longer periods of time and the factors influencing nonparticipation and withdrawal. In this study, participants completed fewer entries in the second week of sampling. Additionally, only a third of participants said they would be willing to complete the procedure for 5 or more weeks with the native smartphone application, with an even lower percentage for the SMS text-only implementation (21%). In the future, it may be necessary to employ machine learning in order to tailor the choice of questions and sampling rates to the service user [25], while placing particular emphasis on symptoms of primary concern to their clinical team. Person-tailored sampling could increase the feasibility of conducting longer-term real-time assessment. Automated and clinician-delivered feedback, or monetary incentives, may also promote acceptance and compliance.

In conclusion, this study provides data to suggest that both native smartphone applications and SMS text-only implementation represent acceptable technologies for facilitating

real-time assessment in individuals with nonaffective psychosis. However, the native smartphone application was found to be preferable to SMS text-only implementation in terms of greater data point completion and shorter response times. Limitations of this study include the relatively modest length of the sampling procedure and moderate rates of nonparticipation by those approached to take part. In the future, it will be important to upload software applications onto individual's own phone rather than issuing them with an additional device.

## Acknowledgments

This research was funded by the UK Medical Research Council's Developmental Pathways Funding Scheme and was supported by the Manchester Academic Health Science Centre (MAHSC). Thanks to all of the participants who took part in this study.

## Conflicts of Interest

Lewis has received speakers' honoraria from the pharmaceutical companies AstraZeneca and Janssen. Shitij Kapur has had grant support from the pharmaceutical industry. Emma Barkus has received funding from P1vital a Precompetitive Consortium. There are no other declarations of interest. Shitij Kapur received partial salary support via the National Institute of Health Research Biomedical Research Centre at the South London and Maudsley NHSFT. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health

## References

1. Mangalore R, Knapp M. Cost of schizophrenia in England. J Ment Health Policy Econ 2007 Mar;10(1):23-41. [Medline: 17417045]
2. Department of Health. No health without mental health: A cross-government mental health outcomes strategy for people of all ages - supporting document. 2011. URL: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_123766 [accessed 2013-03-06] [WebCite Cache ID 6EvTNj7uE]
3. Ben-Zeev D, McHugo GJ, Xie H, Dobbins K, Young MA. Comparing retrospective reports to real-time/real-place mobile assessments in individuals with schizophrenia and a nonclinical comparison group. Schizophr Bull 2012 May;38(3):396-404. [doi: 10.1093/schbul/sbr171] [Medline: 22302902]
4. Kimhy D, Myin-Germeys I, Palmier-Claus J, Swendsen J. Mobile assessment guide for research in schizophrenia and severe mental disorders. Schizophr Bull 2012 May;38(3):386-395. [doi: 10.1093/schbul/sbr186] [Medline: 22287280]
5. Csikszentmihalyi M, Larson R. Validity and reliability of the Experience-Sampling Method. J Nerv Ment Dis 1987 Sep;175(9):526-536. [Medline: 3655778]
6. Kimhy D, Corcoran C. Use of Palm computer as an adjunct to cognitive-behavioural therapy with an ultra-high-risk patient: a case report. Early Interv Psychiatry 2008 Nov;2(4):234-241 [FREE Full text] [doi: 10.1111/j.1751-7893.2008.00083.x] [Medline: 19884956]
7. Swendsen J, Ben-Zeev D, Granholm E. Real-time electronic ambulatory monitoring of substance use and symptom expression in schizophrenia. Am J Psychiatry 2011 Feb;168(2):202-209. [doi: 10.1176/appi.ajp.2010.10030463] [Medline: 21078705]
8. Ennis L, Rose D, Denis M, Pandit N, Wykes T. Can't surf, won't surf: the digital divide in mental health. J Ment Health 2012 Aug;21(4):395-403 [FREE Full text] [doi: 10.3109/09638237.2012.689437] [Medline: 22712756]
9. Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. J Biomed Inform 2012 Feb;45(1):184-198 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]
10. Spaniel F, Vohlídka P, Kozený J, Novák T, Hrdlicka J, Motlová L, et al. The Information Technology Aided Relapse Prevention Programme in Schizophrenia: an extension of a mirror-design follow-up. Int J Clin Pract 2008 Dec;62(12):1943-1946 [FREE Full text] [doi: 10.1111/j.1742-1241.2008.01903.x] [Medline: 18795967]
11. Granholm E, Ben-Zeev D, Link PC, Bradshaw KR, Holden JL. Mobile Assessment and Treatment for Schizophrenia (MATS): a pilot trial of an interactive text-messaging intervention for medication adherence, socialization, and auditory hallucinations. Schizophr Bull 2012 May;38(3):414-425. [doi: 10.1093/schbul/sbr155] [Medline: 22080492]
12. Mosa AS, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
13. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull 1987;13(2):261-276 [FREE Full text] [Medline: 3616518]
14. Ainsworth J, Harper R. The PsyGrid Experience: using web services in the study of schizophrenia. International Journal of Healthcare Information Systems and Informatics (IJHISI) 2007;2(2):1-20. [doi: 10.4018/jhisi.2007040101]
15. Bell M, Milstein R, Beam-Goulet J, Lysaker P, Cicchetti D. The Positive and Negative Syndrome Scale and the Brief Psychiatric Rating Scale. Reliability, comparability, and predictive validity. J Nerv Ment Dis 1992 Nov;180(11):723-728. [Medline: 1431824]
16. Kimhy D, Delespaul P, Corcoran C, Ahn H, Yale S, Malaspina D. Computerized experience sampling method (ESMc): assessing feasibility and validity among individuals with schizophrenia. J Psychiatr Res 2006 Apr;40(3):221-230 [FREE Full text] [doi: 10.1016/j.jpsychires.2005.09.007] [Medline: 16300791]

17. Appelbaum PS, Robbins PC, Roth LH. Dimensional approach to delusions: comparison across types and diagnoses. Am J Psychiatry 1999 Dec;156(12):1938-1943. [Medline: 10588408]
18. Peters ER, Joseph SA, Garety PA. Measurement of delusional ideation in the normal population: introducing the PDI (Peters et al. Delusions Inventory). Schizophr Bull 1999;25(3):553-576 [FREE Full text] [Medline: 10478789]
19. Palmier-Claus JE, Myin-Germeys I, Barkus E, Bentley L, Udachina A, Delespaul PA, et al. Experience sampling research in individuals with mental illness: reflections and guidance. Acta Psychiatr Scand 2011 Jan;123(1):12-20. [doi: 10.1111/j.1600-0447.2010.01596.x] [Medline: 20712828]
20. Statacorp. Programming Reference Manual, Release 10.0. College Stn, TX: Stata Corp; 2007.
21. Mooney CZ, Duval RD, Duval R. Bootstrapping: a nonparametric approach to statistical inference. Newbury Park, California: Sage Publications, Inc; 1993.
22. Myin-Germeys I, van Os J, Schwartz JE, Stone AA, Delespaul PA. Emotional reactivity to daily life stress in psychosis. Arch Gen Psychiatry 2001 Dec;58(12):1137-1144. [Medline: 11735842]
23. Granholm E, Loh C, Swendsen J. Feasibility and validity of computerized ecological momentary assessment in schizophrenia. Schizophr Bull 2008 May;34(3):507-514 [FREE Full text] [doi: 10.1093/schbul/sbm113] [Medline: 17932087]
24. Rogers A, Kirk S, Gately C, May CR, Finch T. Established users and the making of telecare work in long term condition management: implications for health policy. Soc Sci Med 2011 Apr;72(7):1077-1084. [doi: 10.1016/j.socscimed.2011.01.031] [Medline: 21397373]
25. Kelly J, Gooding P, Pratt D, Ainsworth J, Welford M, Tarrier N. Intelligent real-time therapy: harnessing the power of machine learning to optimise the delivery of momentary cognitive-behavioural interventions. J Ment Health 2012 Aug;21(4):404-414. [doi: 10.3109/09638237.2011.638001] [Medline: 22251028]

## Abbreviations

**PANSS:** Positive and Negative Syndrome Scale.
**PDA:** personal digital assistant
**SMS:** short message service

*Edited by G Eysenbach; submitted 28.08.12; peer-reviewed by D Kimhy, S Langrial; comments to author 08.11.12; revised version received 23.11.12; accepted 12.02.13; published 05.04.13.*

*Please cite as:*
*Ainsworth J, Palmier-Claus JE, Machin M, Barrowclough C, Dunn G, Rogers A, Buchan I, Barkus E, Kapur S, Wykes T, Hopkins RS, Lewis S*
*A Comparison of Two Delivery Modalities of a Mobile Phone-Based Assessment for Serious Mental Illness: Native Smartphone Application vs Text-Messaging Only Implementations*
*J Med Internet Res 2013;15(4):e60*
*URL: http://www.jmir.org/2013/4/e60/*
*doi:10.2196/jmir.2328*
*PMID:23563184*

Original Paper

# Adherence to a Smartphone Application for Weight Loss Compared to Website and Paper Diary: Pilot Randomized Controlled Trial

Michelle Clare Carter[1], MA, RD; Victoria Jane Burley[1], MSc, PhD; Camilla Nykjaer[1], MSc; Janet Elizabeth Cade[1], MSc, PhD

Nutritional Epidemiology Group, School of Food Science and Nutrition, University of Leeds, Leeds, United Kingdom

**Corresponding Author:**
Michelle Clare Carter, MA, RD
Nutritional Epidemiology Group
School of Food Science and Nutrition
University of Leeds
Food Science Building
Leeds, LS2 9JT
United Kingdom
Phone: 44 113 343 ext 8908
Fax: 44 1133434877
Email: m.carter@leeds.ac.uk

## Abstract

**Background:** There is growing interest in the use of information communication technologies to treat obesity. An intervention delivered by smartphone could be a convenient, potentially cost-effective, and wide-reaching weight management strategy. Although there have been studies of texting-based interventions and smartphone applications (apps) used as adjuncts to other treatments, there are currently no randomized controlled trials (RCT) of a stand-alone smartphone application for weight loss that focuses primarily on self-monitoring of diet and physical activity.

**Objective:** The aim of this pilot study was to collect acceptability and feasibility outcomes of a self-monitoring weight management intervention delivered by a smartphone app, compared to a website and paper diary.

**Methods:** A sample of 128 overweight volunteers were randomized to receive a weight management intervention delivered by smartphone app, website, or paper diary. The smartphone app intervention, My Meal Mate (MMM), was developed by the research team using an evidence-based behavioral approach. The app incorporates goal setting, self-monitoring of diet and activity, and feedback via weekly text message. The website group used an existing commercially available slimming website from a company called Weight Loss Resources who also provided the paper diaries. The comparator groups delivered a similar self-monitoring intervention to the app, but by different modes of delivery. Participants were recruited by email, intranet, newsletters, and posters from large local employers. Trial duration was 6 months. The intervention and comparator groups were self-directed with no ongoing human input from the research team. The only face-to-face components were at baseline enrollment and brief follow-up sessions at 6 weeks and 6 months to take anthropometric measures and administer questionnaires.

**Results:** Trial retention was 40/43 (93%) in the smartphone group, 19/42 (55%) in the website group, and 20/43 (53%) in the diary group at 6 months. Adherence was statistically significantly higher in the smartphone group with a mean of 92 days (SD 67) of dietary recording compared with 35 days (SD 44) in the website group and 29 days (SD 39) in the diary group ($P$<.001). Self-monitoring declined over time in all groups. In an intention-to-treat analysis using baseline observation carried forward for missing data, mean weight change at 6 months was -4.6 kg (95% CI –6.2 to –3.0) in the smartphone app group, –2.9 kg (95% CI –4.7 to –1.1) in the diary group, and –1.3 kg (95% CI –2.7 to 0.1) in the website group. BMI change at 6 months was –1.6 kg/m$^2$ (95% CI –2.2 to –1.1) in the smartphone group, –1.0 kg/m$^2$ (95% CI –1.6 to –0.4) in the diary group, and –0.5 kg/m$^2$ (95% CI –0.9 to 0.0) in the website group. Change in body fat was –1.3% (95% CI –1.7 to –0.8) in the smartphone group, –0.9% (95% CI –1.5 to –0.4) in the diary group, and –0.5% (95% CI –0.9 to 0.0) in the website group.

**Conclusions:** The MMM app is an acceptable and feasible weight loss intervention and a full RCT of this approach is warranted.

**Trial Registration:** ClinicalTrials.gov NCT01744535; http://clinicaltrials.gov/ct2/show/NCT01744535 (Archived by WebCite at http://www.webcitation.org/6FEtc3PVB)

*(J Med Internet Res 2013;15(4):e32)* doi:10.2196/jmir.2283

**KEYWORDS**

smartphone; obesity; text message; app

## Introduction

Obesity is estimated by the World Health Organization to be the fifth leading risk for global deaths [1], and it is associated with a range of serious and difficult to treat conditions, such as diabetes, some cancers, and heart disease. In the United Kingdom, obesity is a major public health concern reported to affect a quarter of the adult population [2]. The economic burden to the National Health Service (NHS) is significant, with the direct cost of spending on overweight and obesity estimated at £4.2 billion in 2007 [3]. The effective treatment of obesity and overweight is challenging and NHS primary care struggles to provide effective support to meet demand [4]. A large community-based survey showed that individuals desire alternatives to face-to-face weight loss treatments and, if given the choice, some would be interested in engaging with minimal contact weight management programs [5].

Weight management interventions based on information and communication technology (ICT) provide an opportunity to engage a wide audience in a potentially flexible and cost-effective way. In recent years, research into mobile devices to facilitate dietary and physical activity self-monitoring and weight-related behavior change has grown. Mobile phones, in particular, are an intuitively appealing intervention platform given that they are ubiquitous, engaging, and portable [6]. Researchers have investigated text message short message service (SMS) interventions to promote change in diet [7-9] and physical activity [10,11]. For example, in a small randomized controlled trial (RCT) (N=75) an SMS intervention lasting 16 weeks led to a mean weight loss of 2.1 kg (95% CI -3 to -1) in a group receiving daily text messages compared to 0.4 kg (95% CI -1 to 1) in the control group [7]. However, in a follow-up study, a larger 12-month RCT of an SMS intervention in 170 overweight and obese adults showed no statistically significant difference in weight loss between the intervention and control group [12]. In that study, adherence to the text-messaging intervention was found to be related to greater weight loss and the authors concluded that text messages might be a useful adjunct to a weight loss program. Researchers have also investigated dietary self-monitoring-based electronic interventions using personal digital assistants (PDAs), electronic portable devices that share some of the features of mobile phones. A 6-month RCT compared a PDA alone and a PDA with feedback to a paper diary in a sample of 210 overweight adults. The PDA group (combined with feedback) had the highest proportion of participants achieving greater than 5% weight loss (63% compared with 46% in the PDA alone and 49% in the food diary alone) after 6 months [13].

A number of smartphone applications (apps) that use the computational abilities of the phone for self-monitoring rather than just the SMS component have been developed (14-17) and tested. For example, a mobile app developed for a Nokia platform, Wellness Diary, allows users to record health-related data such as weight, sleep, and physical activity, and receive feedback on input [14-16]. The Patient-Centered Assessment and Counseling Mobile Energy Balance (PmEB) app allows users to log food intake from a limited database of foods and track calorie balance [18]. Another app, UbiFit, was developed to promote change in physical activity [19]. However, none of these apps have been formally evaluated in an RCT. In a recent 6-month RCT, 96 overweight and obese participants were randomized to either a group receiving podcasts only or an enhanced group using the podcasts, Twitter, and a smartphone app called FatSecret for self-monitoring [20]. In this study, the enhanced group were not found to have greater weight loss than the podcast-only group.

We developed a smartphone app for weight loss called My Meal Mate (MMM). The enhanced computational ability of a smartphone allows detailed self-monitoring (of diet, physical activity, and weight) and feedback via text message to be combined into 1 intervention. The MMM app uses an Android operating system so it can be trialed on an up-to-date and popular handset. The app has been benchmarked against commercially available systems, such as MyFitnessPal [21], and contains a large, detailed UK-branded food database [22]. These factors are important to engage users with the app in a real-life setting. Although there have been RCTs using text-messaging interventions for weight management, PDAs for self-monitoring and smartphone apps as adjuncts to other weight management interventions, to our knowledge there have not been any RCTs of a smartphone app as a weight loss intervention in itself using both self-monitoring and text-messaging functions. A trial of this type is necessary because smartphone apps are readily available to the public to download and likely to be used as a stand-alone intervention rather than as an adjunct to another intervention (such as podcasts or face-to-face advice). The aim of this pilot was to test the acceptability and feasibility (recruitment, dropout, and adherence) of MMM with a view to informing a larger trial.

## Methods

### Recruitment Strategy

Participants were recruited from large employers within Leeds, United Kingdom, by advertising through email, intranet, posters, and newsletters. Advertising material encouraged participants to contact the research team, following which they were emailed information sheets and an eligibility questionnaire. The eligibility criteria was a body mass index (BMI) of ≥27 kg/m$^2$; aged 18 to 65 years; willing to commit the necessary time and effort to the study; employed by a large employer in Leeds; not pregnant, breast-feeding, or planning a pregnancy; not taking antiobesity medication or medication/insulin for diabetes; never had surgery for weight loss; not taking the antidepressant sertraline (due to associations with weight gain); able to read and write in English; able to access the Internet; and willing to

be randomized to 1 of 3 groups. An inclusion cutoff BMI of ≥27 kg/m$^2$, as opposed to the more familiar cutoff point of 25 kg/m$^2$, was chosen to ensure that participants had a reasonable amount of weight to lose in 6 months before maintenance of weight loss and also as a safety measure so that they would be unlikely to lose so much weight that they fell below the defined healthy BMI range given that the app would be used for 6 months without any clinical supervision.

## Interventions

The researchers developed the MMM smartphone app for weight loss to be used on an Android operating system. Figures 1 and 2 are screenshots of the app. During development of MMM, several commercially available systems, such as MyFitnessPal and Calorie Counter, were informally evaluated by the researchers and by discussion in focus groups with potential system users. The MMM app was benchmarked in this way to produce an app of equivalent appearance and functionality as other apps available to the general public to download. Current UK evidence-based obesity guidelines advocate a lifestyle change approach to treatment [23]; therefore, in-line with this, the key behavioral strategies of goal setting, self-monitoring, and feedback underpin the MMM app. The MMM app allows users to set a weight loss goal and self-monitor daily calorie intake toward achieving that goal. Users select the food and drink consumed from a database and log items in an electronic food diary. Physical activity can also be recorded in the diary enabling the user to receive instant feedback on their energy expenditure. Progress is tracked graphically and further support is provided through tailored weekly text messages. A library of text messages was created and each message was triggered according to progress toward the users' calorie targets. The messages aimed to enhance the users' self-efficacy by encouraging the users to rehearse their weight loss goals and reinforce positive behavioral beliefs (about competence, confidence, and mastery). The MMM app has several usability features, such as the ability to take photographs of food to serve as a memory aid, and store favorite meal combinations and recently used items. The app has an associated Web interface to upload the recorded data. A unique feature of MMM is the large UK-specific branded food database which was provided by Weight Loss Resources, a commercial company [22]. The database contains 23,000 food and drink records that reflect both generic and branded items. The diet measures captured on MMM have been found to correlate well with a reference measure of diet [24]. There are a series of YouTube videos which give a detailed account of each feature of the MMM app that participants were able to directly link to for help [25].

The MMM app was compared against 2 other self-monitoring interventions to allow comparison of self-monitoring on a mobile phone against other approaches. The comparison groups used either a self-monitoring slimming website [22] or a food diary accompanied by a calorie-counting book [26]. The comparison interventions provided an opportunity to deliver a similar self-monitoring intervention by different mediums because each provides goal setting and self-monitoring by using the same Weight Loss Resources food database.

## Procedure

The trial design was a 3-armed parallel group randomized trial. As a pilot trial, the primary outcomes were feasibility and acceptability outcomes of adherence to the trial and adherence to the interventions (frequency of use). Secondary outcomes were anthropometric, which were objectively measured to give an idea of effect size. Eligible participants were invited to attend a baseline enrollment session at the University of Leeds where height, weight, and percentage body fat (BF) were measured by research assistants, and baseline questionnaires self-completed. The questionnaires were designed to collect information on demographics, technology usage, attitudes toward weight loss, physical activity [27], eating behavior [28], and a variety of psychosocial variables [29,30]. Weight (without shoes) and BF were measured by using Weight Watchers 8958U Body Analyser Scale portable weighing scales. Height (without shoes) was measured using a portable stadiometer to the nearest 0.1 cm. After measurements were taken, participants were randomized by a process of minimization using the Minim software package [31] to 1 of 3 groups. The minimization balanced equally at the medians on 3 factors: starting BMI, age, and gender. Minimization was used because this method has the advantage over simple or stratified randomization of providing very similar balanced groups in small samples [32].

After randomization, groups of participants were taken to separate rooms to receive standardized training in their allocated study equipment. Participants were instructed to use the study equipment every day for a week and then to use it as much as they desired over the trial period. The smartphone group were given a HTC Desire smartphone with the MMM app predownloaded, the website group were given a voucher providing 6 months access to the Weight Loss Resources website, and the food diary group were given a paper food diary, a calorie-counting book, and a calculator. All participants were given access to an Internet forum for social support. The baseline enrollment sessions took place over the month of June 2011 with participants enrolling in small groups at a time. Participants returned for repeat measures at 6 weeks and 6 months after randomization. Evaluation questionnaires were also administered at 6 weeks and 6 months. At 6 months, study equipment was returned. Because of the nature of the interventions, it was not possible to blind participants to their assignment. Fieldworkers carrying out measurements on participants were blinded to group assignment and participants were asked not to discuss their group allocation when measurements were taken.

## Sample Size Determination

This was not a phase III trial; therefore, a formal sample size calculation was not appropriate and there are few published guidelines on recommended sample sizes for pilot trials. The trial aimed to recruit a sample size of 135, which was a pragmatic decision based partly on the amount of available study equipment.

## Statistical Analysis

Statistical analysis was carried out using Stata Statistical Software Release 11 (StataCorp LP, College Station, TX, USA). Most analyses are descriptive because this was a pilot trial and

not powered to detect weight change. The effectiveness of the minimization procedure was assessed by determining baseline balance among the groups. When analyzing differences among the 3 intervention groups, 1-way analysis of variance (ANOVA) was used for continuous outcomes found to be normally distributed or the Kruskal-Wallis test when data were not normally distributed. For the analysis of completers versus noncompleters, *t* tests were used for continuous outcomes that were normally distributed and the Wilcoxon rank sum test for non-normally distributed outcomes. Differences among groups for categorical data were analyzed by using Chi-square tests.

This pilot trial was not statistically powered to detect change in anthropometric measures; however, results are displayed for interest and to provide information on effect size. A regression analysis was used to test between-group differences in change in anthropometric measures adjusting for the 3 factors used in randomization at baseline (age, gender, and starting BMI). Two analyses were conducted because there was a proportion of missing data and unequal dropout between groups: an intention-to-treat analysis in which all are included but using baseline weight carried forward for missing data, and an analysis in just those who completed 6-month follow-up.

## Ethical Approval

This trial was conducted according to the guidelines laid down in the Declaration of Helsinki and all procedures involving human subjects were approved by the University of Leeds, Faculty of Medicine and Health Research ethics committee (ethics reference no: HSLTLM/10/002). Written informed consent was obtained from all trial participants. In accordance with the requirements of the International Committee of Medical Journal Editors (ICMJE), this trial was registered (ClinicalTrials.gov NCT01744535) and reported in accordance with the CONSORT-EHEALTH checklist [33]. The version number of the app tested in the pilot trial was 1.0.23.

**Figure 1.** Screenshot of the My Meal Mate (MMM) homepage.

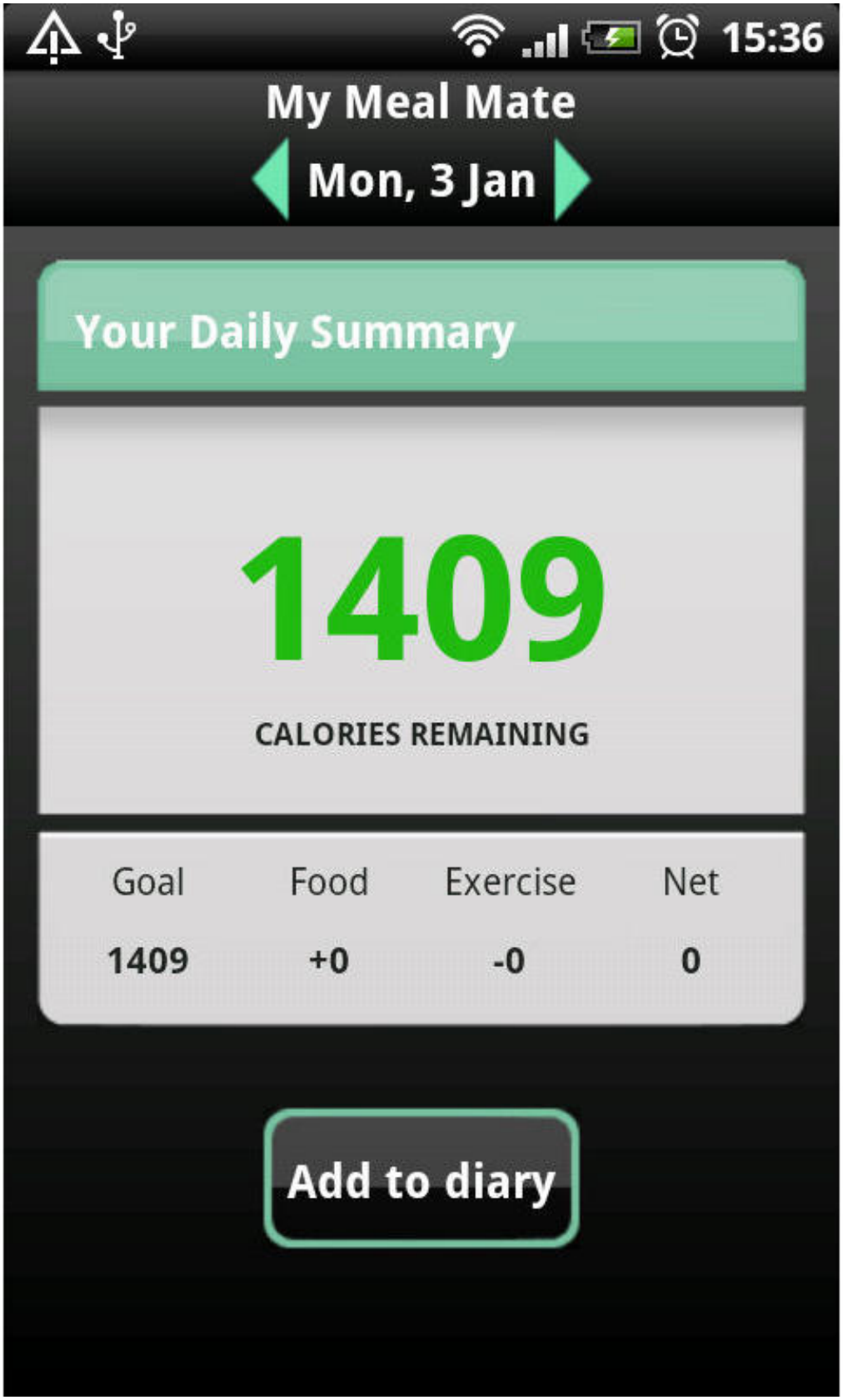
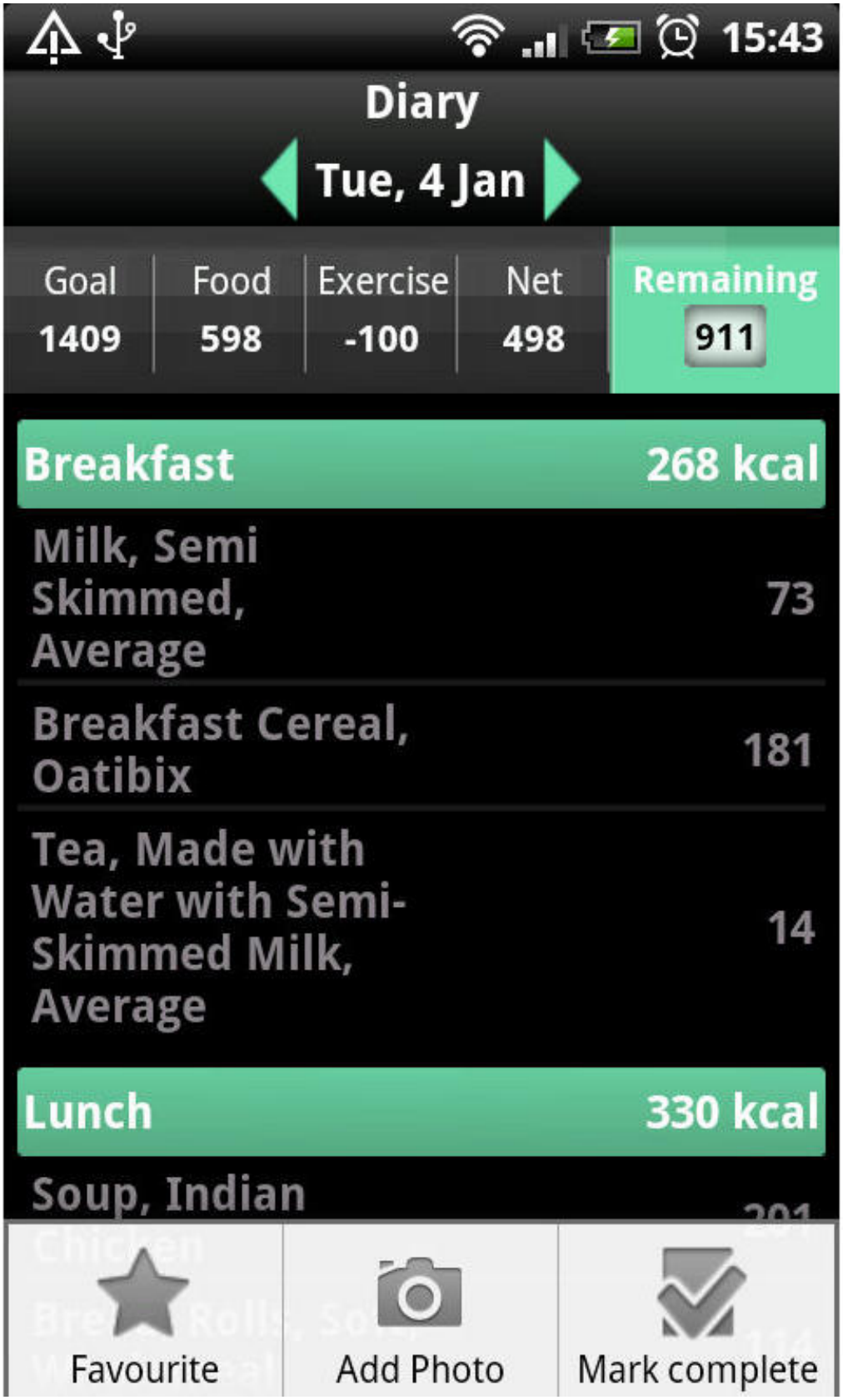


**Figure 2.** Screenshot of the My Meal Mate (MMM) food diary page.



## Results

### Baseline Characteristics

Table 1 shows the baseline characteristics of the pilot study participants by group. There were no statistically significant differences found among the 3 intervention groups for the factors balanced at minimization: gender (*P*=.97), age (*P*=.82), and BMI (*P*=.74). Of the 128 adults enrolled, 77% (99/128) were female and 91% (117/128) were white. The mean age of participants was 42 years (SD 9) and over half (58%, 74/127) were employed in managerial and professional occupations. The mean participant BMI was 34 kg/m$^2$ (SD 5) with over three-quarters of participants (77%, 98/128) classified as obese (BMI≥30 kg/m$^2$).

### Recruitment

Recruitment to the trial took 3 months between March and May 2011. Figure 3 is a CONSORT diagram [34] showing the participant flow through the trial. A total of 336 (73.5% female) people initially expressed an interest in taking part in the trial and 231 (68.8%) of these were assessed for eligibility to take part. The largest proportion (43.2%) of people who responded were from Leeds City Council, followed by the University of Leeds (27.4%). Of those screened, 49 (21.2%) were excluded for not meeting the inclusion criteria, and almost half (49.0%) were excluded because their self-reported BMI was less than 27 kg/m$^2$.

In total, 182 people met the eligibility criteria and were invited to a baseline appointment. Of those invited, 21 (11.5%) declined to participate, 13 (7.1%) did not respond, and 19 (10.4%) agreed to attend but did not show up to appointments. This left 129 people who attended baseline appointments. One person was excluded at baseline because their BMI was found to be below 27 kg/m$^2$. In total, 128 people were randomly allocated to 1 of the 3 groups. These 128 represented 38.1% of those who had originally expressed an interest in taking part, and 70.3% of those who had been invited to take part who met the eligibility criteria. With regard to sources of recruitment, the University of Leeds provided the most study participants (42.2%) and Leeds City Council provided the second highest proportion (39.0%). Most participants (83.6%) heard about the study from an electronic source, either by email (61.7%) or intranet (21.9%).

### Adherence to the Trial

In terms of trial retention, 94 (73.4%) people returned for 6-week follow-up measurements and 79 (61.7%) returned at 6 months. Table 2 shows the differences between those who completed 6-month follow-up compared to noncompleters. Compared to trial completers, noncompleters had a statistically significantly greater baseline BMI and BF. There was a statistically significant difference in self-reported health status at baseline between completers and noncompleters, with more completers reporting their health status as good or excellent (*P*=.001). Trial dropout was statistically significantly different among the groups (*P*=.001) with 3 people not attending 6-month follow-up in the smartphone group compared with 23 people not attending 6-month follow-up in the diary group and 23 people not attending 6-month follow-up in the website group. The reasons given for nonattendance are shown in Table 3. The most popular reasons given for nonattendance were dislike of the study equipment (n=12) and personal issues (n=6).

**Table 1.** Baseline characteristics of the participants enrolled in the 3 arms (smartphone application, website, or diary) of the 6-month pilot randomized controlled trial (N=128).

| | Smartphone (n=43) | Diary (n=43) | Website (n=42) |
|---|---|---|---|
| Age (years), mean (SD) | 41.2 (8.5) | 42.5 (8.3) | 41.9 (10.6) |
| Weight (kg), mean (SD) | 96.4 (16.0) | 97.9 (18.7) | 96.4 (19.9) |
| Body mass index (kg/m$^2$), mean (SD) | 33.7 (4.2) | 34.5 (5.7) | 34.5 (5.6) |
| Body fat (%), mean (SD) | 35.9 (3.8) | 35.9 (4.8) | 36.2 (3.9) |
| Gender (female), n (%) | 33 (76.7) | 33 (76.7) | 33 (78.6) |
| Race/ethnicity (white), n (%) | 43 (100.0) | 35 (83.3) | 39 (92.9) |
| Smoking status (current smokers), n (%) | 2 (4.8) | 8 (19.1) | 2 (4.8) |
| Occupation (managerial professions), n (%)[a] | 32 (74.4) | 22 (51.2) | 20 (48.8) |
| Has a university degree, n (%) | 31 (72.1) | 24 (55.8) | 22 (52.4) |
| Owns a smartphone, n (%) | 18 (41.9) | 19 (44.2) | 14 (34.2) |

[a] The occupation variable was dichotomized into managerial professions or not; it was originally measured as managerial and professional occupations, intermediate occupations, small employers and own account workers, lower supervisory and technical occupation, and semiroutine and routine occupations.

**Figure 3.** Flow of participants through a randomized, 3-armed, 6-month pilot trial of the My Meal Mate (MMM) smartphone application for weight loss (N=128).

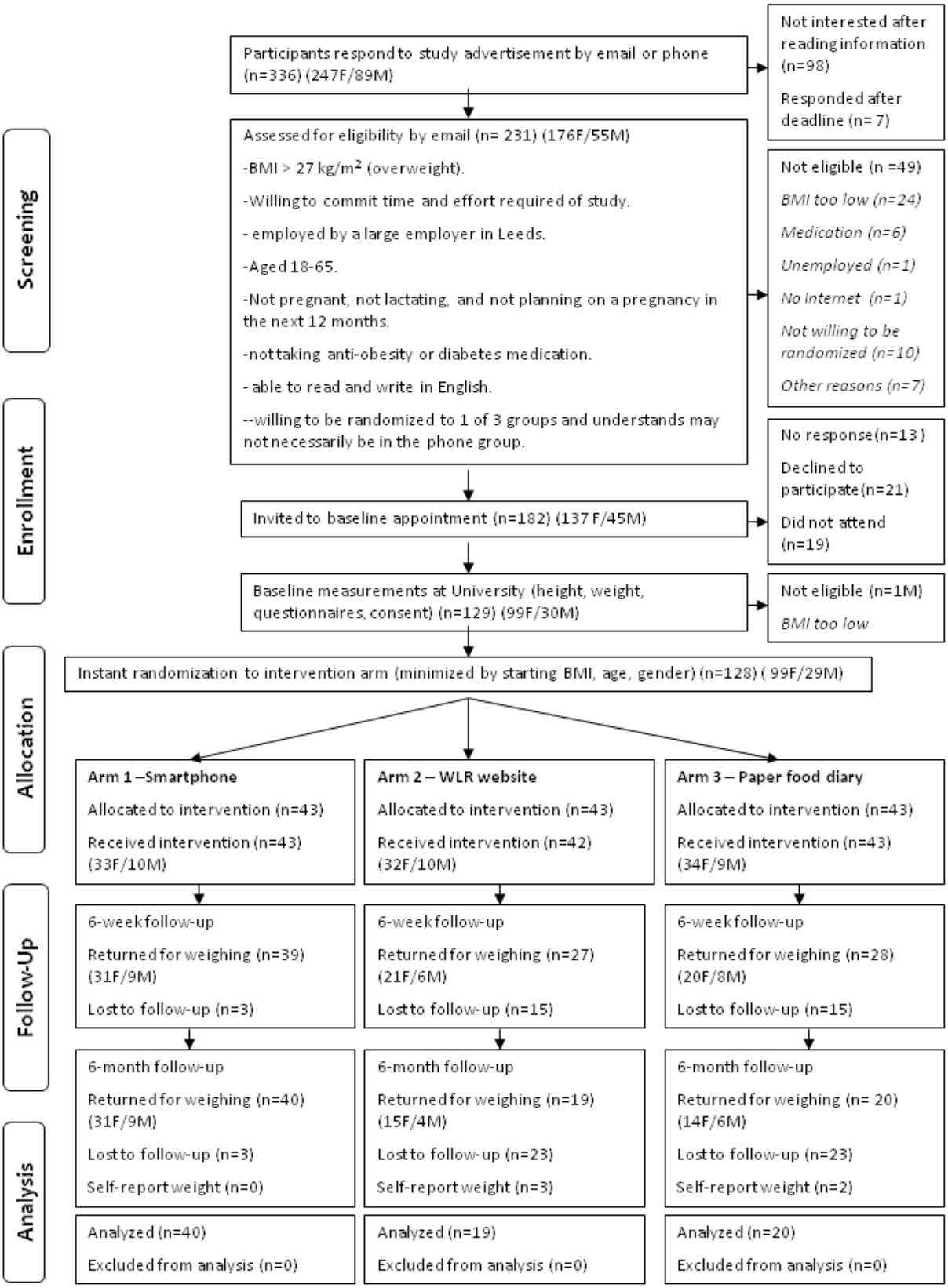


**Table 2.** Differences in baseline characteristics between trial completers and noncompleters at 6-month follow-up.

| Participant characteristics | Noncompleters (n=49) | Completers (n=79) | *P* [a] |
|---|---|---|---|
| Age (years), mean (SD) | 43.2 (9.0) | 41.2 (9.3) | .23 |
| Weight (kg), mean (SD) | 101.5 (18.9) | 94.1 (17.1) | .03 |
| BMI (kg/m$^2$), mean (SD) | 36.1 (5.8) | 33.1 (4.5) | .001 |
| Body fat (%), mean (SD) | 37.4 (4.2) | 35.3 (4.0) | .01 |
| Baseline physical activity (MET-min/week), mean (SD)[b] | 1468.8 (1207.9) | 1638.5 (1412.3) | .67 |
| Motivation to lose weight, mean (SD)[c] | 8 (2) | 8 (1) | .40 |
| Confidence in ability to lose weight, mean (SD)[d] | 7 (2) | 7 (2) | .74 |
| Number of previous weight loss attempts, mean (SD) | 11.9 (16.1) | 6.9 (7.9) | .11 |
| Consideration of future consequence score, mean (SD)[e] | 32.5 (8.5) | 30.7 (7.2) | .22 |
| Conscientiousness score, mean (SD)[f] | 79.6 (11.8) | 76.3 (11.3) | .15 |
| Female, n (%) | 38 (78) | 61 (77) | .96 |
| Obese (BMI≥30) (yes), n (%) | 39 (86.7) | 55 (69.6) | .05 |
| Race/ethnicity (white), n (%) | 39 (88.6) | 74 (93.7) | .44 |
| Smoking status (current smokers), n (%) | 7 (15.9) | 5 (6.4) | .09 |
| Occupation (managerial professions), n (%) | 21 (46.7) | 50 (64.1) | .06 |
| Reported health status as excellent or good, n (%) | 26 (59.1) | 68 (86.1) | .001 |
| Main shopper (yes), n (%) | 34 (75.6) | 66 (83.5) | .27 |
| Main preparer (yes), n (%) | 33 (73.3) | 58 (73.4) | .99 |
| Currently dieting (yes), n (%) | 31 (68.9) | 59 (74.7) | .69 |
| Constant dieter (yes), n (%) | 25 (56.8) | 36 (46.2) | .38 |
| Ever kept a food diary (yes), n (%) | 26 (57.8) | 47 (59.5) | .85 |

[a] Significant difference between completers and noncompleters assessed by 1-way *t* test.

[b] Measured by International Physical Activity Questionnaire (IPAQ); MET-min/week=metabolic equivalent of task (MET) level × minutes of activity × events per week.

[c] Based on a 10-point scale (1=not motivated at all; 10=extremely motivated).

[d] Based on a 10-point scale (1=not confident at all; 10=extremely confident).

[e] Measured by consideration of future consequences scale.

[f] Measured by International Personality Item Pool (IPIP) conscientiousness scale.

**Table 3.** Reasons for nonattendance at 6-month follow-up for trial noncompleters.

| Reason for nonattendance | Smartphone (n=43) | Diary (n=43) | Website (n=42) | Total (n=128) |
|---|---|---|---|---|
| Unable to contact to determine reason, n (%) | 0 (0) | 9 (20.9) | 9 (21.4) | 18 (14.1) |
| Did not like study equipment, n (%) | 3 (6.9) | 5 (11.6) | 4 (9.5) | 12 (9.4) |
| Holiday during follow-up, n (%) | 0 (0) | 2 (4.6) | 0 (0) | 2 (1.6) |
| Illness, n (%) | 0 (0) | 2 (4.6) | 2 (4.7) | 4 (3.1) |
| Personal issues, n (%) | 0 (0) | 3 (6.9) | 3 (7.1) | 6 (4.7) |
| Study too time-consuming, n (%) | 0 (0) | 1 (2.3) | 1 (2.4) | 2 (1.6) |
| Pregnancy, n (%) | 0 (0) | 0 (0) | 1 (2.4) | 1 (0.8) |
| Willing to self-report weight only, n (%) | 0 (0) | 1 (2.3) | 3 (7.1) | 4 (3.1) |
| Total | 3 | 23 | 23 | 49 |

## Frequency of Use of the Interventions

Table 4 shows the total number of days the interventions were used for each group at 6 weeks and 6 months (a complete day is considered as a day with ≥500 and ≤5000 kcal energy recorded). Intervention usage was highest in the smartphone group at 6 months with a mean of 92 days (SD 67) completed compared with 29 days (SD 39) in the diary group and 35 (SD 44) in the website group. There was found to be a statistically significant difference in the number of days usage among the groups at 6 weeks ($P$<.001) and 6 months ($P$=.001). Pairwise comparison showed that this difference lies between the smartphone group and the diary group ($P$<.001), between the smartphone group and the website group ($P$<.001), but not between the website group and the diary group ($P$=.14). At 6 months, 7 people had completed the smartphone electronic diary every day, no participants were found to have complete daily usage in the website and diary groups. Usage within each intervention arm declined over time as shown in Figure 4. In the smartphone group, 2 people recorded ≤7 days of food entry compared with 19 in the diary group (assuming 0 entries for 16 nonreturned diaries at 6 weeks) and 10 people in the website group. The median number of log-ins to the website over the 6-month period was 33 (interquartile range [IQR] 11-75). The frequency of website log-ins ranged from 2 to 375.

## Acceptability of Randomization and Satisfaction With Equipment

Of those who completed the 6-week questionnaires (n=93), 91.2% of smartphone participants reported that they were initially satisfied or very satisfied with their group allocation at baseline compared with 23.1% in the diary group and 71.4% in the website group ($P$=.01). When asked about how satisfied they were with the study equipment at 6 weeks, 86.8% reported that they were satisfied or very satisfied with the smartphone, compared with 57.7% in the diary group and 50.0% in the website group ($P$=.02). At 6 months, of those who completed questionnaires (n=77), 63.2% of smartphone participants were satisfied or very satisfied with the study equipment compared with 50.0% in the diary group and 42.1% in the website group ($P$=.05). At 6 months, 23 (30.0%) completers reported that they would not have volunteered for the trial if there had been no offer of using a smartphone.

No statistically significant difference was seen between the groups for self-reported ease of use of study equipment. In the smartphone group, 86.8% reported that they found their study equipment easy to use, compared with 65.0% in the diary group and 83.3% in the website group ($P$=.63). However, a statistically significant difference between the groups was found for self-reported convenience of use with 64.9% reporting that they found the smartphone convenient, (compared with 35.0% in the diary group and 52.6% in the website group, $P$<.001). In the smartphone group, 76.3% of participants agreed or strongly agreed that they felt comfortable using the study equipment to record their diet in social settings compared with 40.0% in the diary group and 21.1% in the website group ($P$<.001).

## Change in Anthropometric Measures

The pilot trial is not statistically powered to detect change in anthropometric measures; however, results are displayed to give an idea of effect size. As there is a proportion of missing data and unequal dropout, an intention-to-treat analysis was completed with baseline observations carried forward for missing data (Table 5). In the intention-to-treat analysis using all of the participants assigned to their original group, the mean weight change was –4.6 kg (95% CI –6.2 to –3.0) in the smartphone group, –2.9 kg (95% CI –4.7 to –1.1) in the diary group, and –1.3 kg (95% CI –2.7 to 0.1) in the website group. There was found to be a statistically significant difference in follow-up weight between the groups at 6 months ($P$=.004). At 6 months, weight change over time was statistically significantly greater in the smartphone group compared to the website group (–3.3 kg, 95% CI –5.4 to 1.2), but not when the smartphone group was compared to the diary group ($P$=.12).

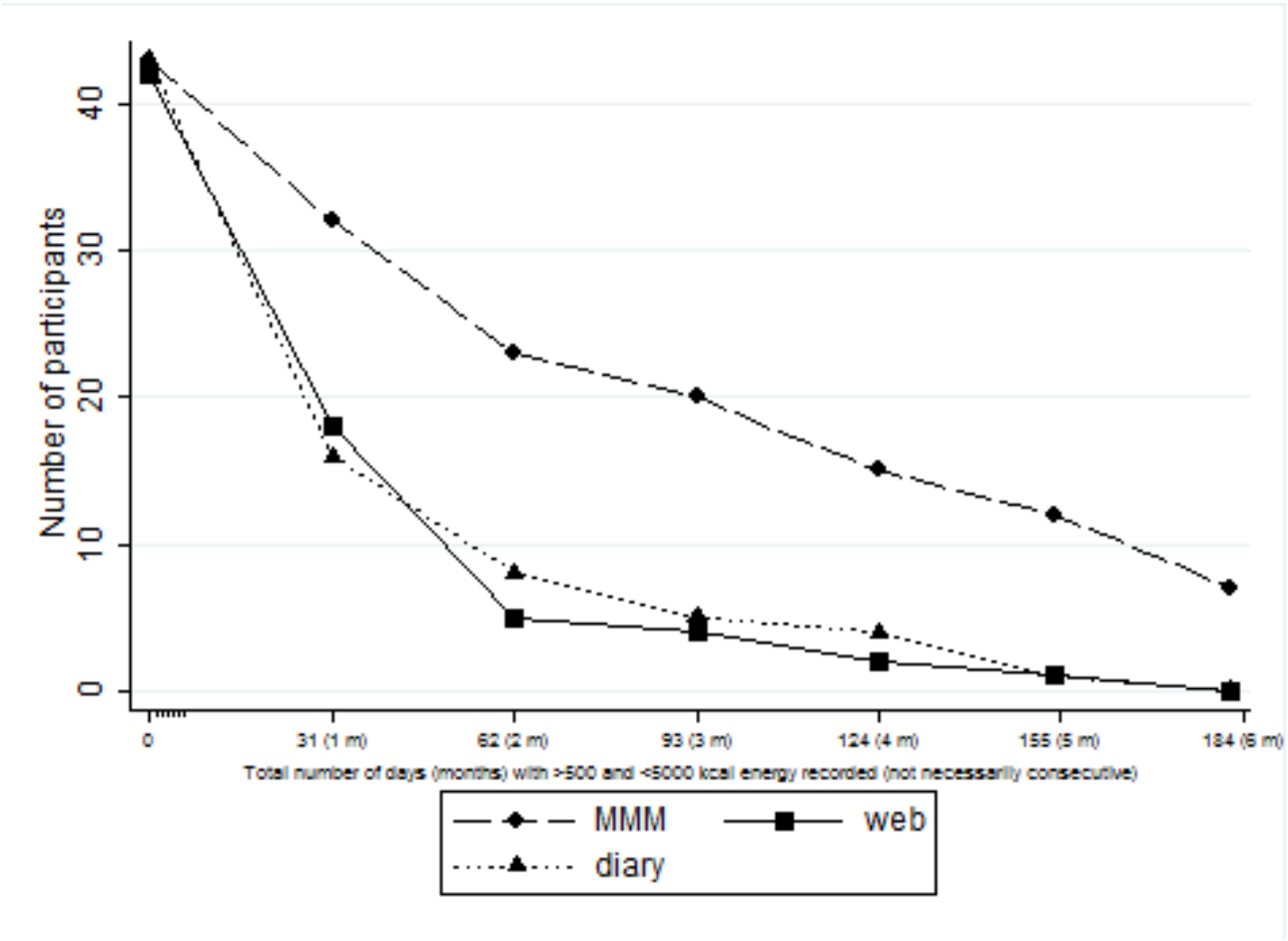
**Table 4.** Total number of days that the interventions were used (N=128).

| Intervention use | Smartphone (n=43) | Diary (n=43) | Website (n=42) | $P$ [a] |
|---|---|---|---|---|
| Total number of days intervention used[b] | | | | |
| 6 weeks (42 days), median (IQR) | 36 (21-42) | 29 (0-38) | 15 (6-33) | .004 |
| Completing every day, n (%) | 14 (33) | 8 (19) | 3 (7) | |
| 6 months (184 days), median (IQR) | 82 (28-172) | 18 (0-37) | 15 (7-45) | <.001 |
| Completing every day, n (%) | 7 (16 ) | 0 (0) | 0 (0) | |
| Completing 0 days/not returning paper diary, n (%) | 1 (2) | 31 (78) | 3(7) | |

[a] Significant difference between groups assessed by Kruskal-Wallis equality-of-populations rank test because adherence variable not normally distributed and not improved after log transformation; significant difference at $P$<.05.

[b] A usage day is considered to be a day with ≥500 and ≤5000 kcal energy recorded.

**Figure 4.** Intervention use in a randomized 3-arm pilot trial (N=128) of My Meal Mate (MMM). Adherence to the intervention arms (smartphone application, website, and diary) over the trial duration (6 months) is shown by total number of days completed in each intervention group. Data collection was conducted between May to December 2011 (total 184 days for each participant in the trial). A complete day is considered to be one with ≥500 kcal and ≤5000 kcal energy recorded. Intervention use is for overall total days completed and not necessarily consecutive days.




In the intention-to-treat analysis with baseline observation carried forward, change in BMI at 6 months was –1.6 kg/m$^2$ (95% CI –2.2 to –1.1) in the smartphone group, –1.0 kg/m$^2$ (95% CI –1.6 to –0.4) in the diary group, and –0.5 kg/m$^2$ (95% CI –0.9 to 0.0). Change in BF was –1.3% (95% CI –1.7 to –0.8) in the smartphone group, –0.9% (95% CI –1.5 to –0.4) in the diary group, and –0.5% (95% CI –0.9 to 0.0) in the website group.

Table 6 is a subanalysis that shows the anthropometric measures for study completers only (participants who attended follow-up at 6 months). In those who completed the trial, the mean weight change was –5.0 kg (95% CI –6.7 to –3.3) in the smartphone group, –6.2 kg (95% CI –9.8 to –2.7) in the diary group, and –2.8 kg (95% CI –5.9 to 0.2) in the website group. One person allocated to the diary group reported that they had actually used a commercially available weight loss smartphone app during the trial rather than the paper diary. This person lost 32 kg overall; if they are excluded from the diary group analysis, the mean weight change in completers is –4.8 kg (95% CI –7.1 to –2.7). There were not found to be statistically significant differences in follow-up weight between the groups at 6 months ($P$=.63) or in difference in change over time (smartphone-diary, $P$=.99; smartphone-website, $P$=.40; diary-website, $P$=.47). A similar trend in results was seen for BMI and BF.

Assuming baseline observation carried forward for those who did not return for follow-up at 6 months, 35/128 (27.3% of all participants randomized) achieved a clinically significant weight loss (≥5% of initial weight). This included 16/43 participants (37.2%) in the smartphone group, 12/43 (27.9%) participants in the diary group, and 7/42 (16.7%) participants in the website group.

**Table 5.** Change in anthropometric measures using an intention-to-treat[a] analysis.

| Anthropometric measurements | Smartphone | | Diary | | Website | | *P* [b] |
|---|---|---|---|---|---|---|---|
| | n | Mean (95% CI) | n | Mean (95% CI) | n | Mean (95% CI) | |
| **Weight (kg)** | | | | | | | |
| Baseline | 43 | 96.8 (91.9-101.8) | 43 | 97.9 (92.2-103.6) | 42 | 96.4 (90.2-102.6) | |
| 6 weeks | 43 | 93.9[c] (89.0-99.0) | 43 | 95.9[c] (89.8-101.7) | 42 | 95.1[c] (89.0-101.2) | .001 |
| 6 months | 43 | 92.2[c] (87.0-97.4) | 43 | 95.0[c] (89.0-101.0) | 42 | 95.1 (89.0-101.3) | <.001 |
| **BMI (kg/m$^2$)** | | | | | | | |
| Baseline | 43 | 33.7 (32.4-35.0) | 43 | 34.5 (32.7-36.2) | 42 | 34.5 (32.7-36.2) | |
| 6 weeks | 43 | 32.6[c] (31.3-33.9) | 43 | 33.7[c] (31.9-35.5) | 42 | 34.0 (32.3-35.7) | <.001 |
| 6 months | 43 | 32.1[c] (30.7-33.5) | 43 | 33.4 (31.5-35.4) | 42 | 34.0 (32.3-35.8) | <.001 |
| **Body fat (%)** | | | | | | | |
| Baseline | 42 | 35.9 (34.7-37.1) | 42 | 36.0 (34.5-37.5) | 42 | 36.3 (35.1-37.5) | |
| 6 weeks | 42 | 35.0[c] (33.7-36.2) | 42 | 35.3[c] (33.8-36.9) | 42 | 36.0 (34.7-37.2) | .01 |
| 6 months | 42 | 34.7[c] (33.5-35.9) | 42 | 35.1 (33.4-36.7) | 42 | 35.9 (34.5-37.2) | .02 |

[a] The baseline measures recorded have been carried forward for missing data.

[b] Significant difference between baseline and 6-week and 6-month follow-up assessed by paired *t* test. The regression analyses for difference in endpoints between the groups adjusts for starting weight and the 3 covariates randomized on at baseline (age, baseline BMI, and gender).

[c] Statistically significant *P* value of <.01.

**Table 6.** Change in recorded anthropometric measures at 6 weeks (n=95) and 6 months (n=79) for trial completers.

| Anthropometric measurements | Smartphone | | Diary | | Website | | *P* [a] |
|---|---|---|---|---|---|---|---|
| | n | Mean (95% CI) | n | Mean (95% CI) | n | Mean (95% CI) | |
| **Weight (kg)** | | | | | | | |
| Baseline | 40 | 96.8 (91.9-101.8) | 20 | 97.9 (92.2-103.6) | 19 | 96.4 (90.2-102.6) | |
| 6 months | 40 | 92.1[b] (86.6-97.6) | 20 | 86.1[b] (78.1-94.2) | 19 | 87.0 (79.5-94.6) | .62 |
| **BMI (kg/m$^2$)** | | | | | | | |
| Baseline | 40 | 33.7 (32.4-35.0) | 20 | 34.5 (32.7-36.2) | 19 | 34.5 (32.7-36.2) | |
| 6 months | 40 | 32.0[b] (30.5-33.5) | 20 | 30.4 (28.2-32.6) | 19 | 31.0 (28.9-33.2) | .58 |
| **Body fat (%)** | | | | | | | |
| Baseline | 39 | 35.9 (34.7-37.1) | 20 | 36.0 (34.5-37.5) | 19 | 36.3 (35.1-37.5) | |
| 6 months | 39 | 34.6[b] (33.4-35.9) | 20 | 32.5 (30.1-34.8) | 19 | 33.7 (31.7-35.8) | .89 |

[a] Significant difference between baseline and 6-week and 6-month follow-up assessed by paired *t* test. The regression analysis for difference in endpoints between the groups adjusts for starting weight and the 3 covariates randomized on at baseline (age, baseline BMI, and gender).

[b] Statistically significant *P* value of <.01.

## Discussion

This pilot trial has shown the MMM app to be a feasible and acceptable weight loss intervention.

### Recruitment and Response

In terms of recruitment and response, we were able to recruit 128 participants to the pilot, which was 95% of the original recruitment target. As is common to many weight loss trials, a large proportion of the sample (77%) were women and white (91%). The initial response rate was lower than expected and the recruitment period was extended to 3 months. Electronic media was the most successful recruitment strategy.

### Trial Retention

The pilot trial suffered from 38% attrition overall. Attrition is a serious difficulty in weight loss trials because of its potential to bias results [35]. Missing data may reflect a person's

dissatisfaction with the dietary intervention and a rebound in weight gain. To put this attrition figure into context, a systematic review of long-term weight loss trials in obese adults reported losses to follow-up in the range of 30% to 60% [36]. A review focusing specifically on Web-based interventions for weight loss found most had attrition rates greater than 20% [37]. In this trial, attrition was not equal among the groups, with more noncompleters at follow-up in the diary and Web groups compared to the smartphone group ($P$<.001). In fact, the smartphone group had extremely high retention with 93% returning for follow-up at 6 months (compared with 53% in the diary group and 55% in the website group).

Unequal dropout among groups is likely to be intervention-related [32], and a dislike of the study equipment was the most popular reason given for nonattendance at follow-up. Questionnaire data collected at follow-up also supports dissatisfaction with treatment group because at 6 weeks and 6 months satisfaction with group allocation was statistically significantly lower in the diary and Web groups. Unequal dropout is a potential source of bias in a large RCT so this will need to be considered for the full trial. Another explanation for differences in group retention may be that the smartphone group felt a greater sense of responsibility to the trial given that they had been provided with a costly piece of study equipment and had signed an agreement that they would return it. The diary and website group may have felt less obliged to return for follow-up because they did not need to physically return any equipment. This may be avoidable in a future study when it is likely that a large proportion of the population will own a smartphone (given the rising trend in smartphone ownership in the United Kingdom) so the app could be downloaded onto existing phones.

The noncompleters in the trial were more likely to have a higher BMI at baseline and report poorer health status. Other studies have shown mixed results with regard to attrition and initial body weight and a review of the behavioral approach to weight loss reports that both a higher and lower initial BMI have been linked to attrition in weight loss trials [38]. It may be that this minimal care approach is more acceptable to patients with a lower initial baseline BMI and a perception of good health, but interpretation should be cautious given the small sample size.

### Frequency of Usage of the Interventions

Adherence to dietary self-monitoring was found to be statistically significantly higher in the smartphone group than the website and paper diary group ($P$<.001). Participants were free to use the study equipment as often as they liked so the relatively high usage in the smartphone group is interesting. In all 3 groups, self-monitoring declined over time so that by 6 months only 7 participants (16% of the group) in the smartphone group had managed to record their dietary intake every day (no participants in the diary and Web group had done this). Adherence to self-monitoring is an important process outcome because it has been consistently linked to weight loss [39]. Researchers have taken different approaches to measuring adherence in studies investigating technology for weight loss so direct comparison of results is difficult.

A similar decline in adherence to dietary self-monitoring over time has been reported in other studies. In a recent RCT [13] comparing a PDA, a PDA with feedback, and a paper diary, 53% of the PDA group were adherent at 6 months compared with 60% of the PDA with feedback group and 31% of the paper diary group. Adherence was measured in that study as >50% of weekly calorie goal achieved so although the result is not directly comparable, the trend is similar. Also supporting the results of this pilot trial, the aforementioned study found that the PDA groups were statistically significantly more adherent to self-monitoring than the paper diary groups. However, in another study of dietary self-monitoring via PDA, no statistically significant difference in adherence was found between a PDA and a paper diary [40].

A key strength of this pilot is the use of a smartphone app for a high-end smartphone that is able to build on the research with PDAs (having similar self-monitoring functions) but is likely to be a more familiar technology to users. There has been a recent surge in smartphone ownership in the United Kingdom with 51% of the population reporting to own a smartphone [41]. It is evident that there is consumer demand for diet tracking apps due to the popularity of commercial systems such as MyFitnessPal [21] and Lose It! [42]. Investigating a researcher-developed app gives a unique opportunity to collect data on usage directly from the participants. In terms of acceptability, MMM was more highly rated in comparison to the diary and website on a range of acceptability measures, including overall satisfaction, convenience, and acceptability of use in social settings.

### Weight Loss

Although the pilot trial was not statistically powered to detect a difference in weight change among the groups, it has provided some data on effect size. Completers in the smartphone group had a mean weight loss of -5.0 kg (95% CI -6.7 to -3.3) after 6 months. This is comparable to the weight loss achieved in a large multicentered RCT of popular commercial diet programs that reported an average weight loss of -5.9 kg at 6 months across all diets [43]. The diary and website group had a comparable mean weight change at 6 months as those who returned for 6-month follow-up. When an intention-to-treat analysis is used with baseline observation carried forward for missing data, the mean weight change in the diary and Web groups is more modest.

### Strengths

This pilot trial has several strengths including its randomized design. Although researchers have investigated dietary self-monitoring as an adjunct or follow-up to a behavioral weight loss intervention [44] or used a smartphone app to enhance adherence to another intervention [45], this pilot trial has taken a minimal contact approach with no dietary advice at baseline. The weight loss seen in the smartphone arm is encouraging given that a minimal contact approach could be a cost-effective and wide-reaching strategy. This approach could also be especially beneficial to those who would prefer not to attend face-to-face meetings. Another strength of the trial is the up-to-date app for tracking diet and physical activity which is comparable in appearance and functionality to commercial diet

tracking apps. Despite their apparent popularity, these commercial apps have not been comprehensively evaluated to date.

### Limitations

Generalizability of the pilot results is limited given that the sample are predominantly white, female, and employed in managerial/professional occupations. The MMM app was a prototype app and participants reported that they frequently encountered bugs that caused the app to close. This may have affected participant engagement. Twenty people in the trial also reported that they had used another intervention (either instead of or in addition to their originally allocated intervention) during the trial. Seven participants from the smartphone group reported using a weight loss website, 7 people from the diary group reporting using a website, and 4 using a smartphone app and 2 from the website group reported to have used a smartphone app. One participant originally randomized to the diary group enjoyed self-monitoring but wanted to make it more convenient, so downloaded the commercially available MyFitnessPal app and used this for the duration of the trial. This person went on to lose 32 kg and had a strong influence on the mean weight change seen in the diary group. The degree of contamination seen in the trial is a serious issue and has implications for the design of a definitive RCT. In the pilot trial, participants knew what interventions were available in the trial and although they had all agreed to sign up with the understanding that they would be randomized to a group and not necessarily receive the intervention of their choice, it is a possibility that the trial raised their awareness of newer ICT-based methods of weight loss which they may not have already been aware of. In a definitive trial, the design would need to be altered to address contamination. A delayed control may be used so that participants in the control group could be asked not to use other weight management interventions during the trial and participants would be recruited in such a way that did not reveal what other groups were receiving.

### Implications

Further analysis will be performed on the pilot data to investigate the characteristics of successful users in the trial to see if there is any scope for tailoring this approach. Given that some participants have more success in behavioral weight loss programs than others [38], knowing who will do well with this smartphone approach is key to tailoring it appropriately. This pilot trial has several implications for a future trial. Given the unequal dropout in the comparator group, a larger trial may need to consider what if any retention strategies are appropriate. Two control groups were used in the pilot, but because participants had comparable adherence and weight loss in the diary group this may be the most cost-effective for a full trial. Further research would also benefit from an economic analysis to investigate the cost of implementing a smartphone app intervention compared with other types of weight management intervention.

### Conclusion

This pilot trial of a smartphone app for weight loss has shown that it is both an acceptable and feasible intervention. Adherence to the intervention and to the trial was greater in the smartphone group than the comparator groups and the app was rated highly in satisfaction and acceptability. To our knowledge, there have been no large RCTs of smartphone apps for weight loss and this pilot trial provides valuable data that could be used to inform such a trial.

## Acknowledgments

The study was funded by a National Prevention Research Initiative grant (grant number G0802108). MC was involved with the design and running of the study, data collection, analysis and interpretation of the data, and wrote the initial draft of the manuscript. JC and VB assisted in designing the study, interpretation of the data, supervision of the project, and preparation of the manuscript. CN helped with data collection and formatting of tables. Also, thank you to Mrs Claire Mcloughlin for clerical support and help with data collection and entry.

## Conflicts of Interest

The authors developed the My Meal Mate (MMM) application for the purposes of this trial by working with a software company (Blueberry Consultants). The University of Leeds owns full intellectual property of the MMM app. The researchers developed the application and objectively evaluated it, but have no commercial intent with the app, which is available for free download.

## Multimedia Appendix 1

CONSORT EHEALTH checklist V1.6.2 [33].

[PDF File (Adobe PDF File), 1MB - jmir_v15i4e32_app1.pdf ]

## References

1. World Health Organization. Global database on body mass index URL: http://apps.who.int/bmi/index.jsp?introPage=intro_3.html [accessed 2012-07-24] [WebCite Cache ID 69OQbEoVe]
2. The Information Centre for Health and Social Care. 2012 Feb 23. Statistics on Obesity, Physical Activity and Diet: England, 2012 URL: http://www.ic.nhs.uk/pubs/opad12 [accessed 2012-07-25] [WebCite Cache ID 69PoNojPy]

3. Butland B, Jebb S, Kopelman P, McPherson K, Thomas S, Mardell J. Foresight Tackling Obesities: Future Choices-Project Report. UK: Government Office for Science; 2007. URL: http://www.bis.gov.uk/assets/foresight/docs/obesity/17.pdf [accessed 2013-02-08] [WebCite Cache ID 6EHkWfoi2]
4. Jolly K, Lewis A, Beach J, Denley J, Adab P, Deeks JJ, et al. Comparison of range of commercial or primary care led weight reduction programmes with minimal intervention control for weight loss in obesity: lighten Up randomised controlled trial. BMJ 2011;343:d6500 [FREE Full text] [Medline: 22053315]
5. Sherwood NE, Morton N, Jeffery RW, French SA, Neumark-Sztainer D, Falkner NH. Consumer preferences in format and type of community-based weight control programs. Am J Health Promot 1998;13(1):12-18. [Medline: 10186930]
6. Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. J Biomed Inform 2012 Feb;45(1):184-198 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]
7. Patrick K, Raab F, Adams MA, Dillon L, Zabinski M, Rock CL, et al. A text message-based intervention for weight loss: randomized controlled trial. J Med Internet Res 2009;11(1):e1 [FREE Full text] [doi: 10.2196/jmir.1100] [Medline: 19141433]
8. Joo NS, Kim BT. Mobile phone short message service messaging for behaviour modification in a community-based weight control programme in Korea. J Telemed Telecare 2007;13(8):416-420. [doi: 10.1258/135763307783064331] [Medline: 18078554]
9. Haapala I, Barengo NC, Biggs S, Surakka L, Manninen P. Weight loss by mobile phone: a 1-year effectiveness study. Public Health Nutr 2009 Dec;12(12):2382-2391. [doi: 10.1017/S1368980009005230] [Medline: 19323865]
10. Fjeldsoe BS, Miller YD, Marshall AL. MobileMums: a randomized controlled trial of an SMS-based physical activity intervention. Ann Behav Med 2010 May;39(2):101-111. [doi: 10.1007/s12160-010-9170-z] [Medline: 20174902]
11. Hurling R, Catt M, Boni MD, Fairley BW, Hurst T, Murray P, et al. Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. J Med Internet Res 2007;9(2):e7 [FREE Full text] [doi: 10.2196/jmir.9.2.e7] [Medline: 17478409]
12. Shapiro JR, Koro T, Doran N, Thompson S, Sallis JF, Calfas K, et al. Text4Diet: a randomized controlled study using text messaging for weight loss behaviors. Prev Med 2012 Nov;55(5):412-417. [doi: 10.1016/j.ypmed.2012.08.011] [Medline: 22944150]
13. Burke LE, Conroy MB, Sereika SM, Elci OU, Styn MA, Acharya SD, et al. The effect of electronic self-monitoring on weight loss and dietary intake: a randomized behavioral weight loss trial. Obesity (Silver Spring) 2011 Feb;19(2):338-344 [FREE Full text] [doi: 10.1038/oby.2010.208] [Medline: 20847736]
14. Mattila E. Design and evaluation of a mobile phone diary for personal health management. Vuorimiehentie, Finland: VTT Technical Research Centre of Finland; 2010. URL: http://www.vtt.fi/inf/pdf/publications/2010/P742.pdf [accessed 2013-02-08] [WebCite Cache ID 6EHl7jYCB]
15. Mattila E, Korhonen I, Salminen JH, Ahtinen A, Koskinen E, Särelä A, et al. Empowering citizens for well-being and chronic disease management with wellness diary. IEEE Trans Inf Technol Biomed 2010 Mar;14(2):456-463. [doi: 10.1109/TITB.2009.2037751] [Medline: 20007055]
16. Mattila E, Lappalainen R, Pärkkä J, Salminen J, Korhonen I. Use of a mobile phone diary for observing weight management and related behaviours. J Telemed Telecare 2010;16(5):260-264. [doi: 10.1258/jtt.2009.091103] [Medline: 20483880]
17. Lee W, Chae YM, Kim S, Ho SH, Choi I. Evaluation of a mobile phone-based diet game for weight control. J Telemed Telecare 2010;16(5):270-275. [doi: 10.1258/jtt.2010.090913] [Medline: 20558620]
18. Tsai CC, Lee G, Raab F, Norman GJ, Griswold WG, Patrick K. Usability and feasibility of PmEB: a mobile phone application for monitoring real time caloric balance. 2006 Presented at: Pervasive Health Conference and Workshops, 2006; Nov 29 2006-Dec 1 2006; Innsbruk, Austria p. 1-10. [doi: 10.1109/PCTHEALTH.2006.361659]
19. Consolvo S, McDonald DW, Toscos T, Chen MY, Froehlich J, Harrison B, et al. Activity sensing in the wild: a field trial of ubifit garden. In: Proceedings of the twenty-sixth annual SIGCHI conference on human factors in computing systems. 2008 Presented at: Conference on Human Factors in Computer Systems; April 5-10, 2008; Florence, Italy p. 1797-1806.
20. Turner-McGrievy G, Tate D. Tweets, Apps, and Pods: Results of the 6-month Mobile Pounds Off Digitally (Mobile POD) randomized weight-loss intervention among adults. J Med Internet Res 2011;13(4):e120 [FREE Full text] [doi: 10.2196/jmir.1841] [Medline: 22186428]
21. MyFitnessPal. URL: http://www.myfitnesspal.com/ [accessed 2012-07-24] [WebCite Cache ID 69OQhV7f5]
22. Weight Loss Resources. URL: http://www.weightlossresources.co.uk/ [accessed 2012-07-24] [WebCite Cache ID 69OQlvXnX]
23. National Institute for Health and Clinical Excellence, National Collaborating Centre for Primary Care. Obesity: the prevention, identification, assessment and management of overweight and obesity in adults and children. 2006 Dec. URL: http://www.ncbi.nlm.nih.gov/books/NBK63696/pdf/TOC.pdf [accessed 2013-02-08] [WebCite Cache ID 6EHmUsiDu]
24. Carter MC, Burley VJ, Nykjaer C, Cade JE. 'My Meal Mate' (MMM): validation of the diet measures captured on a smartphone application to facilitate weight loss. Br J Nutr 2012 May 3:1-8. [doi: 10.1017/S0007114512001353] [Medline: 22717334]
25. YouTube. My Meal Mate videos URL: http://www.youtube.com/user/michcart1?feature=mhee [accessed 2012-10-19] [WebCite Cache ID 6BWpTZiR7]

26. Kellow J, Costain L, Beeken L. Calorie, Carb & Fat Bible. Peterborough, UK: Weight Loss Resources; 2011.
27. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
28. Stunkard AJ, Messick S. The three-factor eating questionnaire to measure dietary restraint, disinhibition and hunger. J Psychosom Res 1985;29(1):71-83. [Medline: 3981480]
29. Strathman A, Gleicher F, Boninger DS, Edwards CS. The consideration of future consequences: weighing immediate and distant outcomes of behavior. Journal of Personality and Social Psychology 1994 1994;66(4):742-752. [doi: 10.1037/0022-3514.66.4.742]
30. Goldberg LR, Johnson JA, Eber HW, Hogan R, Ashton MC, Cloninger CR, et al. The international personality item pool and the future of public-domain personality measures. Journal of Research in Personality 2006 Feb 2006;40(1):84-96. [doi: 10.1016/j.jrp.2005.08.007]
31. Evans S, Royston P, Day S. Minim: allocation by minimisation in clinical trials. URL: http://www-users.york.ac.uk/~mb55/guide/minim.htm [accessed 2013-02-08] [WebCite Cache ID 6EHn7EFx0]
32. Altman DG. Practical Statistics for Medical Research. London: Chapman and Hall; 1991.
33. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
34. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c332 [FREE Full text] [Medline: 20332509]
35. Ware JH. Interpreting incomplete data in studies of diet and weight loss. N Engl J Med 2003 May 22;348(21):2136-2137. [doi: 10.1056/NEJMe030054] [Medline: 12761370]
36. Douketis JD, Macie C, Thabane L, Williamson DF. Systematic review of long-term weight loss studies in obese adults: clinical significance and applicability to clinical practice. Int J Obes (Lond) 2005 Oct;29(10):1153-1167. [doi: 10.1038/sj.ijo.0802982] [Medline: 15997250]
37. Neve M, Morgan PJ, Jones PR, Collins CE. Effectiveness of web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: a systematic review with meta-analysis. Obes Rev 2010 Apr;11(4):306-321. [doi: 10.1111/j.1467-789X.2009.00646.x] [Medline: 19754633]
38. Wing RP, Phelan S. Behavioural treatment of obesity: strategies to improve outcome and predictors of success. In: Obesity: Mechanisms and Clinical Management. Philadelphia, PA: Lippincott Williams & Wilkins; 2003:415-435.
39. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
40. Yon BA, Johnson RK, Harvey-Berino J, Gold BC. The use of a personal digital assistant for dietary self-monitoring does not improve the validity of self-reports of energy intake. J Am Diet Assoc 2006 Aug;106(8):1256-1259. [doi: 10.1016/j.jada.2006.05.004] [Medline: 16863723]
41. Kantar Worldpanel. URL: http://www.kantarworldpanel.com/ [accessed 2012-07-24] [WebCite Cache ID 69OQpMxsu]
42. iTunes. Lose It! By FitNow URL: https://itunes.apple.com/ca/app/lose-it!/id297368629?mt=8 [accessed 2013-02-08] [WebCite Cache ID 6EHnubDpC]
43. Truby H, Baic S, deLooy A, Fox KR, Livingstone MB, Logan CM, et al. Randomised controlled trial of four commercial weight loss programmes in the UK: initial findings from the BBC "diet trials". BMJ 2006 Jun 3;332(7553):1309-1314 [FREE Full text] [doi: 10.1136/bmj.38833.411204.80] [Medline: 16720619]
44. Burke LE, Conroy MB, Sereika SM, Elci OU, Styn MA, Acharya SD, et al. The effect of electronic self-monitoring on weight loss and dietary intake: a randomized behavioral weight loss trial. Obesity (Silver Spring) 2011 Feb;19(2):338-344 [FREE Full text] [doi: 10.1038/oby.2010.208] [Medline: 20847736]
45. Turner-McGrievy GM, Campbell MK, Tate DF, Truesdale KP, Bowling JM, Crosby L. Pounds Off Digitally study: a randomized podcasting weight-loss intervention. Am J Prev Med 2009 Oct;37(4):263-269 [FREE Full text] [doi: 10.1016/j.amepre.2009.06.010] [Medline: 19765496]


## Abbreviations

**ANOVA:** analysis of variance
**BF:** body fat
**BMI:** body mass index
**ICT:** information and communication technology
**IPIP:** International Personality Item Pool
**MET:** metabolic equivalent of task
**MMM:** My Meal Mate
**NHS:** National Health Service
**PDA:** personal digital assistants

**PmEB:** Patient-Centered Assessment and Counseling Mobile Energy Balance
**RCT:** randomized controlled trial
**SMS:** short message service

*Edited by G Eysenbach; submitted 25.07.12; peer-reviewed by J Johnston, S Sethi, D Spruijt-Metz; comments to author 25.09.12; revised version received 23.10.12; accepted 02.01.13; published 15.04.13.*

*Please cite as:*
*Carter MC, Burley VJ, Nykjaer C, Cade JE*
*Adherence to a Smartphone Application for Weight Loss Compared to Website and Paper Diary: Pilot Randomized Controlled Trial*
*J Med Internet Res 2013;15(4):e32*
*URL: http://www.jmir.org/2013/4/e32/*
*doi:10.2196/jmir.2283*
*PMID:23587561*

Original Paper

# Opportunities and Challenges for Smartphone Applications in Supporting Health Behavior Change: Qualitative Study

Laura Dennison[1], PhD; Leanne Morrison[1], PhD; Gemma Conway[1], BSC(Hons); Lucy Yardley[1], PhD

Academic Unit of Psychology, University of Southampton, Southampton, United Kingdom

**Corresponding Author:**
Laura Dennison, PhD
Academic Unit of Psychology
University of Southampton
Shackleton Building, Highfield Campus, Highfield
Southampton, SO171BJ
United Kingdom
Phone: 44 02380 ext 597657
Fax: 44 02380 594597
Email: l.k.dennison@soton.ac.uk

## Abstract

**Background:** There is increasing interest from academics and clinicians in harnessing smartphone applications (apps) as a means of delivering behavioral interventions for health. Despite the growing availability of a range of health-related apps on the market, academic research on the development and evaluation of such apps is in the relatively early stages. A few existing studies have explored the views of various populations on using mobile phones for health-related issues and some studies are beginning to report user feedback on specific apps. However, there remains little in depth research on users' (and potential users') experiences and views on a wide range of features and technologies that apps are, or will soon be, capable of. In particular, research on young adults is lacking, which is an unfortunate omission considering that this group comprises of a good number of mobile technology adoptors.

**Objective:** The current study sought to explore young adults' perspectives on apps related to health behavior change. It sought their experiences and views of features that might support health behavior change and issues that contribute to interest in and willingness to use such apps.

**Methods:** Four focus groups were conducted with 19 students and staff at a University in the United Kingdom. Participants included 13 females and 6 males with a mean age of 23.79 (SD 7.89). The focus group discussions centred on participants' experiences of using smartphone apps to support a healthy lifestyle, and their interest in and feelings about features and capabilities of such apps. The focus groups were recorded, transcribed, and analyzed using inductive thematic analysis.

**Results:** Study findings suggested that young, currently healthy adults have some interest in apps that attempt to support health-related behavior change. Accuracy and legitimacy, security, effort required, and immediate effects on mood emerged as important influences on app usage. The ability to record and track behavior and goals and the ability to acquire advice and information "on the go" were valued. Context-sensing capabilities and social media features tended to be considered unnecessary and off-putting.

**Conclusions:** This study provided insight into the opportunities and challenges involved in delivering health-related behavioral interventions through smartphone apps. The findings suggested a number of valued features and characteristics that app developers may wish to consider when creating health behavior apps. Findings also highlighted several major challenges that appeared to need further consideration and research to ensure the development of effective and well-accepted behavior change apps.

***(J Med Internet Res 2013;15(4):e86)*** doi:10.2196/jmir.2583

**KEYWORDS**

mobile phone; cellular phone; behavior; health; qualitative research; focus groups

## Introduction

### Background

The smartphone market is growing rapidly with statistics from 2011 and 2012 suggesting that 35% of US and 39% of UK adults, respectively, use smartphones, with acceptance increasing quickly [1,2]. These ubiquitous devices are increasingly complex, computationally powerful, sensor-rich, and integrated with social networking [3-5]. Alongside these technological developments, there has been increasing interest from academics and clinicians in harnessing smartphones as a means of delivering behavioral interventions for health.

Various features of smartphones make them good candidates for the delivery of behavioral interventions. First, as portable devices that are highly valued by individuals, they tend to be switched on and remain with the owner throughout the day [3,5,6]. Therefore, they offer the opportunity to bring behavioral interventions into important real life contexts where people make decisions about their health and encounter barriers to behavior change [4,5,7]. Second, smartphone apps may provide cheaper, more convenient, or less stigmatizing interventions that are unavailable elsewhere [4,7]. Third, the connectedness of smartphones facilitates the sharing of behavioral and health data with health professionals or peers [4,6]. Furthermore, the increasing ability of smartphones to use internal sensors to infer context such as user location, movement, emotion, and social engagement (eg, [8-10]) has raised the prospect of continuous and automated tracking of health-related behaviors and timely, tailored interventions for specific contexts.

In addition to the existing body of research on telephone and SMS (short message service) text-messaging-delivered interventions, smartphone software programs, or applications (apps) have stimulated significant attention in recent years. For example, many thousands of commercial apps have already been developed to assist people with managing stress, improving mood, following a healthy diet, increasing physical activity, quitting smoking, and self-managing chronic health problems. Apps aiming to improve health tend to provide information, advice, instruction, prompts, support, encouragement, and interactive tools for individuals to monitor, record, and reflect.

Although there has been much enthusiasm for delivering interventions through smartphone apps, academic research on the development and evaluation of such apps is in the early stages. At this juncture, as well as initial pilot studies of specific individual apps, there is a need for formative research that helps us to better appreciate the interest various groups of people have in using these sorts of apps and factors that may influence acceptability and engagement.

### Previous Studies and Interventions

Some initial qualitative research on mobile phone use for health has already been conducted with various populations in several contexts including both healthy [11,12] and chronically ill [11] adults, and adults at risk of developing lifestyle-related health conditions [13]. Findings to date indicated that using mobile phones for managing physical and mental health and for supporting changes in health-related behavior is acceptable in principle to many people. A number of studies have focused more specifically on experiences of smartphone apps. Some such studies have been nested within pilot studies of research-developed apps. For instance, a feasibility study in 8 users of a mobile intervention for depression that included ecological momentary intervention and context sensing found that participants were satisfied with the intervention, despite encountering substantial technical problems [14]. Other studies have concentrated on understanding useful features of one or more specific apps. In one study, 15 participants downloaded 3 physical activity apps, which they used for one week each. They then provided quantitative ratings of the apps and qualitative feedback about their experiences. Users preferred flexible apps with automatic tracking of physical activity, tracking of goal progress, music features, and well-documented and easy to use features [15]. Other research, nested within an intervention trial, collected data on 119 users' experiences of 3 different apps (a wellness diary, a physical activity coach, and a relaxation assistant) and generated detailed findings relating to motivators and barriers to using each app [16].

Overall, the literature on smartphone app feasibility and acceptability is encouraging. However, in previous research, the exploration of user viewpoints has often been limited and fairly superficial. There is little in depth, qualitative research allowing users to describe their experiences, views, and usage patterns. Several important areas are inadequately addressed by previous research. First, there is an absence of research on how young adults perceive and use apps for behavior change. Few reports exist discussing the development of interventions for this population [17] despite the tendency for young adults to be early adoptors of smartphone technology [2,17]. Recent publications have highlighted the lack of focus on smartphone interventions for young adults and suggested that this is a priority population for a broad range of health behavior issues [17,18]. Second, tens of thousands of health apps are available directly to consumers through the market (eg, Apple App Store, Google Play) and a 2011 report showed that 11% of US adult smartphone users download health apps [19]. Reviews suggested that most direct-to-consumer apps are not developed by health professionals or academics, do not draw on behavior change theories or techniques, and do not have content aligned to clinical guidelines for the condition or behavior in question [20-24]. However, there is a surprising paucity of academic research on user views and experiences of these apps. Finally, there has been little exploration of people's views about context sensing by smartphones in the setting of health apps, although it has been recognized that this advancing technology may raise concerns, including worries around regarding privacy and security [4,25].

### Current Study

The current study sought to explore some of these under-researched areas. We conducted a series of focus groups with healthy young adult smartphone users in order to explore: (1) their existing experiences of using health-related smartphone apps, and (2) their views about a range of different features, technologies, and capabilities that characterize currently available or future apps. We sought their views on features that might support them in making changes to behaviors relevant to

health, factors that contribute to interest in apps, willingness to use the apps, and issues leading to disinclination to using the apps.

## Methods

### Participants

A total of 19 participants were recruited through advertisements placed around Southampton University campus. Eligible participants were 18 years of age or older, and owned a smartphone. Prior to arrival at the focus group, all participants completed a brief online questionnaire providing demographic, lifestyle, and smartphone-related data. Participants comprised of 13 females, 6 males, 16 students, and 3 junior members of staff. The age range was 18 to 50 (mean 23.79, SD 7.89) and 16/19 (84%) were White British. At the time of the study, 10 participants owned iPhones, 5 owned Android phones, and 4 owned Blackberries. All participants reported that they had previously used smartphone apps. The participants were relatively frequent users of their phones with all but two reporting at least one hour of use per day. Sixteen reported using non-call features (eg, apps, camera, Internet, multimedia) "often" or "very often" and only one used these "rarely". Fifteen participants indicated that they had previously sought health-related advice or information from the Internet or other forms of technology. Participants rated themselves as relatively "up to date with the latest technology" (minimum 3, mean 3.84, SD 6.02 on a 1-5 Likert scale). In terms of health and lifestyle, 7/19 (37%) participants reported themselves as "overweight", whereas 12/19 (63%) classified themselves as "about right". Around half (10/19, 53%) reported that they "lead an active lifestyle". Most (14/19, 74%) believed themselves to "eat a healthy diet".

### Procedure

Four focus groups were conducted (with 3, 5, 5, and 6 participants in each group). The session was guided by an interview schedule (Multimedia Appendix 1). The discussion began with questions around participants' personal experiences of smartphone apps for health. To prompt further discussion, particularly among participants with limited experience of relevant apps, the facilitator presented trigger materials. Images showing features and capabilities of health-related apps were displayed on a laptop computer using PowerPoint. This included examples of apps that: (1) assisted with setting goals and reviewing progress towards goals, (2) gave advice, tips, and information on health, (3) used tools to monitor behavior, mood, and well-being, (4) issued reminders and prompts for healthy behaviors, (5) allowed the sharing of progress through online social media, and (6) made use of context sensing to issue prompts or interventions (Multimedia Appendix 2). Participants were then asked to describe their thoughts and feelings about the features of these types of apps, including their perceived usefulness, relevance, and concerns. Participants themselves tended to guide the conversation, while the facilitator provided prompts to pursue more detailed discussions on some topics and to keep the discussions on track.

The focus groups were 55 to 70 minutes. Two researchers participated in each group—the first author (a post-doctoral researcher in health psychology) and the third author (a psychology student trained in focus group data collection methods). In each group, one researcher facilitated the discussion and presented the trigger materials, while the other assisted and took notes.

### Data Analysis

The focus groups were audio recorded and transcribed verbatim. The transcripts were then analyzed using inductive thematic analysis [26]. Following initial coding, where labels were attached to text segments which appeared to indicate important material in relation to the research questions, analysis progressed in an iterative fashion to develop a set of themes that captured the essence of the focus group discussions. The analyst compared the raw data with the emerging theme labels and definitions, and further refined the themes by merging, adding, and removing redundant themes.

## Results

### Overview

Many participants in this study were positive about the potential for phones to help people adopt healthier lifestyles. Most had already discovered and tried various apps to support healthy lifestyle practices. Many female participants had used calorie counting apps to help manage their weight. The male participants tended to describe using apps that supported physical fitness. Many participants expressed interest in trying the apps discussed within the focus groups. Despite appearing to be a health-conscious and enthusiastic group, the participants nonetheless perceived considerable barriers, concerns, and opportunities related to health apps. Table 1 summarizes the key themes identified from the focus groups. The following sections describe the themes with illustrative quotations. Pseudonyms have been used to preserve anonymity.

**Table 1.** Overview of themes.

| Theme | Summary of key points identified |
|---|---|
| Smartphones as valuable information sources | Smartphones are valued for providing a quick and efficient link to a wealth of information "on the go". |
| Tracking progress and raising awareness | Apps that support monitoring, tracking, and reviewing of behavior are interesting and potentially useful.<br>Tracking progress must be both accurate and require little effort.<br>Tracking features have the potential to trigger negative emotions. |
| Behavior change apps and social networks | A social undesirability of using health-related apps exists.<br>There is an aversion to sharing information related to health behavior via online social networks (except in circumscribed contexts). |
| Scepticism over context sensing | Context sensing can be perceived as "gimmicky", unreliable, and unnecessary.<br>Interventions triggered by context sensing could have negative consequences including irritation, poor mood, and more unhealthy behavior. |
| Useful prompts or harassment? | Prompts or reminders from apps can be useful, but can also become annoying and perceived as "nagging" |
| Motivation and necessity of behavior change | The appeal and usefulness of apps and their features is dependent on users' existing intentions and motivation to change.<br>Currently healthy young adults may perceive heath apps as of limited relevance to themselves and more suitable for people with "more serious" need for behavior change interventions (eg, people with existing health problems) |
| Necessity for efficiency and convenience | Smartphones are expected to be efficient and pleasurable to interact with.<br>Health apps are often perceived as time-consuming or burdensome and thus unlikely to be used for extended periods of time. |
| Disposability | There are plenty of free apps to choose from with little commitment to long-term use of any particular app.<br>Apps are easily downloaded and just as easily uninstalled. |
| Credibility and accuracy | Reputable and legitimate sources are considered important.<br>Safe and accurate app content is a concern to many. |
| Privacy and security concerns | Users may feel uneasy about whether apps keep health-related data secure and private.<br>Context sensing features create concerns about intrusion of privacy and personal safety. |
| Keeping control over apps | Users may feel uneasy about what apps could do without user awareness or permission.<br>Users desire to be aware of all features and in control of all app settings. |

## Smartphones as Valuable Information Sources

A commonly mentioned and uncontested use of smartphones for supporting the adoption of healthy behaviors was its ability to provide a wealth of information quickly and efficiently. Participants accessed information via the Internet regularly through their smartphones, including material relating to health. Participants described that they used their smartphones to look up potential symptoms to decide if they needed to consult a doctor, to search for healthy recipes to cook when at home, and to find advice on specific exercises they could do while at the gym. Some described a preference for multimedia information formats, for example video or audio of a fitness instructor providing guidance on an exercise. Many participants expressed interest in apps that gave information and advice that they could access "on the go".

## Tracking Progress and Raising Awareness

Participants liked apps that provided convenient tools to help them to monitor, track, and review attempts to change or improve a health behavior.

*Most people have their phones with them most of the time, so if you're out and about and want to check how much you're doing or what you're burning when you're walking, it's a good idea.* [Hannah, Focus Group (FG) 1]

Participants who had tried apps with these sorts of features found it interesting and insightful to set and record goals, track details of progress, and as a result notice trends and patterns.

> *It provides a lot of kind of... it helps you keep track. The exercises you do. Because I try and set myself certain targets, do say, a certain amount of running a week, a certain amount of cycling a week.* [Gareth, FG 4]

Some participants considered that simply viewing records of tracked behavioral data could prompt healthier behavior. Others felt that tools to monitor or track behavior alone may be insufficient to modify behavior in the absence of some advice or intervention to help them change.

> *If it gave you some hints to help you out or something along those lines. Only if it had a real benefit rather than just tracking.* [Leona, FG2]

Participants placed importance on accuracy with regard to data tracked and recorded by apps. They were concerned about errors within the app itself, or human error (forgetting, "fooling yourself", or "lying to yourself") leading to inaccurate records of things like food intake, physical activity, or alcohol

consumption. They were also keen that a high level of detail was preserved (eg, recording exact products/brands of foods consumed and their nutritional profiles). However, as described below, they considered entering such information burdensome and believed themselves likely either to forget to do this regularly or to give up due to boredom.

Another concern was that an app that tracks behavior and progress towards goals might reveal disappointing or upsetting results related to a behavior that could negatively affect mood, and be demotivating.

> *If someone is really trying to work hard on this and then it's telling them that they have not done very well, or that they have not reached their goals then it could go either way: it could motivate them or it could just make them feel like they're not achieving anything.* [Hannah, FG1]

Participants did not like the possibility that an app might explicitly "tell you off", and were also concerned about the emotional and motivational impact of simply viewing a log that revealed large discrepancies between goals and actual achievement.

> *You'd probably be like, "oh well I might as well give up then, it's too depressing".* [Amanda, FG3]

### Behavior Change Apps and Social Networks

Participants appeared to perceive decisions and actions related to a healthy lifestyle to be a private activity. In general, participants viewed health apps as slightly embarrassing and being seen using one seemed socially undesirable.

> *You might not want the whole world to know that as it flashes up on your phone.* [Aled, FG4]

> *If this popped up, I think people would laugh at me.* [Hannah, FG1]

One participant had previously used a weight management app that sent reports and updates to a nominated support person. Having chosen her boyfriend to take this role, she was later embarrassed and frustrated as he continues to receive notifications of what she is *not* doing and *not* achieving.

> *Believe me that was a mistake, never do that! Because he still gets emails saying "it's been 427 days since Izzy logged her calories on the program" and I'm not even using it!* [Isabella, FG1]

Considering apps that link and share information with social networks more generally, participants were very clear that they would not want to use app features that allow the unselective broadcasting of information about health-related goals or behaviors to friends via social media sites (eg, Facebook). Such sites were considered inappropriate places to share very personal information or to request help or support, and participants were keen to avoid presenting themselves as weak or vulnerable. Despite this aversion to apps that share data with social networks, a few specific instances where support could be harnessed in a more socially acceptable way were recognized and discussed. For example, as Isabella's boyfriend took up running, he posted about the distances he had run on Facebook.

> *He really liked doing that, to share with people, and he had loads of people "liking" it, and he felt that was quite inspiring.* [Isabella, FG1]

Katie and others agreed that simple reports of tangible achievements were more appropriate than broadcasting health-related goals or making overt requests for emotional support. Another context where social sharing was considered acceptable was within a group of people who were working towards similar goals (eg, people using the same health app or participating in the same face-to-face intervention program). Some considered introducing an element of competition through sharing progress with others with similar goals potentially useful. Others considered tips and advice from people with similar goals to be of potential value. Participants also suggested that connection with health professionals could be useful for apps that focus on monitoring and/or changing certain health conditions.

### Scepticism Over Context Sensing

After seeing the trigger materials, participants were curious about the ability of smartphones to detect location, mood, social situation, and activity levels in order to offer specific advice and suggestions useful for that context. However, there was a strong sentiment that this technology was "gimmicky" and had little relevance to supporting healthy behavior change.

> *I might try it just to test it and see what it did but just to see that and see how accurate it is…Even though it's really clever, I don't think I'd pay attention to it. It wouldn't make me do anything I wouldn't normally do.* [Charlotte, FG3]

There was also considerable scepticism about whether phones could sense context accurately enough to be useful. Participants thought that there would be substantial potential for mistakes and agreed that erroneous sensing of mood, location, or context would be extremely irritating and could result in losing faith and interest in an app.

> *It's just gonna tell you things that are wrong and you don't need to know.* [Joe, FG4]

> *If it gets it wrong, you would automatically get really irritated by it…I think the risk of getting it wrong would be really annoying and I'd probably delete the app.* [Isabella, FG1]

Even if accuracy was not a problem, participants had concerns about the consequences of sensed data being used to trigger intervention components. They predicted that context-triggered advice and suggestions would produce counterproductive effects by drawing attention towards unhealthy but attractive behavior choices.

> *What if you were walking past [a fast food chain], you hadn't even noticed it?... It would draw your attention to it.* [Isabella, FG1]

There was also concern about phones recognizing negative mood, inactivity, or lack of social contact. Participants suggested that realizing a mobile phone had recognized these states and offered just-in-time interventions could worsen your mood, or

unwantedly draw attention towards feelings or situations that would otherwise be unnoticed.

> *The last thing you want is your phone being like "relax!".* [Lauren, FG2]

Participants could also imagine many circumstances where any advice or suggestion based on sensed data would be impossible to follow and would simply be irritating.

> *I would get really annoyed if I was trying to do my work or something and it was telling me I hadn't moved, I'd be like "I'm busy!"* [Katie, FG1]

However, some participants envisaged situations when sensed data might helpfully activate a prompt.

> *The amount of times I have sat watching TV or whatever when it is sunny outside, I think if [an app that prompted exercise if it sensed good weather] told me to go outside, I think that is all I would need.* [Bina, FG2]

> *If your phone knows that you're walking then it could say "You're walking. Why not try walking at a faster pace?" or "At the moment you are walking this fast but if you walk a little bit faster", you know telling you things like that, by enhancing on what you're already doing as opposed to telling you to do something you're not.* [Isabella, FG1]

### Useful Prompts or Harassment?

Several participants talked about needing or wanting to be reminded of certain things, or prompted to take certain action and felt their phone could do this successfully.

> *I would really like it, I think I need pushing into doing anything, I'm not very motivated myself, that's why I have all these apps to try and motivate me… I think I'm a person that responds well to little prompts, I don't really get up and do things myself.* [Ellie, FG2]

Having choice in the frequency and timing of reminders/prompts was considered important, as was the tone and message content itself, with a preference towards positivity and praise.

> *If they can send me a message every now and then saying that I am doing well rather than just reminders. Because I quite like messages like that.* [Zoe, FG2]

The majority of male participants maintained that they would not find such messages from a phone supportive, with the exception of one.

> *If I got a message saying "you've done really well, you've done X", I think that if I'm honest I would probably quite like it.* [Aled, FG4]

Despite the potential for an app to prompt helpful behavior at useful times and places, several participants discussed annoyance caused by receiving alerts, reminders, or prompts via their smartphone. They talked about their phone "nagging" or "harassing" them and described abandoning use of an app as a result.

### Motivation and Necessity of Behavior Change

An important and common theme of the discussions was that the appeal and usefulness of apps and their features depends on whether a user was *already* motivated to change lifestyle and health habits. Features such as self-monitoring tools, reminders, and prompts were often considered to be of potential use, provided that the person was committed to engaging in an effort to change and had decided to "sign up to it", but unnecessary or irritating in the absence of pre-existing genuine motivation.

Although most participants displayed interest in making various healthy lifestyle changes, their comments suggested that they were not motivated enough to bother to consistently and "seriously" use an app to support these endeavours. Several participants discussed how behavior change apps may be most suited to people diagnosed with and being treated for either physical health or mental health problems. This included people with chronic health conditions with a higher perceived necessity for monitoring diet and exercise, ill or elderly people who might need reminding to take medications, or people engaging in psychological therapies, who might benefit from apps to help them monitor mood or complete "homework". Participants tended to believe that they themselves were not the type of person who would want or need that level of behavior change support from a smartphone app.

> *The kind of people that would sign up to these are people knowing what they are signing up for and they are people who want this kind of level of support.* [Isabella, FG1]

> *If, for example, they were particularly overweight or obese and needed to lose weight for a health-related reason and needed to monitor what they were eating in that aspect then maybe they would be useful, but for me personally I don't have any real need to use it.* [Gareth, FG4]

### Necessity for Efficiency and Convenience

Participants described how smartphone apps were most appealing and useful when they were well integrated with how they used their phones naturally. Participants disliked apps that drain battery power, take up excessive space/memory, or cannot run in the background without affecting other phone functions.

Participants spoke about their smartphones as being valuable because they make things quicker, easier, and more pleasurable to do. However, many health-related apps that they have tried or heard about were time-consuming, burdensome, or even stressful to set up and interact with. Participants displayed very little patience for using these apps.

> *It's quite easy to lose interest really because it is quite an effort and, like you said, nobody wants to spend all their life writing down what they want on their phone.* [Aled, FG4]

> *After about a week I was just forgetting to do it and just, this is too much hassle.* [Gareth, FG2]

### Disposability

The need for both appealing features and very low burden became particularly clear as participants described how they choose to use or discontinue use of apps**.** Given that there is plenty of choice of free apps, participants were not particularly committed to any individual app and might try out several that seem appealing or that they are curious about. Participants were very clear that apps were easily discarded if they did not meet expectations. Interest in apps appeared to be fleeting and participants did not expect to use them on a regular or long-term basis.

> *Sometimes if you download them and they are not as good as you thought they'd be then it's a bit rubbish, but they're free so it doesn't really matter.* [Jessica, FG3]

### Credibility and Accuracy

Participants expressed concern about whether apps they used were from reputable and legitimate sources. Apps developed by experts were considered preferable and more persuasive than those from unknown or less reputable sources.

> *If you knew that the app was being controlled or something by some kind of doctor it might make you listen more, rather than just anyone could be sitting at a computer and writing these things.* [Marie, FG3]

Participants were also concerned about whether information and advice given was safe and accurate. Several had encountered apps that they believed had provided inconsistent or inappropriate advice, which had caused them to abandon use.

### Privacy and Security Concerns

Participants voiced a number of apprehensions related to personal privacy and security. Many were uneasy about apps being able to keep their health data secure and private. Some preferred password access, but several participants also complained about the effort involved in creating accounts and entering passwords and suggested that this may dissuade them from using an app.

A key worry was whether data entered or sensed by the smartphone would get into the hands of third parties, including the app developer and other companies. They were particularly sensitive about physical and mental health data being used to tailor advertisements to them.

Participants were also uneasy about some of the context-sensing capabilities of smartphones, describing them as creepy and intrusive, and many suggested these features would discourage their use of an app.

> *I'd be wondering if it could actually hear your conversation, you'd have no sense of privacy.* [Marie, FG3]

Despite some concerns, there were participants who were very relaxed about how their data might be used by an app for advertising. Anthony described himself as "desensitized" to how the Internet uses data about you while Aled was unconcerned, believing that his physical and mental health were not interesting enough to be monitored by third parties. Sensing of location, however, prompted considerable discussion amongst female participants about personal safety (specifically, stalkers and burglars seeking opportunistic moments using GPS data to establish that someone was away from home).

### Keeping Control Over Apps

Most participants had experienced both smartphone apps and websites (especially online social networking sites) behaving in unexpected and undesirable ways. Participants were uneasy about what smartphones might do without their awareness or permission and its consequences given the personal and sensitive nature of health information.

> *You forget that it's connected to the Internet and that it can do anything with it, imagine if it just puts [information about your health] up on your Facebook.* [Isabella, FG1]

Some participants mentioned they would want to be made fully aware of what the app would do prior to using it. However, participants also said they were unlikely to want to read lengthy instructions or terms and conditions.

Participants also wanted to be in control of settings and to personalize the app during setup (and review and edit settings later on) depending on what suited them.

> *You want to be the one who sets up what it's going to do.* [Leona, FG2]

Some participants talked about wanting to be the ones that initiated interaction with apps at convenient times for them, rather than the phone alerting, or prompting them to use it.

> *If I wanted to hear from the app, I'd go into the app rather than it just pop up.* [Ellie, FG2]

Participants also discussed wanting the ability to turn off features of the app (eg, reminders or prompts) or indeed silence or disable the entire app at certain times when behavior change was not considered a priority.

> *Something that you could turn on and off rather than having something that's constantly in your face.* [Leona, FG2]

## Discussion

### Principal Findings and Questions for Future Investigations

This study explored young adults' experiences and views about health-related smartphone apps, including how they perceive various features and their willingness to use these apps. In overall concordance with the limited pre-existing research in this area [1-8], our participants displayed muted enthusiasm about the role of smartphone apps in assisting with health-related behavior change, some positive experiences of using such apps, and tentative willingness to try apps, even though many of them did not perceive a pressing need to change their health habits. The focus group data suggested a number of features of apparent importance and interest to young people. In Textbox 1 these are depicted as a preliminary checklist of app features and characteristics that app developers may wish to consider when creating health behavior apps for this population.

The findings from this study that we consider to be particularly interesting and novel relate to the apathy, concerns, and frustrations around health apps. The remainder of this discussion therefore focuses on these issues. We present 5 major challenges identified from the present study that appear to need further consideration and research in order to inform the development of acceptable and effective behavior change apps (Textbox 2).

**Textbox 1.** A preliminary checklist of valuable features of health behaviors apps.

- Low effort and pleasant to use
- Sustaining interest over long periods of time
- Cost and effort free to download and set up
- Developed by legitimate experts and the developer's credentials made explicit
- Includes features to help users track health-related behavior, including setting and monitoring goals
- Provides feedback and advice that guide people in how they can change behavior
- Generates positively framed alerts and reminders that are relevant and timely but not too frequent
- Easily turned off or disabled (certain settings and the entire app)
- Accurate and reliable information and tracking functions
- Discrete and with adequate privacy settings
- Use of the app does not negatively impact or restrict any other uses of the smartphone
- Clarity about what app will do—no surprises

**Textbox 2.** Challenges for acceptable and effective behavior change apps.

- (How) Can we keep people using behavior change apps for an extended period of time?
- (How) Can we give users features that are desirable and effective without requiring unacceptable levels of effort?
- (How) Can we provide accurate and timely information, feedback, and advice without adverse effects on mood?
- (How) Can we harness context sensing in a way that users feel comfortable with, trust, and find useful?
- (How) Can we harness social media to make interventions engaging and provide social support in a way that users are willing to engage with?

### *Challenge 1: (How) Can We Keep People Using Behavior Change Apps for an Extended Period of Time?*

An important insight from the current study was that participants lacked commitment to using any particular app and seemed likely to engage in only transient, casual use. This finding is of enormous relevance to apps that aim to support long term behavior change (eg, improving diet or physical activity levels to manage weight). This study raises doubts around whether users will use behavior change apps for long periods of time, a critical issue that will affect the effectiveness of many behavior change apps. Future research will need to examine duration of use and features that influence this. From this research, 2 factors that might be relevant to discontinued use of an app were the effort required to use the app and emotional responses to the app. These are discussed further below.

### *Challenge 2: (How) Can We Give Users Features That Are Desirable and Effective Without Requiring Unacceptable Levels of Effort?*

An important and encouraging finding from this research was that many participants were positive about using smartphones to track behavior, set targets, review progress, and receive graphical or verbal commentary on success. Previous research on the views of healthy adults as well as at-risk and chronically ill participants also suggested that monitoring and tracking features are acceptable and valuable [1,3,4,9]. Encouragingly, these features also map onto key behavior change techniques that behavioral science research has established as effective for supporting behavior change, namely self-monitoring, goal-setting, and receiving feedback on performance [10,11]. Importantly, this study suggested that although these features were attractive to users in principle, they might prove to be overly burdensome. Our participants were keen for their behavior recorded accurately and in detail, yet did not want to enter this information regularly. Furthermore, they predicted that they were likely to forget to monitor and track, yet would be irritated by prompts and reminders and would ignore them. In addition, they wanted to understand the features of an app and not be concerned that it would do something unexpected, yet they were unwilling to spend time reading explanations or instructions.

Previous research on smartphone apps in general suggested that simplicity, efficiency, and pleasure influences continued use [27]. When considered in conjunction with the low commitment that characterized the participants in this study, features that were perceived as effortful and burdensome were likely to influence use negatively. The challenge for future research is to establish whether it is possible to incorporate attractive and effective behavior change techniques into phone apps while maintaining a low user burden.

### *Challenge 3: (How) Can We Provide Accurate and Timely Information, Feedback, and Advice Without Adverse Effects on Mood?*

Another key issue that the current study exposed was the potential for interactions with health behavior apps to trigger negative emotional reactions. Participants described becoming annoyed with receiving prompts, alerts, reminders, and messages. They described irritation or disappointment as a consequence of inaccurate, untimely, or irrelevant notifications or advice. They described becoming upset or demotivated by viewing logs or records that showed they were not succeeding in meeting a goal, and by feedback with a punitive or didactic tone. They also predicted heightened distress in response to app-induced awareness of health-related states. Importantly, participants suggested that these reactions might lead them to discontinue app use. Negative emotional responses to apps would benefit from more detailed investigation in a naturalistic setting to ascertain whether they are widespread and whether they are indeed related to app discontinuation. A challenge will be to discover whether there are optimal ways that apps can communicate with the user to engage them in behavior change, keep them aware of their progress towards goals, and provide relevant and timely advice and support without generating adverse emotional reactions and threatening adherence.

### *Challenge 4: (How) Can We Harness Context Sensing in a Way That Users Feel Comfortable With, Trust, and Find Useful?*

To our knowledge, this study was the first to report potential users' views on the use of context sensing to support timely and appropriate behavioral interventions for health. Computer scientists and behavior change experts have been enthusiastic about developments in context sensing, considering it as a potential solution to the problem that self-monitoring is burdensome and unappealing for users, and seeing it as a way of encouraging timely engagement with digital interventions. In contrast, the views of the study sample were characterized by scepticism and concern. It seems that several substantial issues need considerable research and attention if context sensing is to be applied in a way that users perceive as useful and acceptable.

First, the study showed that participants lacked faith in the accuracy with which a smartphone could sense relevant states (eg, mood, activity levels) and expected that incorrect and irritating suggestions would make them mistrust the app and cease using it. Given such low expectations and the potentially adverse consequences of losing a users' trust in an app, demonstrating consistent accuracy of context sensing and that it has face validity may be an important precursor to using it to trigger advice and intervention within health apps. Second, participants predicted that even accurate sensing might produce counter-productive emotional and behavioral consequences. Future research should include a careful monitoring of unanticipated effects from sensing-driven interventions and advice. Third, in order to promote willingness to use context-sensing apps for sensitive topics such as health, careful consideration of potential users' security and privacy concerns is necessary. This might include establishing whether the risks they anticipate are realistic, possible to eliminate or reduce, and how information about security and privacy is communicated to users.

Although this study exposed mostly negative attitudes towards context sensing it did suggest opportunities to use this technology in ways that were more interesting and acceptable. Specifically, using sensing to recognize opportunities and successes such as establishing when somebody is in a receptive mood, suitable location, and/or already engaging in the target behavior may be worthwhile to pursue. Overall, careful exploration of if and how health behavior change apps can use context sensing in a way that users perceive as acceptable and useful is now needed.

### *Challenge 5: (How) Can We Harness Social Media to Make Interventions Engaging and Provide Social Support in a Way That Users Are Willing to Engage With?*

It has been proposed that social networking media may facilitate social support for behavior change and keep people interested and engaged in digital interventions [12]. However, a key finding from this study was a disinclination to use health apps that linked to online social networks. Our participants described an aversion to involving existing online social contacts in their efforts to change behavior and considered doing so to be socially unacceptable, except under certain circumscribed conditions. Previous research on views about this has been mixed. Obtaining real time social support from digital networks was a key benefit of mobile interventions cited by overweight/sedentary pre-diabetic patients [13]. Other research has highlighted how participants in a physical activity intervention had positive experiences of sharing progress and competitiveness using their mobile device [28] and how participants were interested in and enjoyed recording and sharing emotional states and participating in socially supportive activities via a mobile app [29]. Conversely, other studies suggested more negative attitudes, that users of an app targeted to reduce sedentary behavior were not interested in sharing their progress with their social networks [30], and users of apps to increase physical activity enjoyed sharing with social networks when they were being active but were uncomfortable doing so when they were not achieving high levels [31].

Given the large body of evidence implicating social support in successful behavior change, the scope for online social networks to be harnessed in behavior change apps deserves further detailed examination. Specifically, attempts to better understand the conditions under which it is acceptable and useful to link health smartphone apps to online social networks and the most acceptable and effective ways to facilitate and encourage support from these networks would be beneficial.

## Study Limitations and Conclusion

Several key limitations influence the strength of the conclusions and recommendations drawn from the current study. First, the sample was small, self-selected, and drawn from a university population. This allowed, as intended, a detailed exploration of the views of an age group who have not been the focus of extant health app research yet are likely to own and use smartphones

[2] and are likely to have scope for improving behaviors such as diet, physical activity, and alcohol consumption. However, the sample may differ from other young people and the extent to which the themes discerned here are applicable across other populations is unclear. Another limitation of this study was that the views and experiences it uncovered were either retrospective accounts of experiences with apps or discussions of *hypothetical* app use. Neither are necessarily accurate portrayals of how apps are actually used and the affective, cognitive, and behavioral responses that they generate. Further research could establish whether findings emerging in this study are observed when actual app use is studied. Useful data collection methods might include phone-generated records of the nature of actual app use, and quantitative and qualitative data on changes in behavior, emotions, and thoughts measured frequently and repeatedly as people use apps to support behavior change.

Given the limitations of the current study, its findings serve as research questions for future investigations, rather than definitive solutions to developing effective health apps. This study presents important messages to developers and behavioral scientists involved in creating smartphone apps that aim to support long-term changes to health-related behavior. A number of unresolved challenges remain, threatening efforts to develop apps that are effective in changing behavior to improve health. We recommend further investigation to resolve issues with lack of long-term use and commitment. Efforts should be made to explore how appealing and effective behavior change techniques can be incorporated into health behavior change apps while avoiding content, features, and technologies that embarrass, irritate, upset, worry, or burden users.

## Acknowledgments

This study was supported by EPSRC grant funding (Ref: EP/I032673/1 “UBhave: ubiquitous and social computing for positive behaviour change”). Thank you to the 19 participants who took the time to contribute their experiences and viewpoints within our focus groups. We would also like to acknowledge Gulcan Garip for her assistance with running one of the focus groups and Stacey Winter for transcribing.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Interview schedule.

[PDF File (Adobe PDF File), 3KB - jmir_v15i4e86_app1.pdf ]

## Multimedia Appendix 2

Trigger materials.

[PDF File (Adobe PDF File), 329KB - jmir_v15i4e86_app2.pdf ]

## References

1. Pew Internet. 2011. Smartphone adoption and usage URL: http://www.webcitation.org/6ES3uUHPv [accessed 2013-04-13]
2. OFCOM 2012. Communications market report 2012 URL: http://stakeholders.ofcom.org.uk/binaries/research/cmr/cmr12/CMR_UK_2012.pdf [accessed 2013-02-15] [WebCite Cache ID 6ES32TiXn]
3. Miller G. The smartphone psychology manifesto. Perspectives on Psychological Science 2012 May;7(3):221-237. [doi: 10.1177/1745691612441215]
4. Morris ME, Aguilera A. Mobile, social, and wearable computing and the evolution of psychological practice. Prof Psychol Res Pr 2012 Dec;43(6):622-626 [FREE Full text] [Medline: 17296997]
5. Boschen MJ, Casey LM. The use of mobile telephones as adjuncts to cognitive behavioral psychotherapy. Professional Psychology: Research and Practice 2008 Oct;39(5):546-552 [FREE Full text] [doi: 10.1037/0735-7028.39.5.546]
6. Patrick K, Griswold WG, Raab F, Intille SS. Health and the mobile phone. Am J Prev Med 2008 Aug;35(2):177-181 [FREE Full text] [doi: 10.1016/j.amepre.2008.05.001] [Medline: 18550322]
7. Preziosa A, Grassi A, Gaggioli A, Riva G. Therapeutic applications of the mobile phone. British Journal of Guidance & Counselling 2009 Aug 2009;37(3):313-325. [doi: 10.1080/03069880902957031]
8. Lane N, Mohammod M, Lin M, Yang X, Lu H, Ali S, et al. BeWell: A smartphone application to monitor, model and promote wellbeing. 2011 Presented at: 5th International Conference on Pervasive Computing Technologies for Healthcare; 2011; Dublin.
9. Mascolo C, Musolesi M, Rentfrow PJ. Mobile sensing for mass-scale behavioural intervention. 2011 Presented at: NSF Workshop on Pervasive Computing at Scale (PeCS); January 2011; University of Washington, Seattle p. 27-28.

10. Rachuri KK, Musolesi M, Mascolo C, Rentfrow PJ, Longworth C, Aucinas A. EmotionSense: a mobile phones based adaptive platform for experimental social psychology research. 2010 Presented at: Proceedings of the 12th ACM international conference on Ubiquitous computing; September 2010; Copenhagen p. 281-290.
11. Loo J. University of North Carolina; Doctoral Dissertation. Acceptance of health services on mobile phones URL: http://www.jeffloo.com/stuff/2009/Jeffery_Loo_dissertation.pdf [accessed 2013-04-13] [WebCite Cache ID 6FqvzXNam]
12. Proudfoot J, Parker G, Hadzi Pavlovic D, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]
13. Fukuoka Y, Kamitani E, Bonnet K, Lindgren T. Real-time social support through a mobile virtual community to improve healthy behavior in overweight and sedentary adults: A focus group analysis. J Med Internet Res 2011 Sep;13(3):e49. [Medline: 21752785]
14. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011 Sep;13(3):e55. [Medline: 21840837]
15. Rabin C, Bock B. Desired features of smartphone applications promoting physical activity. Telemed J E Health 2011 Dec;17(10):801-803. [doi: 10.1089/tmj.2011.0055] [Medline: 22010977]
16. Ahtinen A, Mattila E, Vaatanen A, Hynninen L, Salminen J, Koskinen E. User experiences of mobile wellness applications in health promotion: user study of Wellness Diary, Mobile Coach and SelfRelax. 2009 Presented at: 3rd International Conference on Pervasive Computing Technologies for Healthcare; April 2009; London p. 1-8.
17. Buhi ER, Trudnak TE, Martinasek MP, Oberne AB, Fuhrmann HJ, McDermott RJ. Mobile phone-based behavioural interventions for health: A systematic review. Health Education Journal 2012 Jul 2012;5(5):371 [FREE Full text] [doi: 10.1177/0017896912452071] [Medline: 22973420]
18. Hebden L, Cook A, van der Ploeg HP, Allman-Farinelli M. Development of smartphone applications for nutrition and physical activity behavior change. JMIR Res Protoc 2012 Aug;1(2):e9. [doi: 10.2196/resprot.2205]
19. Pew Internet 2011. Apps update URL: http://pewinternet.org/Reports/2011/Apps-update.aspx [accessed 2013-02-15] [WebCite Cache ID 6ES3ApiZQ]
20. West JH, Hall PC, Hanson CL, Barnes MD, Giraud-Carrier C, Barrett J. There's an app for that: content analysis of paid health and fitness apps. J Med Internet Res 2012;14(3):e72 [FREE Full text] [doi: 10.2196/jmir.1977] [Medline: 22584372]
21. Cowan LT, Van Wagenen SA, Brown BA, Hedin RJ, Seino-Stephan Y, Hall PC. Apps of Steel: Are Exercise Apps Providing Consumers With Realistic Expectations? A Content Analysis of Exercise Apps for Presence of Behavior Change Theory. Health Educ Behav :2012. [doi: 10.1177/1090198112452126]
22. Breton ER, Fuemmeler BF, Abroms LC. Weight loss—there is an app for that! But does it adhere to evidence-informed practices? Behav. Med. Pract. Policy Res 2011 Sep 2011;1(4):523-529. [doi: 10.1007/s13142-011-0076-5] [Medline: 23580463]
23. Abroms LC, Padmanabhan N, Thaweethai L, Phillips T. iPhone apps for smoking cessation: a content analysis. Am J Prev Med 2011 Mar;40(3):279-285 [FREE Full text] [doi: 10.1016/j.amepre.2010.10.032] [Medline: 21335258]
24. Chomutare T, Fernandez-Luque L, Arsand E, Hartvigsen G. Features of mobile diabetes applications: review of the literature and analysis of current applications compared against evidence-based guidelines. J Med Internet Res 2011;13(3):e65 [FREE Full text] [doi: 10.2196/jmir.1874] [Medline: 21979293]
25. Yardley L. The potential of Internet-delivered behaviour change interventions. The European Health Psychologist 2011 Sep;13(3):40-43.
26. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
27. Verkasalo H, López-Nicolás C, Molina-Castillo FJ, Bouwman H. Analysis of users and non-users of smartphone applications. Telematics and Informatics 2010;27(3):242-255.
28. Anderson I, Maitland J, Sherwood S, Barkhuus L, Chalmers M, Hall M, et al. Shakra: tracking and sharing daily activity levels with unaugmented mobile phones. Mobile Netw Appl 2007 Aug;12(2-3):185-199. [doi: 10.1007/s11036-007-0011-7] [Medline: 22811232]
29. Gay G, Pollak J, Adams P, Leonard JP. Pilot study of Aurora, a social, mobile-phone-based emotion sharing and recording system. J Diabetes Sci Technol 2011 Mar;5(2):325-332 [FREE Full text] [Medline: 21527101]
30. Dantzig S, Geleijnse G, Halteren AT. Toward a persuasive mobile application to reduce sedentary behavior. Pers Ubiquit Comput 2012 Jul 2012. [doi: 10.1007/s00779-012-0588-0]
31. Klasnja P, Consolvo S, McDonald DW, Landay JA, Pratt W. Using mobile & personal sensing technologies to support health behavior change in everyday life: lessons learned. AMIA Annu Symp Proc 2009:338-342 [FREE Full text] [Medline: 20351876]


## Abbreviations

**FG:** focus group
**SMS:** short message service

*Edited by G Eysenbach; submitted 22.02.13; peer-reviewed by J Sheats, S Winter; comments to author 21.03.13; revised version received 28.03.13; accepted 07.04.13; published 18.04.13.*

*Please cite as:*
*Dennison L, Morrison L, Conway G, Yardley L*
*Opportunities and Challenges for Smartphone Applications in Supporting Health Behavior Change: Qualitative Study*
*J Med Internet Res 2013;15(4):e86*
*URL: http://www.jmir.org/2013/4/e86/*
*doi:10.2196/jmir.2583*
*PMID:23598614*

Original Paper

# Hispanic Migrant Farm Workers' Attitudes Toward Mobile Phone-Based Telehealth for Management of Chronic Health Conditions

Matthew Price[1*], PhD; Deborah Williamson[2*], DHA, MSN, CNM; Romina McCandless[2*], MPH; Martina Mueller[2*], PhD; Mathew Gregoski[2*], PhD; Brenda Brunner-Jackson[2*], MPH; Eveline Treiber[3*]; Lydia Davidson[3*]; Frank Treiber[2,4*], PhD

[1]National Crime Victims Center, Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina, Charleston, SC, United States

[2]Technology Applications Center for Healthy Lifestyles, College of Nursing, Medical University of South Carolina, Charleston, SC, United States

[3]College of Charleston, Charleston, SC, United States

[4]College of Medicine, Medical University of South Carolina, Charleston, SC, United States

[*]all authors contributed equally

**Corresponding Author:**
Frank Treiber, PhD
Technology Applications Center for Healthy Lifestyles
College of Nursing
Medical University of South Carolina
19 Hagood Avenue
Suite 1002
Charleston, SC, 29425
United States
Phone: 1 843 792 8852
Fax: 1 843 792 8852
Email: treiberf@musc.edu

## Abstract

**Background:** Mobile phone–based interventions present a means of providing high quality health care to hard-to-reach underserved populations. Migrant farm workers (MFWs) are among the most underserved populations in the United States due to a high prevalence of chronic diseases yet limited access to health care. However, it is unknown if MFWs have access to mobile phone devices used in mobile health (mHealth) interventions, or if they are willing to use such technologies.

**Objective:** Determine rates of ownership of mobile devices and willingness to use mHealth strategies in MFWs.

**Methods:** A demonstration of mHealth devices and a survey were individually administered to 80 Hispanic MFWs to evaluate use of mobile phones and mHealth devices and willingness to use such technologies.

**Results:** Of the 80 participants, 81% (65/80) owned cell phones capable of sending and receiving health-related messages. Most participants (65/80, 81%) were receptive to using mHealth technology and felt it would be helpful in enhancing medication adherence, self-monitoring health conditions, and receiving quicker medication changes from their doctors (median scores ≥4 on 5-point Likert scales). Relations between age and attitudes toward using mHealth were not statistically significant.

**Conclusions:** Hispanic MFWs have access to mobile phones and are willing to use mHealth devices. Future work is needed to comprehensively evaluate the degree to which these devices could be used.

***(J Med Internet Res 2013;15(4):e76)*** doi:10.2196/jmir.2500

**KEYWORDS**

mobile phone; hypertension; health care disparities; rural health

## Introduction

Approximately 2 million migrant farm workers (MFWs) perform agricultural duties in the United States annually [1]. This group has been shown to be at increased risk for serious injury and chronic health conditions, including essential hypertension (EH) and type 2 diabetes [2-4]. Rates of uncontrolled EH are elevated in young adult MFW (33%), which is higher than the national average for Hispanics (23%) [5]. EH is particularly problematic because its symptoms often go unnoticed until the individual experiences a cerebrovascular or cardiac event [6]. As such, there is a clear need to improve the care of EH in this population given the high risk of complications.

A primary challenge in the treatment of chronic conditions in MFWs is identifying methods to offer timely and consistent access to care given the mobility of their occupation [2,7,8]. Repeated relocations typically spanning several states prevents this group from establishing long-term relationships with care networks, which has been associated with fragmented treatment, diminished care for chronic conditions, and poorer overall health outcomes [2]. Similarly, MFWs must navigate new health care systems in each location, including identifying care centers and pharmacies, and overcoming substantial transportation difficulties [9]. Most MFWs are non-English speaking (84%) [9], which can limit communication when interacting with primarily English-speaking providers. These issues prevent MFWs from accessing medications to manage chronic conditions, disrupt medication adherence schedules, and increase risk for more serious health outcomes. Innovative strategies to provide routine and seamless contact between patients and providers for MFWs in the treatment of EH and other chronic conditions are needed [10]. Migrant community health centers established through the Federally Qualified Health Center (FQHC) system have helped improve care, but these centers are often underutilized, suggesting that alternative care models are needed [11].

Mobile health (mHealth) technology is a viable option to facilitate timely communication between providers and patients with the goal of providing consistent care for MFWs. Such technologies can deliver automated summary reports of health conditions to health care providers and patients in real time (eg, blood pressure and glucose), deliver interventions targeting motivation for behavior change, and promote adherence to medication regimens. Recent findings indicate that mHealth programs are capable of enhancing patient control of chronic diseases, such as EH and type 2 diabetes [12,13]. Such interventions capitalize on the ability to communicate with patients without interfering with their routines. For MFWs, such tools could prove invaluable toward adherence to recommended treatment. Additionally, such approaches could sustain partnerships with the FQHC providers across their migratory work patterns.

A necessary first step in devising a novel care model is to determine the extent that the new population is willing to adopt the strategy [14]. Prior work with other populations has suggested that most individuals are receptive to the use of mobile phone-administered treatments for elevated blood pressure (BP) [6], heart failure alert [15], asthma [16], stroke recovery [17], multiple sclerosis [18], depression [19], and physical therapy [20]. Findings indicated that patients were enthusiastic to use such services (yet expressed the need for technical assistance if needed), were concerned with malfunctioning equipment, and were worried about the confidentiality of their information. Most of these studies have not involved low-income minority populations. George and colleagues [21] examined similar issues in underserved African Americans and Latinos and found that these groups had elevated concerns about confidentiality of information, quality of care received via a mobile device, and were less trusting of the medical community. Similarly, a recent study found that low socioeconomic status was associated with a reduced willingness to use an Internet-based monitoring system for EH [22]. These results indicate that disadvantaged groups, those who are most likely to benefit from mHealth because of reduced access, may be the most skeptical of this treatment strategy.

In MFWs, concerns regarding confidentiality of information, trust of medical care centers, and fear of poor quality of care are likely to be elevated. Given that many of these workers are undocumented, there are concerns about providing identifying and tracking information that could have adverse consequences. Many have concerns about sharing identifying information because of immigration status [2]. For example, such concerns have limited this group from seeking emergency care in critical situations. Additionally, MFWs have limited resources; therefore, they may not have regular access to mobile devices. Thus, it remains unknown whether migrant farmers have access to mobile devices or are willing to adopt mHealth technology for management of health care-related needs, despite the potential benefits.

The current study used a formal survey developed based upon previous measures administered to patients with varying chronic diseases and our findings from interviews with MFWs and their health care providers. We aimed to determine rates of mobile phone ownership including smartphones, utilization of various phone features, and awareness of mHealth technology within a sample of MFWs at 2 worksites in rural South Carolina. The participants were then given a brief overview of how various medical devices, such as weight scales, glucometers, and BP monitors, can be enabled to send data to a smartphone and then to a secure computer via transmission across cell towers. The demonstration of a prototype mHealth system for BP control was provided and acceptability and willingness to utilize such technology was assessed. The demonstration involved a brief explanation of what information would be transmitted to their doctor (BP, blood glucose, weight) and how to use the BP monitor. Based upon the rapidly expanding ubiquity of cell phone usage across ethnic groups, it was hypothesized that MFWs would have a high rate of ownership of mobile devices (>75%). Additionally, based on the outcomes of several other studies that demonstrated a positive attitude toward mHealth technology [6,15-17,20], it was hypothesized that MFWs with EH would be more willing to use the technology than those without a diagnosis. Lastly, it was hypothesized that those who use a wider array of features on their mobile device, such as

sending text messages or downloading applications, would have more positive perceptions about mHealth.

## Methods

### Participants

The participants in this study were 80 MFWs stationed at 1 of 2 agricultural complexes in Charleston County, South Carolina, during the 2011 and 2012 spring harvesting seasons. Participants were predominately male (70%) with a mean age of 29.76 years (SD 9.82). All participants were Hispanic and spoke fluent Spanish.

### Measures

A questionnaire was designed to evaluate demographic information, EH status, and self-reported medication adherence. A series of 9 questions to assess attitudes toward mobile phone remote monitoring for chronic disease management was adapted from prior studies [6,15,23]. These items assessed willingness to use a mHealth service: (1) for EH and diabetes care in general, (2) if they were taught how to use the devices, and (3) if technical support was available. Additional questions assessed beliefs about the effectiveness of mHealth practices for EH and diabetes and concerns about confidentiality. Ratings were made for each question on a 5-point Likert scale with higher scores indicating increased willingness to use such technology. Internal consistency for the measure was excellent with alpha =.92. The tenth item queried their a priori awareness of health-related remote monitoring technology. The questionnaire was administered in an interview format in Spanish. Another native speaker was available to assist any participants who needed assistance.

### Procedure

The MFWs were approached in 8 groups of 6 to 12 participants in the residential location of their worksite over a 4-month period. No participants declined to participate. Participants were given a brief description and demonstration of a BP device with Bluetooth wireless technology (AND model 9025 BT) and a Motorola Droid X smartphone with an installed software application for BP signal reception and transfer to secure server. The survey was approved by the Institutional Review Board of the Medical University of South Carolina.

## Results

### Clinical Characteristics

Seventeen (21%) farmers had EH based upon a previous BP evaluation by an on-site physician, but only 7 (41%) of these 17 patients reported having received prescriptions for this condition. Those classified as EH had higher systolic/diastolic BP (systolic mean 140.6, SD 20.2; diastolic mean 86.5, SD 13.9) than non-EH participants (systolic mean 116.1, SD 13.5; diastolic mean 73.7, SD 10.8). Self-reported medication adherence using the Morisky scale revealed that among those with EH, only 5 of 17 (29%) were completely adherent within the past week (eg, did not miss a dose).

### Mobile Phone Utilization

Cell phone ownership among this sample was substantial with 81% (65/80) having access to 1 or more mobile phones. Over one-third (31/80, 39%) owned a smartphone capable of using mobile phone applications. A random subsample of 10 farmers who reported owning a smartphone were asked to show their phones so they could be verified as Internet-capable smartphones. All were verified as being smartphones with Internet connectivity. Text messaging was the most commonly used mobile phone feature of MFWs (62/80, 78%). Participants reported using their phones for email (28/80, 35%), accessing the Internet (36/80, 45%), and downloading applications (38/80, 48%). Few participants (12/80, 15%) had prior knowledge of the use of mobile phones in health care.

### Attitudes and Willingness Toward mHealth Technology

As shown in Table 1, most participants (65/80, 81%) reported they would likely or definitely use mHealth services if available. There was an increase in willingness if free technical support was offered (75/80, 94%). Most endorsed mHealth services as being helpful in maintaining timely linkages with their doctors (68/80, 84%) and reported having minimal or no doubts about the security of their health information on mHealth services (61/80, 76%). Security about health information was defined as the transfer of BP values, medication adherence, and other personal health information not being viewed by anyone other their health care provider team and themselves.

A series of multiple regressions were used to identify predictors and barriers toward positive perceptions of using mHealth services (Table 2). Effect sizes for each model were moderate with the multivariate coefficients of determination ($R^2$) ranging from 0.12 to 0.28. There were no differences in willingness to use mHealth across those with EH and those without EH. There were no significant predictors of willingness to use mHealth without assistance or willingness to use mHealth with a tutorial. Accessing the Internet via a phone was associated with increased willingness to use mHealth with continued technical support (beta=0.39, *P*=.02). Those who had prior knowledge of mHealth were less willing to continue to use the service with continued technical support (beta=–0.27, *P*=.02), were more likely to believe that mHealth can improve patient-provider communication (beta=0.38, *P*=.03), and were less concerned about confidentiality (beta=0.39, *P*=.02). The negative association between age and beliefs that mHealth can improve patient-provider communication did not achieve statistical significance (beta=–0.27, *P*=.05). Finally, women were less concerned about confidentially through mHealth (beta=0.34, *P*=.006).

**Table 1.** Responses to self-report survey assessing attitudes toward mHealth technology in migrant farm workers (N=80).

| Questions assessing attitudes | | Proportion of responses, n (%) |
|---|---|---|
| **Prior awareness of mHealth technology** | | |
| | Yes | 10 (13) |
| | No | 70 (87) |
| **Willingness to use mHealth to manage essential hypertension (EH) and diabetes** | | |
| | Definitely would not use it | 4 (5) |
| | Not likely to use | 3 (4) |
| | Neither likely nor unlikely | 8 (10) |
| | Likely would use | 37 (46) |
| | Definitely would use it | 28 (35) |
| **Willingness to use a mHealth with initial tutorial on the application** | | |
| | Definitely would not use it | 2 (3) |
| | Not likely to use | 3 (4) |
| | Neither likely nor unlikely | 4 (6) |
| | Likely would use | 34 (41) |
| | Definitely would use it | 37 (46) |
| **Willingness to use mHealth with continued technical support** | | |
| | Definitely would not use it | 2 (3) |
| | Not likely to use | 0 (0) |
| | Neither likely nor unlikely | 3 (4) |
| | Likely would use | 34 (43) |
| | Definitely would use it | 41 (51) |
| **Confident mHealth can improve communication with provider about EH and diabetes** | | |
| | No it cannot | 2 (3) |
| | Yes, but have many doubts | 2 (3) |
| | Not sure | 8 (10) |
| | Yes but have some doubts | 10 (13) |
| | Yes without a doubt | 58 (71) |
| **Confident privacy protected when using mHealth system** | | |
| | No trust | 2 (3) |
| | Many doubts | 5 (7) |
| | Neither trust nor distrust | 10 (13) |
| | Few doubts | 16 (20) |
| | Complete trust | 45 (56) |

**Table 2.** Predictors of attitudes toward mHealth technologies among migrant farm workers.

| Variable | Willingness to use mHealth | | Willingness to use mHealth with initial tutorial on the app | | Willingness to use mHealth with continued technical support | | Belief that mHealth can improve communication with provider | | Concerns about confidentiality | |
|---|---|---|---|---|---|---|---|---|---|---|
| | β | *P* | β | *P* | β | *P* | β | *P* | β | *P* |
| Age | –0.02 | .87 | 0.14 | .29 | 0.05 | .67 | –0.27 | .05 | 0.05 | .70 |
| Male gender | 0.06 | .62 | 0.11 | .37 | –0.02 | .89 | –0.01 | .99 | 0.34 | .006 |
| Diagnosis of EH | <0.01 | .99 | 0.06 | .62 | –0.03 | .79 | –0.12 | .37 | –0.03 | .82 |
| Own cell phone | 0.24 | .09 | 0.15 | .29 | 0.22 | .08 | 0.22 | .12 | 0.13 | .33 |
| Own smartphone | –0.17 | .26 | –0.25 | .10 | –0.12 | .41 | –0.13 | .42 | –0.03 | .84 |
| Sends text messages | 0.21 | .13 | 0.32 | .06 | 0.148 | .23 | –0.15 | .28 | –0.14 | .28 |

## Discussion

The current study demonstrated that MFWs have high access to mobile devices. This suggests that the mHealth infrastructure, via short message service (SMS) text messages, phone calls, or emails, exists within this population. Over one-third (39%) had an Internet-capable smartphone which is consistent with the rising national average, as well as estimates of smartphone ownership among Hispanics from a current nationally representative sample (49%) [24]. This proportion is expected to steadily increase as such devices become more affordable and available on a growing number of flexible (month-to-month) plans. Current estimates project that virtually all cell phones in the United States will be smartphones with Internet access capability and Bluetooth-enabled within 1.5 to 2 years [25].

Most participants indicated they would use mHealth services (particularly if free technical support was available), believed it would be helpful for managing chronic diseases, such as hypertension and/or diabetes, and expressed few concerns about the security of their medical data across the Internet. Participants who indicated they had no prior knowledge of mHealth were more receptive to being involved in mHealth if they received an initial personal tutorial to help facilitate use of a mHealth service. This highlights a potential barrier: reluctance to engage in use of such technology due to lack of exposure and unfamiliarity with mHealth programs. Designers of mHealth applications would benefit from engaging patients in the development process to help ensure instructions are easily understood. Indeed, recent work has highlighted that problems caused by user error and technological errors are likely to increase frustration and prevent continued use of such approaches [16]. Incorporation of a patient-centered approach will further increase the likelihood of a user-friendly solution that will educate and encourage novice users to engage in the intervention [26-30]. However, additional work is needed to determine if such tutorials, which may be perceived as frustrating to those with higher levels of software proficiency, are a barrier to those with knowledge of such treatments.

There were no differences in any of the willingness measures across those with EH and those without EH. The perceived lack of need for continued care among those with such a condition may further highlight the need to develop mHealth programs that are user-friendly and minimally burdensome. Those with EH may not perceive the management of their condition as particularly relevant to their daily functioning [6]. Thus, mHealth systems cannot rely on participant motivation to ensure the use of such technology.

These findings suggest that mHealth is a promising method of providing health care to the difficult-to-reach group of MFWs. The findings from this study helped facilitate development of a recently activated 3-month proof-of-concept mHealth medical regimen enhancement trial among uncontrolled EH patients [31]. Electronic medication trays provide reminder signals and smartphone messages remind patients to measure their BP with the Bluetooth-enabled monitor used in the demonstration. Patients receive personalized motivational and reinforcement messages based upon adherence levels to the mHealth program. All who have received the mHealth program thus far have showed increases in medication intake (≥95% across the 3 months), reduced resting systolic BPs (from 163.8 to ≤120.9 mmHg at each monthly clinic evaluation and at 3-month follow-up), and reduced 24-hour systolic BP (from 151.3 to 122.7 mmHg at the completion of 3 months). These promising findings require completion of the proof-of-concept trial and subsequent refinement of the mHealth system based upon patient and provider feedback. This suggests that such a system should be evaluated with MFWs.

Although informative, the current findings must be interpreted cautiously. The sample size was relatively small compared with prior work [6,16]; it is unclear if these findings will apply across MFW communities and to other health care issues common to Hispanic MFWs [2]. Additional epidemiological work is needed to more fully gauge ownership of mobile phones in MFWs and attitudes toward the use of such approaches in medical care with empirically validated measures. Similarly, only attitudes were evaluated, and the extent these mHealth-driven interventions will truly be used is unknown, particularly within a framework of multisite, interlinked health centers, such as FQHCs. Finally, related work in this area has highlighted the need to identify provider attitudes to use such services [15]. Such work should focus on the willingness of providers to offer care to patients who they may be unable to contact face-to-face in the event of an emergency and the frequency with which contact should occur.

There are numerous barriers in managing chronic conditions among MFWs that mHealth has the potential to overcome.

Remote access would allow MFWs and their primary providers to maintain contact and improve adherence to health maintenance behaviors [32]. The MFWs' primary community health center provider teams, typically FQHCs, would be able to help navigate the MFWs as they migrate across geographical regions to gain access to other FQHCs and pharmacies that provide discounted or free medications. The consistent connection between patients and providers through the mHealth system would also likely enhance provider competence and cultural awareness for working with this group. A model of seamless interstate health care for MFWs as they traverse the eastern seaboard seems plausible, especially through collaboration with rural FQHCs. If future large-scale efficacy and effectiveness trials reveal similar results to the pilot work, the mHealth program for EH control will be ready for large-scale dissemination among MFWs. Such programs would vastly improve the health care of this highly vulnerable and underserved group. The data from the current study will support future work to develop culturally sensitive efficacious mHealth programs for this underserved population.

## Acknowledgments

This paper was supported by the South Carolina Clinical & Translational Research Institute, with an academic home at the Medical University of South Carolina, CTSA NIH/NCRR, Grant no UL1RR029882, a Duke Endowment Award 6355-SP, a Verizon Foundation Award, and a postdoctoral training fellowship award (T32 MH018869). The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH or NCRR.

## Conflicts of Interest

None declared.

## References

1. Holmes SM. An ethnographic study of the social context of migrant health in the United States. PLoS Med 2006 Oct;3(10):e448 [FREE Full text] [doi: 10.1371/journal.pmed.0030448] [Medline: 17076567]
2. Arcury TA, Quandt SA. Delivery of health services to migrant and seasonal farmworkers. Annu Rev Public Health 2007;28:345-363. [doi: 10.1146/annurev.publhealth.27.021405.102106] [Medline: 17291182]
3. Keenan NL, Rosendorf KA, Centers for Disease Control and Prevention (CDC). Prevalence of hypertension and controlled hypertension - United States, 2005-2008. MMWR Surveill Summ 2011 Jan 14;60 Suppl:94-97 [FREE Full text] [Medline: 21430632]
4. Pabon-Nau LP, Cohen A, Meigs JB, Grant RW. Hypertension and diabetes prevalence among U.S. Hispanics by country of origin: the National Health Interview Survey 2000-2005. J Gen Intern Med 2010 Aug;25(8):847-852 [FREE Full text] [doi: 10.1007/s11606-010-1335-8] [Medline: 20490949]
5. Egan BM, Zhao Y, Axon RN. US trends in prevalence, awareness, treatment, and control of hypertension, 1988-2008. JAMA 2010 May 26;303(20):2043-2050. [doi: 10.1001/jama.2010.650] [Medline: 20501926]
6. Bostock Y, Hanley J, McGown D, Pinnock H, Padfield P, McKinstry B. The acceptability to patients and professionals of remote blood pressure monitoring using mobile phones. Primary Health Care 2009 Jul 2009;10(04):299. [doi: 10.1017/S1463423609990107]
7. Slesinger DP. Health status and needs of migrant farm workers in the United States: a literature review. J Rural Health 1992;8(3):227-234. [Medline: 10121552]
8. Hansen E, Donohoe M. Health issues of migrant and seasonal farmworkers. J Health Care Poor Underserved 2003 May;14(2):153-164. [Medline: 12739296]
9. Carroll DJ, Samardick R, Bernard S, Gabbard S, Hernandez T. Findings from the National Agricultural Workers Survey (NAWS) 2001-2002: A Demographic and Employment Profile of United States Farm Workers. Washington, DC: US Department of Labor, Office of the Assistant to Secretary for Policy; 2005 Mar. URL: http://www.doleta.gov/msfw/pdf/naws_rpt9.pdf [accessed 2013-03-23] [WebCite Cache ID 6FKoQW9GY]
10. Hartley D. Rural health disparities, population health, and rural culture. Am J Public Health 2004 Oct;94(10):1675-1678. [Medline: 15451729]
11. Schmalzried H, Fallon LF. Reducing barriers associated with delivering health care services to migratory agricultural workers. Rural Remote Health 2012;12:2088 [FREE Full text] [Medline: 22827833]
12. Kouris I, Mougiakakou S, Scarnato L, Iliopoulou D, Diem P, Vazeou A, et al. Mobile phone technologies and advanced data analysis towards the enhancement of diabetes self-management. Int J Electron Healthc 2010;5(4):386-402. [Medline: 21041177]
13. Weinstock RS, Teresi JA, Goland R, Izquierdo R, Palmas W, Eimicke JP, IDEATel Consortium. Glycemic control and health disparities in older ethnically diverse underserved adults with diabetes: five-year results from the Informatics for Diabetes Education and Telemedicine (IDEATel) study. Diabetes Care 2011 Feb;34(2):274-279 [FREE Full text] [doi: 10.2337/dc10-1346] [Medline: 21270184]

14. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press; 2001. URL: http://www.nap.edu/openbook.php?isbn=0309072808 [accessed 2013-02-15] [WebCite Cache ID 6ESJYt3N7]
15. Seto E, Leonard KJ, Masino C, Cafazzo JA, Barnsley J, Ross HJ. Attitudes of heart failure patients and health care providers towards mobile phone-based remote monitoring. J Med Internet Res 2010;12(4):e55 [FREE Full text] [doi: 10.2196/jmir.1627] [Medline: 21115435]
16. Cleland J, Caldow J, Ryan D. A qualitative study of the attitudes of patients and staff to the use of mobile phone technology for recording and gathering asthma data. J Telemed Telecare 2007;13(2):85-89. [doi: 10.1258/135763307780096230] [Medline: 17359572]
17. Piron L, Turolla A, Tonin P, Piccione F, Lain L, Dam M. Satisfaction with care in post-stroke patients undergoing a telerehabilitation programme at home. J Telemed Telecare 2008;14(5):257-260. [doi: 10.1258/jtt.2008.080304] [Medline: 18633001]
18. Haase R, Schultheiss T, Kempcke R, Thomas K, Ziemssen T. Use and acceptance of electronic communication by patients with multiple sclerosis: a multicenter questionnaire study. J Med Internet Res 2012;14(5):e135 [FREE Full text] [doi: 10.2196/jmir.2133] [Medline: 23069209]
19. Proudfoot J, Parker G, Hadzi Pavlovic D, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]
20. Tousignant M, Boissy P, Moffet H, Corriveau H, Cabana F, Marquis F, et al. Patients' satisfaction of healthcare services and perception with in-home telerehabilitation and physiotherapists' satisfaction toward technology for post-knee arthroplasty: an embedded study in a randomized trial. Telemed J E Health 2011 Jun;17(5):376-382. [doi: 10.1089/tmj.2010.0198] [Medline: 21492030]
21. George SM, Hamilton A, Baker R. Pre-experience perceptions about telemedicine among African Americans and Latinos in South Central Los Angeles. Telemed J E Health 2009;15(6):525-530 [FREE Full text] [doi: 10.1089/tmj.2009.0152] [Medline: 19566397]
22. Green BB, Anderson ML, Ralston JD, Catz S, Fishman PA, Cook AJ. Patient ability and willingness to participate in a web-based intervention to improve hypertension control. J Med Internet Res 2011;13(1):e1 [FREE Full text] [doi: 10.2196/jmir.1625] [Medline: 21371993]
23. Pinnock H, Slack R, Pagliari C, Price D, Sheikh A. Professional and patient attitudes to using mobile phone technology to monitor asthma: questionnaire survey. Prim Care Respir J 2006 Aug;15(4):237-245 [FREE Full text] [doi: 10.1016/j.pcrj.2006.03.001] [Medline: 16843066]
24. Pew Internet & American Life Project. Trend Data (Adults): Demographics of Internet Users. Washington, DC: Pew Internet & American Life Project; 2012 Apr. URL: http://pewinternet.org/Trend-Data-(Adults)/Whos-Online.aspx [accessed 2013-02-15] [WebCite Cache ID 6ERz64Q6C]
25. Nielson Newswire. 2012 Mar 29. Smartphones account for half of all mobile phones, dominate new phone purchases in the US URL: http://blog.nielsen.com/nielsenwire/online_mobile/smartphones-account-for-half-of-all-mobile-phones-dominate-new-phone-purchases-in-the-us/ [accessed 2012-12-18] [WebCite Cache ID 6D0FbLAkh]
26. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
27. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
28. Price M, Gros DF, McCauley JL, Gros KS, Ruggiero KJ. Nonuse and dropout attrition for a web-based mental health intervention delivered in a post-disaster context. Psychiatry 2012;75(3):267-284. [doi: 10.1521/psyc.2012.75.3.267] [Medline: 22913502]
29. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar 1;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
30. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol Med 2009 May;39(5):705-712. [doi: 10.1017/S0033291708004224] [Medline: 18812006]
31. McGillicuddy JW, Gregoski MJ, Jackson B, Weiland A, Patel SK, Rock RA, et al. Facilitating medication adherence and eliminating therapeutic inertia using wireless technology: proof of concept findings with uncontrolled hypertensives and kidney transplant recipients. Journal of the Association of Computing Machinery 2013:1-10 (forthcoming).
32. Weiss LJ, Blustein J. Faithful patients: the effect of long-term physician-patient relationships on the costs and use of health care by older Americans. Am J Public Health 1996 Dec;86(12):1742-1747. [Medline: 9003131]

## Abbreviations

**BP:** blood pressure
**EH:** essential hypertension
**FQHC:** Federally Qualified Health Center
**MFW:** migrant farm workers
**SMS:** short message service

*Edited by G Eysenbach; submitted 18.12.12; peer-reviewed by E Afari-kumah, C Masino; comments to author 18.01.13; revised version received 29.01.13; accepted 13.02.13; published 26.04.13.*

*Please cite as:*
*Price M, Williamson D, McCandless R, Mueller M, Gregoski M, Brunner-Jackson B, Treiber E, Davidson L, Treiber F*
*Hispanic Migrant Farm Workers' Attitudes Toward Mobile Phone-Based Telehealth for Management of Chronic Health Conditions*
*J Med Internet Res 2013;15(4):e76*
*URL: http://www.jmir.org/2013/4/e76/*
*doi:10.2196/jmir.2500*
*PMID:23624105*

Review

# A New Dimension of Health Care: Systematic Review of the Uses, Benefits, and Limitations of Social Media for Health Communication

S Anne Moorhead[1], PhD, MSc; Diane E Hazlett[1], PhD, MSc; Laura Harrison[1], MSc; Jennifer K Carroll[2], MD, MPH; Anthea Irwin[1], PhD; Ciska Hoving[3], PhD

[1]School of Communication, University of Ulster, Newtownabbey, Northern Ireland, United Kingdom

[2]Family Medicine Research Programs, Department of Family Medicine, University of Rochester Medical Center, Rochester, NY, United States

[3]CAPHRI School for Public Health and Primary Care, Department of Health Promotion, Maastricht University, Maastricht, Netherlands

**Corresponding Author:**
S Anne Moorhead, PhD, MSc
School of Communication
University of Ulster
Shore Road
Newtownabbey, Northern Ireland, BT37 OQB
United Kingdom
Phone: 44 28 90368905
Fax: 44 28 90368251
Email: a.moorhead@ulster.ac.uk

## Abstract

**Background:** There is currently a lack of information about the uses, benefits, and limitations of social media for health communication among the general public, patients, and health professionals from primary research.

**Objective:** To review the current published literature to identify the uses, benefits, and limitations of social media for health communication among the general public, patients, and health professionals, and identify current gaps in the literature to provide recommendations for future health communication research.

**Methods:** This paper is a review using a systematic approach. A systematic search of the literature was conducted using nine electronic databases and manual searches to locate peer-reviewed studies published between January 2002 and February 2012.

**Results:** The search identified 98 original research studies that included the uses, benefits, and/or limitations of social media for health communication among the general public, patients, and health professionals. The methodological quality of the studies assessed using the Downs and Black instrument was low; this was mainly due to the fact that the vast majority of the studies in this review included limited methodologies and was mainly exploratory and descriptive in nature. Seven main uses of social media for health communication were identified, including focusing on increasing interactions with others, and facilitating, sharing, and obtaining health messages. The six key overarching benefits were identified as (1) increased interactions with others, (2) more available, shared, and tailored information, (3) increased accessibility and widening access to health information, (4) peer/social/emotional support, (5) public health surveillance, and (6) potential to influence health policy. Twelve limitations were identified, primarily consisting of quality concerns and lack of reliability, confidentiality, and privacy.

**Conclusions:** Social media brings a new dimension to health care as it offers a medium to be used by the public, patients, and health professionals to communicate about health issues with the possibility of potentially improving health outcomes. Social media is a powerful tool, which offers collaboration between users and is a social interaction mechanism for a range of individuals. Although there are several benefits to the use of social media for health communication, the information exchanged needs to be monitored for quality and reliability, and the users' confidentiality and privacy need to be maintained. Eight gaps in the literature and key recommendations for future health communication research were provided. Examples of these recommendations include the need to determine the relative effectiveness of different types of social media for health communication using randomized control trials and to explore potential mechanisms for monitoring and enhancing the quality and reliability of health communication using social media. Further robust and comprehensive evaluation and review, using a range of methodologies, are required to establish whether social media improves health communication practice both in the short and long terms.

*(J Med Internet Res 2013;15(4):e85)* doi:10.2196/jmir.1933

**KEYWORDS**

health communication; social media; review

## Introduction

There is an ongoing increase in the use of social media globally [1], including in health care contexts [2-9]. When focusing on social media for health communication, it is useful to first outline the general characteristics of social media. Kaplan and Haenlein [10] defined social media as "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user generated content". They suggested that social media can be classified as two components: media-related and social dimension. The media-related component [11] involves how close to synchronous face-to-face communication different types of social media come and how well they reduce ambiguity and uncertainty. The social dimension is based on Goffman's [12] notion of self-presentation, whereby individuals' interactions have the purpose of trying to control others' impressions of them.

Social media provides opportunities for users to generate, share, receive, and comment on social content among multiusers through multisensory communication [1,2,10,13,14]. Although the terms "social media" and "social networking" are often used interchangeably and have some overlaps, they are not really the same. Social media functions as a communication channel that delivers a message, which involves asking for something. Social networking is two-way and direct communication that includes sharing of information between several parties. Social media can be classified in a number of ways to reflect the diverse range of social media platforms, such as collaborative projects (eg, Wikipedia), content communities (eg, YouTube), social networking sites (eg, Facebook), and virtual game and social worlds (eg, World of Warcraft, Second Life) [10].

The relationship between personality traits and engagement with social media has been reported [15]. Gender is a factor in that extraverted women and men are equally likely to engage, but emotional instability increases usage only for men. Age is also a factor in that extraversion is particularly important in younger users, while openness to new experiences is particularly important in older users [15]. Lenhart and colleagues [16] explored various types of Internet usage among teens and young adults in the United States between 2006 and 2010. During this time, social networking sites experienced the biggest rise (an average of around 50%), and the key shift in use came at age 30 years with almost double the number of teens and 18-29 years old using them as those 30 years and over (73% compared with 39%).

Social media is changing the nature and speed of health care interaction between individuals and health organizations. The general public, patients, and health professionals are using social media to communicate about health issues [2-9]. In the United States, 61% of adults search online and 39% use social media such as Facebook for health information [7]. Social media adoption rates vary in Europe; for example, the percentage of German hospitals using social networks is in "single figures", whereas approximately 45% of Norwegian and Swedish hospitals are using LinkedIn, and 22% of Norwegian hospitals use Facebook for health communication [8]. Recent UK statistics reported Facebook as the fourth most popular source of health information [9]. There have been many applications of social media within health contexts, ranging from the World Health Organization using Twitter during the influenza A (H1N1) pandemic, with more than 11,700 followers [4], to medical practices [3] and health professionals obtaining information to inform their clinical practice [5,6].

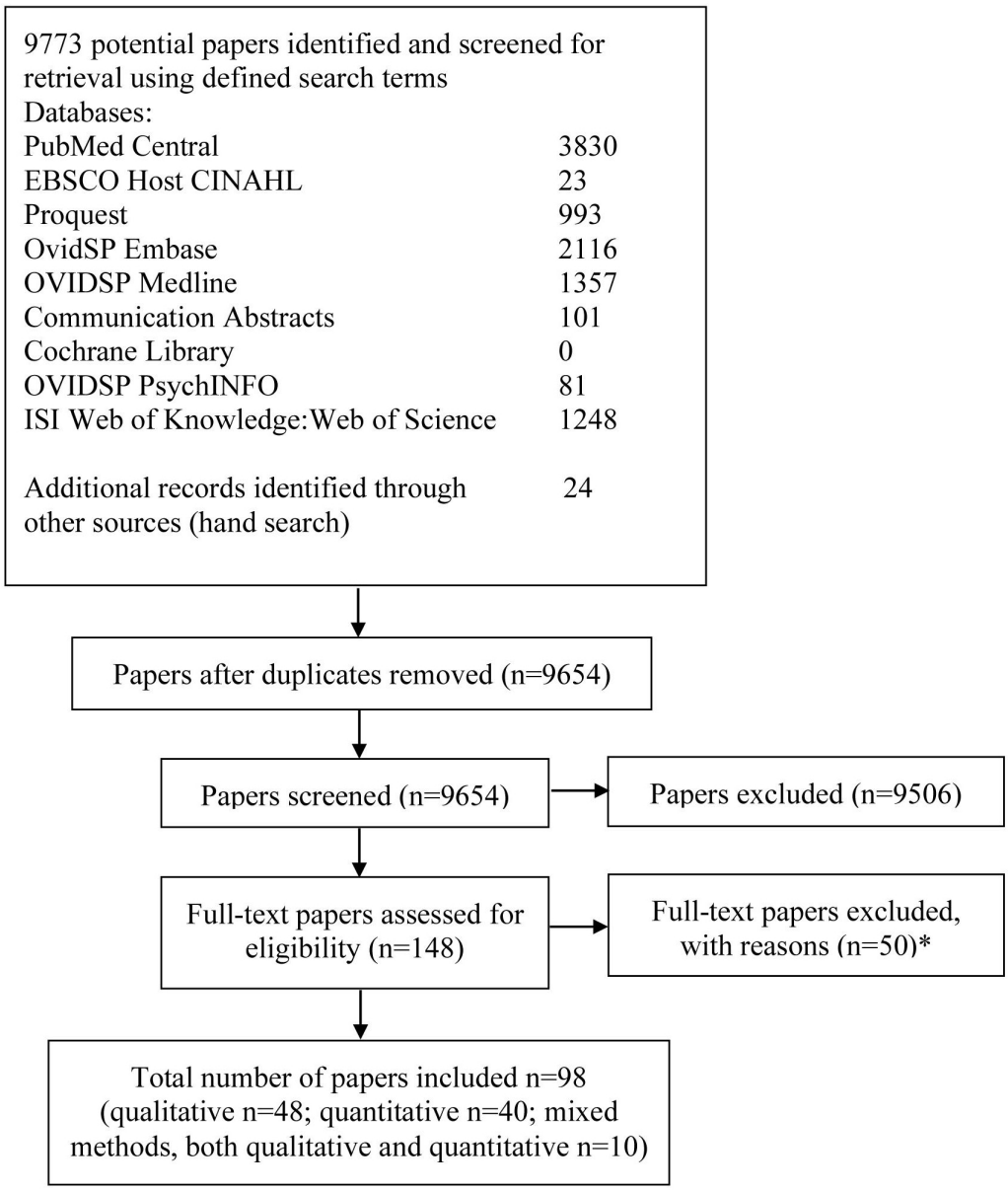
To explore the diversity in form and function of different social media platforms, Keitzmann and colleagues [17] presented the "social media ecology", a honeycomb framework of seven building blocks that are configured by different social media platforms and have different implications for organizations such as health care providers. In developing their model, they have drawn on Butterfield [18], Morville [19], Webb [20], and Smith [21]. The building blocks are (1) identity: the extent to which users reveal themselves, (2) conversations: the extent to which users communicate with each other, (3) sharing: the extent to which users exchange, distribute, and receive content, (4) presence: the extent to which users know if others are available, (5) relationships: the extent to which users relate to each other, (6) reputation: the extent to which users know the social standing of others and content, and (7) groups: the extent to which users are ordered or form communities. Thus organizations, including health care providers, need to recognize and understand the social media landscape, where the conversations about them are already being held, and develop their own strategies where suitable [17]. Similarly, Mangold and Faulds [22] highlighted that social media is changing the relationship between producers and consumers of a message. This suggests that health care providers may need to take a certain degree of control over online health communication to maintain validity and reliability.

In this paper, social media for health communication refers to the general public, patients, and health professionals communicating about health issues using social media platforms such as Facebook and Twitter. Currently, there is a lack of information about the uses, benefits, and limitations of social media for health communication among the general public, patients, and health professionals from primary research. The objective of this paper was to review the current published literature to identify the uses, benefits, and limitations of social media for health communication among the general public, patients, and health professionals and to identify current gaps in the literature to provide recommendations for future health communication research. This is important in order to establish whether social media improves health communication practices.

## Methods

This review paper followed the PRISMA guidelines [23] and used a systematic approach to retrieve the relevant research studies. The review included all study designs in order to identify the best evidence available to address the research objective. The literature search was conducted on February 7, 2012, using the following 10 electronic databases: CSA Illumina, Cochrane Library, Communication Abstracts, EBSCO Host CINAHL, ISI Web of Knowledge, Web of Science, OvidSP Embase, OvidSP MEDLINE, OVIDSP PsycINFO, and PubMeb Central. The searches were performed using the following defined search terms: “social media” OR “social network” OR “social networking” OR “Web 2.0” OR “Facebook” OR “Twitter” OR “MySpace” AND “Health”. From the above database searches, 9749 hits were identified. Manual searches were conducted in the *Journal of Medical Internet Research* (January 2002 to February 2012) where 24 papers were identified; thus, 9773 papers were identified in total. The papers’ titles and abstracts were screened for relevance, duplication, and the selection criteria. The inclusion criteria were (1) primary focus on all communication interactions within and between the general public and/or patients and/or health professionals about health issues using social media, (2) including the uses and/or benefits and/or limitations of social media for health communication, (3) original research studies, (4) published between January 2002 and February 2012, and (5) all study designs. The exclusion criteria were (1) studies not in English, (2) literature reviews, dissertation theses, review papers, reports, conference papers or abstracts, letters (to the editor), commentaries and feature articles, (3) studies only on Web 1.0 ( ie, traditional Internet use), and (4) studies with a primary marketing or advertising focus. In total, 98 original research studies that included the use, and/or benefits, and/or limitations of social media for health communication among the general public, patients, and health professionals were selected for this review [24-121] (see Figure 1). Excluded studies and the reasons for exclusion are listed in Multimedia Appendix 1. Two researchers (AM, LH) independently reviewed and evaluated the studies and reached consensus on the inclusion for the analysis. The interrater reliability between them was 0.90, indicating strong agreement [122]. Any discrepancies were discussed with reference to the research objective until consensus was reached.

**Figure 1.** PRISMA flow diagram illustrating the study selection procedure.




## Results

The 98 selected studies are summarized by study design, social media tool/application, study purpose, participants/sample and sample size, measurement tools, results, conclusion, and use of social media in Multimedia Appendix 2 [24-121]. The diverse studies included the use of a range of social media tools/applications, the most reported being Facebook, blogs, Twitter, and YouTube (the full list is provided in Table 1**)**. The study samples included blogs/forum discussions in which the participants were the general public, patients, and/or health professionals (Multimedia Appendix 2). There was a wide range of health topics, but the most frequently reported on were sexual health [45,46,70,104,107,115,117], diabetes [47,60,68,110,116], flu/H1N1 [54,57,58,111], and mental health issues such as stress or depression [48,83,101].

**Table 1.** Social media tools/applications within the 98 studies[a].

| | |
|---|---|
| Facebook (n=13) | Farmer et al (2009) [37], Ahmed et al (2010) [51], Greene et al (2010) [60], Bender et al (2011) [78], Egan & Moreno (2011a) [83], Egan & Moreno (2011b) [84], Frimmings et al (2011) [86], Gajara et al (2011)[88], Garcia-Romero et al (2011) [89], Jent et al (2011) [92], Kukreja et al (2011) [95], Lord et al (2011) [99], Sajadi & Goldman (2011) [105] |
| Blogs (n=13) | Adams (2008) [26], Kovic et al (2008) [29], Lagu et al (2008) [30], Tan (2008) [32], Denecke & Nedjl (2009) [36], Keelan et al (2009) [41], Kim (2009) [42], Adams (2010) [50], Clauson et al (2010) [53], Hu & Sundar (2010) [61], Sanford (2010) [71], Shah & Robinson (2011) [109], Marcus et al (2012) [119] |
| Twitter (n=8) | Chew & Eysenbach (2010) [54], Scanfeld et al (2010) [72], Heavillin et al (2011) [91], Kukreja et al (2011) [95], Sajadi & Goldman (2011) [105], Salathe & Khandelwal (2011) [106], Signorini et al (2011) [111], Turner-McGrievy & Tate (2011) [112] |
| YouTube (n=7) | Freeman & Chapman (2007) [24], Fernandez-Luque et al (2009) [38], Lo et al (2010) [67], Tian (2010) [74], Chou et al (2011) [80], Sajadi & Goldman (2011) [105], Fernandez-Luque et al (2012) [118] |
| MySpace (n=5) | Moreno et al (2007) [25], Moreno et al (2009a) [44], Moreno et al (2009b) [45], Versteeg et al (2009) [49], Ralph et al (2011) [104] |
| PatientsLikeMe (n=4) | Frost et al (2008) [28], Wicks et al (2010) [75], Doing-Harris & Zeng Treitler (2011) [81], Frost et al (2011) [87] |
| Wikipedia (n=3) | Clauson et al (2008) [27], Morturu & Liu (2011) [100], Rajagopalan et al (2011) [103] |
| Wiki (n=2) | Denecke & Nedjl (2009) [36], Adams (2010) [50] |
| Quitnet / online smoking cessation support group (n=2) | Cobb et al (2010) [55], Selby et al (2010) [73] |
| Physician rating website (not specified) (n=2) | Lagu (2010) [65], Kadry et al (2011) [93] |
| Second Life (n=1) | Beard et al (2009) [34] |
| Daily Strength (n=1) | Morturu & Liu (2011) [100] |
| ArboAntwoord (n=1) | Rhebergen et al (2012) [121] |
| Social media (tool not specified) (n=30) | Chou et al (2009) [35], Jennings et al (2009) [40], Takahashi et al (2009) [48], Avery et al (2010) [52], Colineau & Paris (2010) [56], Corley et al (2010) [57], Ding & Zhang (2010) [58], Hwang et al (2010) [62], Kim & Kwon (2010) [63], Kontos et al (2010) [64], Lariscy et al (2010) [66], Orizio et al (2010) [69], Rice et al (2010) [70], Adrie et al (2011) [76], Baptist et al (2011) [77], Bosslett et al (2011) [79], Dowdell et al (2011) [82], Friedman et al (2011) [85], Hanson et al (2011) [90], Kishimoto & Fukushmima (2011), [94], Lariscy et al (2011) [96], Liang & Scammon (2011) [98], O'Dea & Campbell (2011) [101], Omurtag et al (2011) [102], Selkie et al (2011) [107], Setoyama et al (2011) [108], Shrank et al (2011) [110], Veinot et al (2011) [115], Weitzman et al (2011) [116], Young & Rice (2011) [117], O'Grady et al (2012) [120] |
| Web 2.0 application (not specified) (n=11) | Scotch et al (2008) [31], Timpka et al (2008) [33], Hughes et al (2009) [39], Lupianez-Villanueva et al (2009) [43], Moen et al (2009) [44], Nordqvist et al (2009) [47], Ekberg et al (2010) [59], Nordfeldt et al (2010) [68], Lau (2011) [97], Usher et al (2011) [113], Van Uden-Kraan (2011) [114] |

[a]Some studies included more than one social media tool/application.

## Methodological Quality of Studies

From the searches between January 2002 and February 2012, the selected studies in this review were published from 2007 to 2012 with the vast majority in the last 2 years (Table 1). From the available methodology of bias tools/quality scales, the Downs and Black Instrument [123] has been previously identified as a recommended tool to evaluate the quality of both quantitative randomized and nonrandomized studies [124]. As there are no standard accepted quality scales for studies of proportions [125], only quantitative studies (including mixed methods) were evaluated. Using this Downs and Black instrument [123], the maximum total score that could be achieved was 32, but the scores of the studies in this review ranged from 3 [89] to 22 [121]. Overall, the studies scored low using this scale as they were mainly exploratory and descriptive with three intervention studies [43,112,121] and one randomized controlled trial (RCT) [45]. From the 98 studies, 40 were applied quantitative, 48 qualitative (including studies with content analysis presenting data with descriptive statistics), and 10 mixed methods (both quantitative and quantitative). These studies are presented by methodology in Table 2. Methodological bias of the selected studies using the Downs and Black instrument [123] is presented in Multimedia Appendix 3.

## Characteristics/Profile of Users Accessing Social Media for Health Communication

The characteristics of users of social media for health communication in the selected studies were diverse, covering a range of different population groups. The age of the social media users ranged from school children to older adults aged 65 years and up [24-121], but the majority of the reported ages were 11-34 years [25,35,45,46,64,77,104,107,119]. Some studies reported that there were more female than male users of social network sites [35,40,55,62,64]. A few studies found that social media users were disproportionately from lower-income

households [35,64,72]. Studies within the United States reported that more social media users were African Americans than nonHispanic Whites [35,40]. Chou et al [35] concluded that the population is accessing social media regardless of education and race/ethnicity.

### Uses of Social Media for Health Communication

From the selected studies, seven key uses of social media for health communication were identified for the general public, patients, and health professionals (Table 3). Social media provided health information on a range of conditions to the general public [36,61,71,74,103], patients [47,63,71,75,98,103], and health professionals [36,47,98]. This communication can provide answers to medical questions [34,36,60]. Social media allows information to be presented in modes other than text and can bring health information to audiences with special needs; for example, videos can be used to supplement or replace text and can be useful when literacy is low [50]. A range of social media platforms can facilitate dialogue between patients and patients, and patients and health professionals [56,79,110]. Sites such as PatientsLikeMe enable patients to engage in dialogue with each other and share health information and advice including information on treatment and medication [28,75]. YouTube has been used by the general public to share health information on medications, symptoms, and diagnoses [38], and by patients to share personal cancer stories [80]. Blog sites create a space where individuals can access tailored resources [26] and provide health professionals with an opportunity to share information with patients and members of the public [30,56]. Facebook is being used by the general public, patients, carers, and health professionals to share their experience of disease management, exploration, and diagnosis [37]. Asthma groups are using MySpace to share health information, in particular personal stories and experiences [49,60]. Social media can be used to collect data on patient experiences and opinions such as physician's performance [26,65,109].

Social media have been used for health promotion and health education [25,34,46,59,82,90,113,117] and for delivering a health intervention by providing social support/influence to promote smoking cessation and abstinence [55]. A study has shown that social media can reduce stigma about certain conditions such as epilepsy [67]. In addition, there were some opportunities for health professionals to have online consultations [88].

### Benefits of Social Media for Health Communication

Six overarching benefits of social media for health communication were identified for the general public, patients, and health professionals (Table 4). Social media users have the potential to increase the number of interactions and thus are provided with more available, shared, and tailored information. Social media can generate more available health information as users create and share medical information online [50]. Blog sites create a space where individuals can access tailored resources to deal with health issues [26]. Social media can widen access to those who may not easily access health information via traditional methods, such as younger people, ethnic minorities, and lower socioeconomic groups [35,64,66,83,84,86,99,104,107,115]. An important aspect of using social media for health communication is that it can provide valuable peer, social, and emotional support for the general public [37,43,44,47,48,51,56,73,101,108,114] and patients [28,33,34,44,47,48,56,60,62,68,71,76,88,98,120]. For example, social media can aid health behavior change such as smoking cessation [53,73], and PatientsLikeMe enables patients to communicate with other patients and share information about health issues [28]. Colineau and Paris [56] reported that people used health-related social networking sites to discuss sensitive issues and complex information with health professionals.

In public health surveillance, social media can provide communication in real time and at relatively low cost [31,40,54,57,72,111,116]. Social media can monitor public response to health issues [54], track and monitor disease outbreak [111], identify misinformation of health information [72], identify target areas for intervention efforts [106], and disseminate pertinent health information to targeted communities [57]. Health professionals can aggregate data about patient experiences from blogs and monitor public reaction to health issues [26,40]. Social media may have particular potential for risk communications as they can be used to disseminate personalized messages immediately thus making outreach more effective [58]. There is the potential that information on social media may contribute to health care policy making, as medical blogs are frequently viewed by mainstream media [29].

**Table 2.** List of studies by methodology—quantitative, qualitative, or both (n=98).

| Quantitative (n=40) | Qualitative (n=48) | Mixed methods (n=10) |
|---|---|---|
| Kovic et al (2008) [29] | Freeman & Chapman (2007)[a][24] | Clauson et al (2008) [27] |
| Chou et al (2009) [35] | Moreno et al (2007)[a][25] | Timpka et al (2008) [33] |
| Moreno et al (2009a) [45] | Adams (2008) [26] | Hughes et al (2009) [39] |
| Avery et al (2010) [52] | Frost et al (2008) [28] | Jennings et al (2009) [40] |
| Chew & Eysenbach (2010) [54] | Lagu et al (2008)[a][30] | Lupianez-Villanueva et al (2009) [43] |
| Cobb et al (2010) [55] | Scotch et al (2008) [31] | Takahashi et al (2009) [48] |
| Colineau & Paris (2010) [56] | Tan (2008) [32] | Hwang et al (2010) [62] |
| Hu & Sundar (2010) [61] | Beard et al (2009) [34] | Ralph et al (2011) [104] |
| Kim & Kwon (2010) [63] | Denecke & Nedjl (2009)[a][36] | Selkie et al (2011) [107] |
| Kontos et al (2010) [64] | Farmer et al (2009)[a][37] | O'Grady et al (2012) [120] |
| Lariscy et al (2010) [66] | Fernandez-Luque et al (2009)[a][38] | |
| Lo et al (2010) [67] | Keelan et al (2009)[a][41] | |
| Rice et al (2010) [70] | Kim (2009) [42] | |
| Wicks et al (2010) [75] | Moen et al (2009) [44] | |
| Adrie et al (2011) [76] | Moreno et al (2009b)[a][46] | |
| Baptist et al (2011) [77] | Nordqvist et al (2009) [47] | |
| Bosslett et al (2011) [79] | Versteeg et al (2009)[a][49] | |
| Dowdell et al (2011) [82] | Adams (2010) [50] | |
| Frimmings et al (2011) [86] | Ahmed et al (2010)[a][51] | |
| Garcia-Romero et al (2011) [89] | Clauson et al (2010) [53] | |
| Hanson et al (2011) [90] | Corley et al (2010) [57] | |
| Jent et al (2011) [92] | Ding & Zhang (2010)[a][58] | |
| Kadry et al (2011) [93] | Ekberg (2010) [59] | |
| Kishimoto & Fukushmima (2011) [94] | Greene et al (2010)[a][60] | |
| Kukreja et al (2011) [95] | Lagu (2010)[a][65] | |
| Lau (2011) [97] | Nordfeldt et al (2010) [68] | |
| Lord et al (2011) [99] | Orizio et al (2010)[a][69] | |
| Morturu & Liu (2011) [100] | Sanford (2010) [71] | |
| O'Dea & Campbell (2011) [101] | Scanfeld et al (2010) [72] | |
| Omurtag et al (2011) [102] | Selby et al (2010)[a][73] | |
| Rajagopalan et al (2011) [103] | Tian (2010) [74] | |
| Setoyama et al (2011) [108] | Bender et al (2011)[a][78] | |
| Signorini et al (2011) [111] | Chou et al (2011)[b][80] | |
| Turner-McGrievy & Tate (2011) [112] | Doing-Harris & Zeng-Treitler (2011) [81] | |
| Usher et al (2011) [113] | Egan & Moreno (2011a)[a][83] | |
| Van Uden-Kraan (2011) [114] | Egan & Moreno (2011b)[a][84] | |
| Weitzman et al (2011) [116] | Friedman et al (2011)[a][85] | |
| Young & Rice (2011) [117] | Frost et al (2011) [87] | |

| Quantitative (n=40) | Qualitative (n=48) | Mixed methods (n=10) |
|---|---|---|
| Fernandez-Luque et al (2012) [118] | Gajaria et al (2011) [88] | |
| Rhebergen et al (2012) [121] | Heavillin et al (2011) [91] | |
| | Lariscy et al (2011) [96] | |
| | Liang & Scammon (2011) [98] | |
| | Sajadi & Goldman (2011) [105] | |
| | Salthe & Khandelwal (2011)[a][106] | |
| | Shah & Robinson (2011) [109] | |
| | Shrank et al (2011)[a][110] | |
| | Veinot et al (2011) [115] | |
| | Marcus et al (2012) [119] | |

[a] Qualitative study using content analysis with some findings reported as descriptive statistics.

[b] Descriptive statistics.

## Limitations of Using Social Media for Health Communication

There were 12 limitations of social media for health communication (Table 5). The main recurring limitations of social media are quality concerns [26,39,42,44,47,50,69,85] and the lack of reliability of the health information [26,37,39,40,42,44,47,50,69,74,85,95]. The authors of websites are often unidentifiable, or there can be numerous authors, or the line between producer and audience is blurred [38,50,74]. Thus it is more difficult for individuals to discern the reliability of information found online [50,38]. Regulations may not facilitate health professionals to communicate with patients online, for example, email is not an official medical record and could be vulnerable to security breaches [68]. Policy reactions to address concerns include providing training in how to use and navigate social media technologies and validate accuracy of information found [39,66], or bringing more credible sites into the mainstream and making them fully accessible [39].

The large volume of information available through social media and the possibility for inaccuracies posted on these sites presents challenges when validating information [26]. Several studies highlighted concerns about privacy and confidentiality, data security, and the potential harms that emerge when personal data are indexed [38,44,47,50]. Social media users are often unaware of the risks of disclosing personal information online [26] and with communicating harmful or incorrect advice using social media [26,50]. As information is readily available, there is the potential of information overload for the user [50]. The general public may not know how to correctly apply information found online to their personal health situation [50]. There is the potential that adverse health consequences can result from information found on social media sites, for example, pro-smoking imagery [24]. In addition, there may be negative health risk behaviors displayed online, such as unsafe sexual behavior [45,46]. There is limited evidence that engaging in online communities positively impacts people's health [56]. Health professionals may not often use social media to communicate with their patients [42]. There is also the possibility that social media may act as a deterrent for patients from visiting health professionals [42].

**Table 3.** Uses of social media for health communication among the general public, patients, and health professionals.

| Uses of social media for health communication | Social media user | | |
|---|---|---|---|
| | General Public | Patients | Health Professionals |
| Provide health information on a range of conditions | ✓ | ✓ | ✓ |
| Provide answers to medical questions | ✓ | ✓ | ✓ |
| Facilitate dialogue between patients to patients, and patients and health professionals | | ✓ | ✓ |
| Collect data on patient experiences and opinions | | ✓ | ✓ |
| Used for health intervention, health promotion and health education | ✓ | ✓ | ✓ |
| Reduce stigma | | ✓ | ✓ |
| Provide online consultations | | ✓ | ✓ |

**Table 4.** Benefits of using social media for health communication for the general public, patients, and health professionals.

| Benefits of social media for health communication | Social media user | | |
|---|---|---|---|
| | General Public | Patients | Health Professionals |
| Increase interactions with others | ✓ | ✓ | ✓ |
| More available, shared, and tailored information | ✓ | ✓ | ✓ |
| Increase accessibility & widening access | ✓ | ✓ | ✓ |
| Peer/social/emotional support | ✓ | ✓ | ✓ |
| Public health surveillance | ✓ | ✓ | ✓ |
| Potential to influence health policy | ✓ | ✓ | ✓ |

**Table 5.** Limitations of social media for health communication among the general public, patients, and health professionals.

| Limitations of social media for health communication | Social media user | | |
|---|---|---|---|
| | General Public | Patients | Health Professionals |
| Lack of reliability | ✓ | ✓ | ✓ |
| Quality concerns | ✓ | ✓ | ✓ |
| Lack of confidentiality & privacy | ✓ | ✓ | ✓ |
| Often unaware of the risks of disclosing personal information online | ✓ | ✓ | |
| Risks associated with communicating harmful or incorrect advice using social media | ✓ | ✓ | |
| Information overload | ✓ | ✓ | |
| Not sure how to correctly apply information found online to their personal health situation | ✓ | ✓ | |
| Certain social media technologies may be more effective in behavior change than others | ✓ | | |
| Adverse health consequences | ✓ | | |
| Negative health behaviors | ✓ | | |
| Social media may act as a deterrent for patients from visiting health professionals | | ✓ | ✓ |
| Currently may not often use social media to communicate to patients | | | ✓ |

## Discussion

The 98 research studies in this review provided evidence that social media (most reported applications were Facebook, Blogs, Twitter, and YouTube) can create a space to share, comment, and discuss health information on a diverse range of health issues such as sexual health, diabetes, flu/H1N1, and mental health issues [24-121]. Social media attracts a large number of users thus creating a platform for mass health communication [35] with identified uses, benefits, and limitations for the general public, patients, and health professionals.

### Uses of Social Media for Health Communication

The main uses of social media focus on increasing interactions with others, and facilitating, sharing, and obtaining health messages [24-121]. The general public mainly use social media for themselves, family members, and/or friends to obtain and share information on a wide range of health issues [36,60,61,71,74,103]. Patients can share their experiences through discussion forums, chat rooms and instant messaging, or online consultation with a qualified clinician [26,62,63]. Some health professionals were reported to use social media to collect data on patients [26,65] and to communicate with patients using online consultations [88]; however, this latest use is limited. Recent research reported that female health professionals in Quebec, Canada, believed that Web 2.0 may be a useful mechanism for knowledge transfer but is limited due to their lack of time and technological skills [126]. Perhaps in light of Kaplan and Haenlein's [10] classifications of social media, further work on improving the "social presence", the closeness to synchronous face-to-face communication of such online consultations, would contribute to improving communication between health professionals and patients. Another recent study applied social network analysis to understand the knowledge sharing behavior of practitioners in a clinical online discussion forum and found that although their number is limited, interprofessional and interinstitutional ties are strong [127]. This relates to Gilbert and Karahalios' [128] social tie analysis and suggests that development of mechanisms that evaluate tie strength in social media that in turn impact on its functionality may be useful for health communication. Further technological advances will provide more opportunities to use social media in health care in the future, especially between patients and patients, and also health professionals and

patients. However, both patients and health professionals may require training to fully maximize the uses of using social media in health care.

### Benefits of Social Media for Health Communication

Numerous benefits of using social media for health communication were reported for the general public, patients, and health professionals. A major benefit of social media for health communication is the accessibility and widening access of health information to various population groups, regardless of age, education, race or ethnicity, and locality, compared to traditional communication methods [35,64,95,72]. While these changing patterns may lessen health disparities, traditional inequalities in overall Internet access remain. Furthermore, variation in social media engagement according to personality traits, age, and gender [15] suggests the need for ongoing scrutiny regarding equality of access and effectiveness for different users. Social media can be used to provide a valuable and useful source of peer, social, and emotional support to individuals, including those with various conditions/illnesses [48,62,71]. Hwang and colleagues [62] reported that encouragement, motivation, and shared experience were important social support features of social media sites.

Social media allows users to generate peer-to-peer discussion in a way not enabled by traditional websites [48,50,62,71]. However, this may challenge expectations, relationships, quality, and consistent health care practice. As Moen et al [44] explain, current patterns of collaboration tend to produce an asymmetric patient-health care provider relationship. This highlights a strong need for health providers to maintain a role within social media health communication that is not simply the same as that of patient and general public users. Keitzmann et al [17] have suggested that organizations need to recognize and understand the social media landscape, and where the conversations about them are already being held (cognize), develop strategies that are suitable, work out how often and when they should enter into conversations, and be aware of what others are doing and act accordingly. This review highlights clearly that social media has benefits for health communication but the long-term effects are not known. As the use of social media is expected to increase in the future [1], there may be further benefits of using social media in health care. It is not yet known how effective social media applications are in health communications, which warrants further research.

### Limitations of Social Media for Health Communication

Social media tools remain informal, unregulated mechanisms for information collection, sharing, and promotion, so the information is of varying quality and consistency [26,27,39,40,42,44,47,50,69,74,85,95]. Similar issues exist with traditional Internet sites, but these issues are being heightened by the interactive nature of social media, which allows lay-users to upload information regardless of quality [50]. Reliability may be monitored by responsible bodies using automated processes, employed to signal when content has been significantly edited, and progress is being made in automated quality detection [50]. Further work to improve the “media richness” [10] of social media for health communication, that is, how they may reduce ambiguity and uncertainty, would be valuable. In addition, combining more resources in one site could improve reliability of information. As patients interact and share links, they could compare numerous social media sites and triangulate information to help them discern correct from incorrect information [50]. Despite concerns, information found on some websites is reported to be generally factually accurate [39,62]. A further limitation is that postings can be a permanent record and be viewed by an increasing audience, and perhaps users are unaware of the potential size of the audience base. Regulatory and security issues must be addressed to broach a way forward for best-practice that allows the benefits of social media to be utilized yet still protects patients’ privacy and to therefore improve use of these media in routine clinical care. This is a public policy issue and is already being contested in the United States. Public education is required for the general public, patients, and health professionals to make them more aware of the nature of using social media. Consideration of the variation in social media engagement according to personality traits, age, and gender [15] will be valuable in tailoring education to meet the needs of population groups.

### Gaps in the Research Literature and Recommendations of Research Into Social Media for Health Communication

This literature review has shown that the general public, patients, and health professionals use social media in health care for various purposes with numerous benefits and limitations. The current research’s methodological scoring was low; this was mainly due to the fact that the vast majority of the studies in this review were exploratory and descriptive. To date, there is very limited evidence from RCTs and longitudinal studies. To more fully determine the role of social media for health communication, further research with larger sample sizes and more robust methodologies are required. Based on this review [24-121], several gaps in the literature have been identified that need to be addressed:

- the impact of social media for health communication in specific population groups, such as minority groups, patients groups, culture differences;
- the relative effectiveness of different applications of social media for health communication;
- the longer-term impact on the effectiveness of social media for health communication;
- the most suitable mechanisms to monitor and enhance the quality and reliability of health communication using social media;
- the risks arising from sharing information online, the consequences for confidentiality and privacy, and the most suitable mechanisms for effectively educating users in the maintenance of their confidentiality and privacy;
- the full potential of social media in effectively supporting the patient-health professional relationship;
- the impact of peer-to-peer support for the general public, patients, and health professionals to enhance their interpersonal communication;
- the impact of social media on behavior change for healthy lifestyles.

To address these gaps in the literature, the key recommendations for future health communication research focus on robust and comprehensive evaluation and review, using a range of methodologies. The research priorities are highlighted below:

- To determine the impact of social media for health communication in specific population groups with large sample sizes (representation of population groups).
- To determine the relative effectiveness of different social media applications for health communication using RCTs.
- To determine the longer-term impact on the effectiveness of social media for health communication using longitudinal studies.
- To explore potential mechanisms for monitoring and enhancing the quality and reliability of health communication using social media.
- To investigate the risks arising from sharing information online and the consequences for confidentiality and privacy, coupled with developing the most suitable mechanisms to effectively educate users in the maintenance of their confidentiality and privacy.
- To determine how social media can be effectively used to support the patient-health professional relationship.
- To determine the impact of peer-to-peer support for the general public, patients, and health professionals to enhance their interpersonal communication.
- To explore the potential for social media to lead to behavior change for healthy lifestyles to inform health communication practice.

## Conclusions

Social media brings a new dimension to health care, offering a platform used by the public, patients, and health professionals to communicate about health issues with the possibility of potentially improving health outcomes. Although there are benefits to using social media for health communication, the information needs to be monitored for quality and reliability, and the users' confidentiality and privacy need to be maintained. Social media is a powerful tool that offers collaboration between users and a social interaction mechanism for a range of individuals. With increasing use of social media, there will be further opportunities in health care. Research into the application of social media for health communication purposes is an expanding area because increasing general use of social media necessitates that health communication researchers match the pace of development. Further robust research is required to establish whether social media improves health communication practices in both the short and long terms.

## Authors' Contributions

Dr. Anne Moorhead developed the concept of this paper and selected and evaluated papers and led this manuscript. Laura Harrison conducted the searches for studies, and Dr. Anne Moorhead and Laura Harrison evaluated the papers. All authors evaluated the studies and contributed to this manuscript.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Excluded studies.

[PDF File (Adobe PDF File), 250KB - jmir_v15i4e85_app1.pdf ]

## Multimedia Appendix 2

Summary of the selected studies (n=98).

[PDF File (Adobe PDF File), 518KB - jmir_v15i4e85_app2.pdf ]

## Multimedia Appendix 3

Study quality scores using Downs and Black scale: checklist for measuring study quality (n=50).

[PDF File (Adobe PDF File), 345KB - jmir_v15i4e85_app3.pdf ]

## References

1. Boyd DM, Ellison NB. Social network sites: Definition, history, and scholarship. J Comp Med Commun 2008;13:210-230. [doi: 10.1111/j.1083-6101.2007.00393.x]
2. Thackeray R, Neiger BL, Hanson CL, McKenzie JF. Enhancing promotional strategies within social marketing programs: use of Web 2.0 social media. Health Promot Pract 2008 Oct;9(4):338-343. [doi: 10.1177/1524839908325335] [Medline: 18936268]
3. Eyrich N, Padman ML, Sweetser DS. PR practitioners' use of social media tools and communication technology. Public Relations Review 2008;34:412-414.

4. McNab C. What social media offers to health professionals and citizens. 2009. URL: http://www.who.int/bulletin/volumes/87/8/09-066712/en/ [accessed 2013-04-04] [WebCite Cache ID 6FdBA9AuG]
5. Green B, Hope A. Promoting clinical competence using social media. Nurse Educ 2010;35(3):127-129. [doi: 10.1097/NNE.0b013e3181d9502b] [Medline: 20410751]
6. Giustini D. How Web 2.0 is changing medicine. BMJ 2006 Dec 23;333(7582):1283-1284 [FREE Full text] [doi: 10.1136/bmj.39062.555405.80] [Medline: 17185707]
7. Fox S, Jones S. The Social Life of Health Information. 2009. URL: http://www.pewinternet.org.Reports/2007/Information-Searches- [accessed 2013-04-04] [WebCite Cache ID 6FdBDPOJA]
8. Heidelberger CA. Health Care Professionals' Use of Online Social Networks. 2011. URL: http://cahdsu.wordpress.com/2011/04/07/infs-892-health-care-professionals-use-of-online-social-networks/ [accessed 2013-04-04] [WebCite Cache ID 6FdBOyNJt]
9. Dawson J. Doctors join patients in going online for health information. New Media Age 2010;7.
10. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of social media. Business Horizons 2010;53:59-68.
11. Short J, Williams E, Christie B. The social psychology of telecommunications. Hoboken, NJ: John Wiley & Sons, Ltd; 1976.
12. Goffman E. The presentation of self in everyday life. New York: Doubleday Anchor Books; 1959.
13. Maness JM. Library 2.0 Theory: Web 2.0 and its implications for libraries. 2006. URL: http://www.webology.org/2006/v3n2/a25.html [accessed 2013-04-04] [WebCite Cache ID 6FdBicNxh]
14. Kamel Boulos MN, Wheeler S. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. Health Info Libr J 2007 Mar;24(1):2-23. [doi: 10.1111/j.1471-1842.2007.00701.x] [Medline: 17331140]
15. Correa T, Willard Hinsley A, de Zúñiga HG. Who interacts on the Web?: The intersection of users' personality and social media use. Computers in Human Behavior 2010;26(2):247-253.
16. Lenhart A, Purcell K, Smith A, Zickuhr K. Social Media & Mobile Internet Use Among Teens and Young Adults. USA: Pew Research Center; 2010. URL: http://www.pewinternet.org/~/media/Files/Reports/2010/PIP_Social_Media_and_Young_Adults_Report_Final_with_toplines.pdf [accessed 2013-04-04] [WebCite Cache ID 6FdBrFKd9]
17. Kietzmann JH, Hermkens K, McCarthy IP, Silvestre BS. Social media? Get serious! Understanding the functional building blocks of social media. Business Horizons 2011;54(3):241-251.
18. Butterfield S. Social software. 2003. URL: http://www.sylloge.com/personal/2003_03_01_s.html [accessed 2013-04-04] [WebCite Cache ID 6FdBtXa5e]
19. Morville P. User experience design. 2004. URL: http://semanticstudios.com/publications/semantics/000029.php [accessed 2013-04-04] [WebCite Cache ID 6FdBwP2d9]
20. Webb M. Social software consultancy. 2004. URL: http://interconnected.org/home/2004/04/28/on_social_software [accessed 2013-04-04] [WebCite Cache ID 6FdByLPat]
21. Smith G. Social software building blocks. 2007. URL: http://nform.ca/publications/social-software-building-block [accessed 2013-04-04] [WebCite Cache ID 6FdBzbhIe]
22. Mangold WG, Faulds DJ. Social media: The new hybrid element of the promotion mix. Business Horizons 2009;52(4):357-365.
23. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 2009;6(6):e1000097. [doi: 10.1371/journal.pmed1000097]
24. Freeman B, Chapman S. Is "YouTube" telling or selling you something? Tobacco content on the YouTube video-sharing website. Tob Control 2007 Jun;16(3):207-210 [FREE Full text] [doi: 10.1136/tc.2007.020024] [Medline: 17565142]
25. Moreno MA, Parks M, Richardson LP. What are adolescents showing the world about their health risk behaviors on MySpace? MedGenMed 2007;9(4):9. [Medline: 18311359]
26. Adams SA. Blog-based applications and health information: two case studies that illustrate important questions for Consumer Health Informatics (CHI) research. Int J Med Inform 2010 Jun;79(6):e89-e96. [doi: 10.1016/j.ijmedinf.2008.06.009] [Medline: 18701344]
27. Clauson KA, Polen HH, Boulos MN, Dzenowagis JH. Scope, completeness, and accuracy of drug information in Wikipedia. Ann Pharmacother 2008 Dec;42(12):1814-1821. [doi: 10.1345/aph.1L474] [Medline: 19017825]
28. Frost JH, Massagli MP. Social uses of personal health information within PatientsLikeMe, an online patient community: what can happen when patients have access to one another's data. J Med Internet Res 2008 May;10(3):e15 [FREE Full text] [doi: 10.2196/jmir.1053] [Medline: 18504244]
29. Kovic I, Lulic I, Brumini G. Examining the medical blogosphere: an online survey of medical bloggers. J Med Internet Res 2008 Sep;10(3):e28 [FREE Full text] [doi: 10.2196/jmir.1118] [Medline: 18812312]
30. Lagu T, Kaufman EJ, Asch DA, Armstrong K. Content of weblogs written by health professionals. J Gen Intern Med 2008 Oct;23(10):1642-1646 [FREE Full text] [doi: 10.1007/s11606-008-0726-6] [Medline: 18649110]

31. Scotch M, Yip KY, Cheung KH. Development of grid-like applications for public health using Web 2.0 mashup techniques. J Am Med Inform Assoc 2008 Aug;15(6):783-786 [FREE Full text] [doi: 10.1197/jamia.M2731] [Medline: 18755998]
32. Tan L. Psychotherapy 2.0: MySpace blogging as self-therapy. Am J Psychother 2008;62(2):143-163. [Medline: 18605128]
33. Timpka T, Eriksson H, Ludvigsson J, Ekberg J, Nordfeldt S, Hanberger L. Web 2.0 systems supporting childhood chronic disease management: a pattern language representation of a general architecture. BMC Med Inform Decis Mak 2008;8:54 [FREE Full text] [doi: 10.1186/1472-6947-8-54] [Medline: 19040738]
34. Beard L, Wilson K, Morra D, Keelan J. A survey of health-related activities on second life. J Med Internet Res 2009 May;11(2):e17 [FREE Full text] [doi: 10.2196/jmir.1192] [Medline: 19632971]
35. Chou WY, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009 Nov;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
36. Denecke K, Nejdl W. How valuable is medical social media data? Content analysis of the medical web. Inform Sciences 2009;179:1870-1880.
37. Farmer AD, Bruckner Holt CE, Cook MJ, Hearing SD. Social networking sites: a novel portal for communication. Postgrad Med J 2009 Sep;85(1007):455-459. [doi: 10.1136/pgmj.2008.074674] [Medline: 19734511]
38. Fernandez-Luque L, Elahi N, Grajales FJ. An analysis of personal medical information disclosed in YouTube videos created by patients with multiple sclerosis. Stud Health Technol Inform 2009;150:292-296. [Medline: 19745316]
39. Hughes B, Joshi I, Lemonde H, Wareham J. Junior physician's use of Web 2.0 for information seeking and medical education: a qualitative study. Int J Med Inform 2009 Oct;78(10):645-655. [doi: 10.1016/j.ijmedinf.2009.04.008] [Medline: 19501017]
40. Jennings A, Powell J, Armstrong N, Sturt J, Dale J. A virtual clinic for diabetes self-management: pilot study. J Med Internet Res 2009 Mar;11(1):e10 [FREE Full text] [doi: 10.2196/jmir.1111] [Medline: 21821504]
41. Keelan J, Pavri V, Balakrishnan R, Wilson K. An analysis of the Human Papilloma Virus vaccine debate on MySpace blogs. Vaccine 2010 Feb 10;28(6):1535-1540. [doi: 10.1016/j.vaccine.2009.11.060] [Medline: 20003922]
42. Kim S. Content analysis of cancer blog posts. J Med Libr Assoc 2009 Oct;97(4):260-266 [FREE Full text] [doi: 10.3163/1536-5050.97.4.009] [Medline: 19851489]
43. Lupiáñez-Villanueva F, Mayer MA, Torrent J. Opportunities and challenges of Web 2.0 within the health care systems: an empirical exploration. Inform Health Soc Care 2009 Sep;34(3):117-126. [doi: 10.1080/17538150903102265] [Medline: 19670002]
44. Moen A, Smørdal O, Sem I. Web-based resources for peer support - opportunities and challenges. Stud Health Technol Inform 2009;150:302-306. [Medline: 19745318]
45. Moreno MA, Vanderstoep A, Parks MR, Zimmerman FJ, Kurth A, Christakis DA. Reducing at-risk adolescents' display of risk behavior on a social networking web site: a randomized controlled pilot intervention trial. Arch Pediatr Adolesc Med 2009 Jan;163(1):35-41. [doi: 10.1001/archpediatrics.2008.502] [Medline: 19124701]
46. Moreno MA, Parks MR, Zimmerman FJ, Brito TE, Christakis DA. Display of health risk behaviors on MySpace by adolescents: prevalence and associations. Arch Pediatr Adolesc Med 2009 Jan;163(1):27-34. [doi: 10.1001/archpediatrics.2008.528] [Medline: 19124700]
47. Nordqvist C, Hanberger L, Timpka T, Nordfeldt S. Health professionals' attitudes towards using a Web 2.0 portal for child and adolescent diabetes care: qualitative study. J Med Internet Res 2009 Apr;11(2):e12 [FREE Full text] [doi: 10.2196/jmir.1152] [Medline: 19403464]
48. 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 Jul;11(3):e29 [FREE Full text] [doi: 10.2196/jmir.1142] [Medline: 19632979]
49. Versteeg KM, Knopf JM, Posluszny S, Vockell AL, Britto MT. Teenagers wanting medical advice: Is MySpace the answer? Arch Pediatr Adolesc Med 2009 Jan;163(1):91-92 [FREE Full text] [doi: 10.1001/archpediatrics.2008.503] [Medline: 19124711]
50. Adams SA. Revisiting the online health information reliability debate in the wake of "web 2.0": an inter-disciplinary literature and website review. Int J Med Inform 2010 Jun;79(6):391-400. [doi: 10.1016/j.ijmedinf.2010.01.006] [Medline: 20188623]
51. Ahmed OH, Sullivan SJ, Schneiders AG, McCrory P. iSupport: do social networking sites have a role to play in concussion awareness? Disabil Rehabil 2010;32(22):1877-1883. [doi: 10.3109/09638281003734409] [Medline: 20367328]
52. Avery E, Lariscy R, Amador E, Ickowitz T, Primm C, Taylor A. Diffusion of Social Media Among Public Relations Practitioners in Health Departments Across Various Community Population Sizes. Journal of Public Relations Research 2010 Jul 2010;22(3):336-358. [doi: 10.1080/10627261003614427]
53. Clauson KA, Ekins J, Goncz CE. Use of blogs by pharmacists. Am J Health Syst Pharm 2010 Dec 1;67(23):2043-2048. [doi: 10.2146/ajhp100065] [Medline: 21098377]
54. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
55. Cobb NK, Graham AL, Abrams DB. Social network structure of a large online community for smoking cessation. Am J Public Health 2010;100(7):1282-1289. [doi: 10.2105/AJPH.2009.165449]

56. Colineau N, Paris C. Talking about your health to strangers: understanding the use of online social networks by patients. New Review of Hypermedia and Multimedia 2010 Apr 2010;16(1-2):141-160. [doi: 10.1080/13614568.2010.496131]
57. Corley CD, Cook DJ, Mikler AR, Singh KP. Text and structural data mining of influenza mentions in Web and social media. Int J Environ Res Public Health 2010 Feb;7(2):596-615 [FREE Full text] [doi: 10.3390/ijerph7020596] [Medline: 20616993]
58. Ding H, Zhang J. Social media and participatory risk communication during the H1N1 flu epidemic: A comparative study. China Media Res 2010;6:80-90.
59. Ekberg J, Ericson L, Timpka T, Eriksson H, Nordfeldt S, Hanberger L, et al. Web 2.0 Systems Supporting Childhood Chronic Disease Management: Design Guidelines Based on Information Behaviour and Social Learning Theories. J Med Syst 2008 Oct 2008;34(2):107-117. [doi: 10.1007/s10916-008-9222-0]
60. Greene JA, Choudhry NK, Kilabuk E, Shrank WH. Online social networking by patients with diabetes: a qualitative evaluation of communication with Facebook. J Gen Intern Med 2011 Mar;26(3):287-292 [FREE Full text] [doi: 10.1007/s11606-010-1526-3] [Medline: 20945113]
61. Hu Y, Sundar S. Effects of online health sources on credibility and behavioral Intentions. Commun Res 2010;37:105-132. [doi: 10.1177/0093650209351512]
62. Hwang KO, Ottenbacher AJ, Green AP, Cannon-Diehl MR, Richardson O, Bernstam EV, et al. Social support in an Internet weight loss community. Int J Med Inform 2010 Jan;79(1):5-13 [FREE Full text] [doi: 10.1016/j.ijmedinf.2009.10.003] [Medline: 19945338]
63. Kim K, Kwon N. Profile of e-patients: analysis of their cancer information-seeking from a national survey. J Health Commun 2010 Oct;15(7):712-733. [doi: 10.1080/10810730.2010.514031] [Medline: 21104502]
64. Kontos EZ, Emmons KM, Puleo E, Viswanath K. Communication inequalities and public health implications of adult social networking site use in the United States. J Health Commun 2010 Dec;15 Suppl 3:216-235 [FREE Full text] [doi: 10.1080/10810730.2010.522689] [Medline: 21154095]
65. Lagu T, Hannon NS, Rothberg MB, Lindenauer PK. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010 Sep;25(9):942-946 [FREE Full text] [doi: 10.1007/s11606-010-1383-0] [Medline: 20464523]
66. Lariscy RW, Reber BH, Paek H. Examination of Media Channels and Types as Health Information Sources for Adolescents: Comparisons for Black/White, Male/Female, Urban/Rural. Journal of Broadcasting & Electronic Media 2010 Mar 2010;54(1):102-120. [doi: 10.1080/08838150903550444]
67. Lo AS, Esser MJ, Gordon KE. YouTube: a gauge of public perception and awareness surrounding epilepsy. Epilepsy Behav 2010 Apr;17(4):541-545. [doi: 10.1016/j.yebeh.2010.02.004] [Medline: 20236867]
68. Nordfeldt S, Hanberger L, Berterö C. Patient and parent views on a Web 2.0 Diabetes Portal--the management tool, the generator, and the gatekeeper: qualitative study. J Med Internet Res 2010;12(2):e17 [FREE Full text] [doi: 10.2196/jmir.1267] [Medline: 20511179]
69. Orizio G, Schulz P, Gasparotti C, Caimi L, Gelatti U. The world of e-patients: A content analysis of online social networks focusing on diseases. Telemed J E Health 2010 Dec;16(10):1060-1066. [doi: 10.1089/tmj.2010.0085] [Medline: 21070131]
70. Rice E, Monro W, Barman-Adhikari A, Young SD. Internet use, social networking, and HIV/AIDS risk for homeless adolescents. J Adolesc Health 2010 Dec;47(6):610-613 [FREE Full text] [doi: 10.1016/j.jadohealth.2010.04.016] [Medline: 21094441]
71. Sanford AA. "I Can Air My Feelings Instead of Eating Them": Blogging as Social Support for the Morbidly Obese. Communication Studies 2010 Nov 2010;61(5):567-584. [doi: 10.1080/10510974.2010.514676]
72. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: twitter and antibiotics. Am J Infect Control 2010 Apr;38(3):182-188 [FREE Full text] [doi: 10.1016/j.ajic.2009.11.004] [Medline: 20347636]
73. Selby P, van Mierlo T, Voci SC, Parent D, Cunningham JA. Online social and professional support for smokers trying to quit: an exploration of first time posts from 2562 members. J Med Internet Res 2010 Aug;12(3):e34 [FREE Full text] [doi: 10.2196/jmir.1340] [Medline: 20719739]
74. Tian Y. Organ donation on Web 2.0: content and audience analysis of organ donation videos on YouTube. Health Commun 2010 Apr;25(3):238-246. [doi: 10.1080/10410231003698911] [Medline: 20461609]
75. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010 Jun;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]
76. Dumaij AC, Tijssen EC. On-line health companion contact among chronically ill in the Netherlands. Health Technol (Berl) 2011 Aug;1(1):5-23 [FREE Full text] [doi: 10.1007/s12553-011-0003-2] [Medline: 21909478]
77. Baptist AP, Thompson M, Grossman KS, Mohammed L, Sy A, Sanders GM. Social media, text messaging, and email-preferences of asthma patients between 12 and 40 years old. J Asthma 2011 Oct;48(8):824-830. [doi: 10.3109/02770903.2011.608460] [Medline: 21864099]
78. Bender JL, Jimenez-Marroquin MC, Jadad AR. Seeking support on facebook: a content analysis of breast cancer groups. J Med Internet Res 2011 Feb;13(1):e16 [FREE Full text] [doi: 10.2196/jmir.1560] [Medline: 21371990]

79. Bosslet GT, Torke AM, Hickman SE, Terry CL, Helft PR. The patient-doctor relationship and online social networks: results of a national survey. J Gen Intern Med 2011 Oct;26(10):1168-1174 [FREE Full text] [doi: 10.1007/s11606-011-1761-2] [Medline: 21706268]
80. Chou WY, Hunt Y, Folkers A, Augustson E. Cancer survivorship in the age of YouTube and social media: a narrative analysis. J Med Internet Res 2011 Jan;13(1):e7 [FREE Full text] [doi: 10.2196/jmir.1569] [Medline: 21247864]
81. Doing-Harris KM, Zeng Treitler Q. Computer-assisted update of a consumer health vocabulary through mining of social network data. J Med Internet Res 2011;13(2):e37. [doi: 10.2196/jmir.1636]
82. Dowdell EB, Burgess AW, Flores JR. Original research: online social networking patterns among adolescents, young adults, and sexual offenders. Am J Nurs 2011 Jul;111(7):28-36; quiz 37. [doi: 10.1097/01.NAJ.0000399310.83160.73] [Medline: 21673563]
83. Egan KG, Moreno MA. Prevalence of stress references on college freshmen Facebook profiles. Comput Inform Nurs 2011 Oct;29(10):586-592 [FREE Full text] [doi: 10.1097/NCN.0b013e3182160663] [Medline: 21436681]
84. Egan KG, Moreno MA. Alcohol references on undergraduate males' Facebook profiles. Am J Mens Health 2011 Sep;5(5):413-420 [FREE Full text] [doi: 10.1177/1557988310394341] [Medline: 21406490]
85. Friedman DB, Koskan A, Rose ID. Prostate cancer guidelines on Web 2.0-based sites: the screening dilemma continues online. J Cancer Educ 2011 Mar;26(1):188-193. [doi: 10.1007/s13187-010-0180-0] [Medline: 21153573]
86. Frimmings RE, Polsgrove MJ, Bower GG. Evaluation of a health and fitness social media experience. Am J Health Educ 2011;42(4):222-227.
87. Frost J, Okun S, Vaughan T, Heywood J, Wicks P. Patient-reported outcomes as a source of evidence in off-label prescribing: analysis of data from PatientsLikeMe. J Med Internet Res 2011;13(1):e6 [FREE Full text] [doi: 10.2196/jmir.1643] [Medline: 21252034]
88. Gajaria A, Yeung E, Goodale T, Charach A. Beliefs about attention-deficit/hyperactivity disorder and response to stereotypes: youth postings in Facebook groups. J Adolesc Health 2011 Jul;49(1):15-20. [doi: 10.1016/j.jadohealth.2010.09.004] [Medline: 21700151]
89. Garcia-Romero MT, Prado F, Dominguez-Cherit J, Hojyo-Tomomka MT, Arenas R. Teledermatology via a social networking web site: a pilot study between a general hospital and a rural clinic. Telemed J E Health 2011 Oct;17(8):652-655. [doi: 10.1089/tmj.2011.0038] [Medline: 21790270]
90. Hanson C, West J, Neiger B, Thackeray R, Barnes M, McIntyre E. Use and acceptance of social media among health educators. Am J Health Educ 2011;42(4):197-204.
91. Heaivilin N, Gerbert B, Page JE, Gibbs JL. Public health surveillance of dental pain via Twitter. J Dent Res 2011 Sep;90(9):1047-1051 [FREE Full text] [doi: 10.1177/0022034511415273] [Medline: 21768306]
92. Jent JF, Eaton CK, Merrick MT, Englebert NE, Dandes SK, Chapman AV, et al. The decision to access patient information from a social media site: what would you do? J Adolesc Health 2011 Oct;49(4):414-420 [FREE Full text] [doi: 10.1016/j.jadohealth.2011.02.004] [Medline: 21939873]
93. Kadry B, Chu LF, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. J Med Internet Res 2011 Nov;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960] [Medline: 22088924]
94. Kishimoto K, Fukushima N. Use of anonymous Web communities and websites by medical consumers in Japan to research drug information. Yakugaku Zasshi 2011;131(5):685-695 [FREE Full text] [Medline: 21532265]
95. Kukreja P, Sheehan AH, Riggins J. Use of social media by pharmacy preceptors. Am J Pharm Educ a176. Accession number 2217 2011;75(9):1104.
96. Lariscy R, Reber B, Paek H. Exploration of health concerns and the role of social media information among rural and urban adolescents: A preliminary study. Int Electr J Hea Educ 2011;14:16-36.
97. Lau AS. Hospital-based nurses' perceptions of the adoption of Web 2.0 tools for knowledge sharing, learning, social interaction and the production of collective intelligence. J Med Internet Res 2011 Nov;13(4):e92 [FREE Full text] [doi: 10.2196/jmir.1398] [Medline: 22079851]
98. Liang B, Scammon DL. E-Word-of-Mouth on health social networking sites: An opportunity for tailored health communication. J Consumer Behav 2011 Dec;10(6):322-331. [doi: 10.1002/cb.378]
99. Lord S, Brevard J, Budman S. Connecting to young adults: an online social network survey of beliefs and attitudes associated with prescription opioid misuse among college students. Subst Use Misuse 2011 Jan;46(1):66-76. [doi: 10.3109/10826084.2011.521371] [Medline: 21190407]
100. Morturu ST, Liu H. Quantifying the trustworthiness of social media content. Distrib Parallel Dat 2011;29:239-260.
101. O'Dea B, Campbell A. Healthy connections: online social networks and their potential for peer support. Stud Health Technol Inform 2011;168:133-140. [Medline: 21893921]
102. Omurtag K, Jimenez PT, Ratts V, Odem R, Cooper AR. The ART of social networking: how SART member clinics are connecting with patients online. Fertil Steril 2012 Jan;97(1):88-94. [doi: 10.1016/j.fertnstert.2011.10.001] [Medline: 22088209]

103. Rajagopalan MS, Khanna VK, Leiter Y, Stott M, Showalter TN, Dicker AP, et al. Patient-oriented cancer information on the internet: a comparison of wikipedia and a professionally maintained database. J Oncol Pract 2011 Sep;7(5):319-323 [FREE Full text] [doi: 10.1200/JOP.2010.000209] [Medline: 22211130]
104. Ralph LJ, Berglas NF, Schwartz SL, Brindis CD. Finding teens in their space: Using social networking sites to connect youth to sexual health services. Sex Res Soc Pol 2011;8:38-49.
105. Sajadi KP, Goldman HB. Social networks lack useful content for incontinence. Urology 2011 Oct;78(4):764-767. [doi: 10.1016/j.urology.2011.04.074] [Medline: 21862115]
106. Salathé M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. PLoS Comput Biol 2011 Oct;7(10):e1002199 [FREE Full text] [doi: 10.1371/journal.pcbi.1002199] [Medline: 22022249]
107. Selkie EM, Benson M, Moreno M. Adolescents' Views Regarding Uses of Social Networking Websites and Text Messaging for Adolescent Sexual Health Education. Am J Health Educ 2011 Dec;42(4):205-212 [FREE Full text] [Medline: 22229150]
108. Setoyama Y, Yamazaki Y, Namayama K. Benefits of peer support in online Japanese breast cancer communities: differences between lurkers and posters. J Med Internet Res 2011;13(4):e122. [doi: 10.2196/jmir.1696]
109. Shah SG, Robinson I. Patients' perspectives on self-testing of oral anticoagulation therapy: content analysis of patients' internet blogs. BMC Health Serv Res 2011;11:25 [FREE Full text] [doi: 10.1186/1472-6963-11-25] [Medline: 21291542]
110. Shrank WH, Choudhry NK, Swanton K, Jain S, Greene JA, Harlam B, et al. Variations in structure and content of online social networks for patients with diabetes. Arch Intern Med 2011 Sep 26;171(17):1589-1591. [doi: 10.1001/archinternmed.2011.407] [Medline: 21949173]
111. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS One 2011 May;6(5):e19467 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]
112. Turner-McGrievy G, Tate D. Tweets, Apps, and Pods: Results of the 6-month Mobile Pounds Off Digitally (Mobile POD) randomized weight-loss intervention among adults. J Med Internet Res 2011 Dec;13(4):e120 [FREE Full text] [doi: 10.2196/jmir.1841] [Medline: 22186428]
113. Usher W. Types of social media (Web 2.0) used by Australian allied health professionals to deliver early twenty-first-century practice promotion and health care. Soc Work Health Care 2011 Apr;50(4):305-329. [doi: 10.1080/00981389.2010.534317] [Medline: 21512953]
114. Van Uden-Kraan CF, Drossaert CH, Taal E, Smit WM, Bernelot Moens HJ, Van de Laar MA. Determinants of engagement in face-to-face and online patient support groups. J Med Internet Res 2011 Dec;13(4):e106 [FREE Full text] [doi: 10.2196/jmir.1718] [Medline: 22155649]
115. Veinot TC, Campbell TR, Kruger D, Grodzinski A, Franzen S. Drama and danger: The opportunities and challenges of promoting youth sexual health through online social networks. In: Proceedings of the AMIA Annual Symposium. Improving Health: Informatics and IT Changing the World. 2011 Presented at: AMIA Annual Symposium. Improving Health: Informatics and IT Changing the World; Oct. 22-26, 2011; Washington, DC.
116. Weitzman ER, Adida B, Kelemen S, Mandl KD. Sharing data for public health research by members of an international online diabetes social network. PLoS One 2011 Apr;6(4):e19256 [FREE Full text] [doi: 10.1371/journal.pone.0019256] [Medline: 21556358]
117. Young SD, Rice E. Online social networking technologies, HIV knowledge, and sexual risk and testing behaviors among homeless youth. AIDS Behav 2011 Feb;15(2):253-260 [FREE Full text] [doi: 10.1007/s10461-010-9810-0] [Medline: 20848305]
118. Fernandez-Luque L, Karlsen R, Melton GB. HealthTrust: a social network approach for retrieving online health videos. J Med Internet Res 2012 Jan;14(1):e22 [FREE Full text] [doi: 10.2196/jmir.1985] [Medline: 22356723]
119. Marcus MA, Westra HA, Eastwood JD, Barnes KL, Mobilizing Minds Research Group. What are young adults saying about mental health? An analysis of Internet blogs. J Med Internet Res 2012 Jan;14(1):e17 [FREE Full text] [doi: 10.2196/jmir.1868] [Medline: 22569642]
120. O'Grady L, Wathen CN, Charnaw-Burger J, Betel L, Shachak A, Luke R, et al. The use of tags and tag clouds to discern credible content in online health message forums. Int J Med Inform 2012 Jan;81(1):36-44. [doi: 10.1016/j.ijmedinf.2011.10.001] [Medline: 22030035]
121. Rhebergen MD, Lenderink AF, van Dijk FJ, Hulshof CT. Comparing the use of an online expert health network against common information sources to answer health questions. J Med Internet Res 2012 Feb;14(1):e9 [FREE Full text] [doi: 10.2196/jmir.1886] [Medline: 22356848]
122. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
123. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. Journal of Epidemiology & Community Health 1998 Jun 1998;52(6):377-384. [doi: 10.1136/jech.52.6.377]
124. Lim SM, Shin ES, Lee SH, Seo KH, Jung YM, Jang JE. Tools for assessing quality and risk of bias by levels of evidence. J Korean Med Assoc 2011 2011;54(4):419. [doi: 10.5124/jkma.2011.54.4.419]

125. National Health and Medical Research Council. How to review the evidence: systematic identification and review of the scientific literature. 2000. URL: http://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/cp65.pdf?q=publications/synopses/_files/cp65 [accessed 2013-04-04] [WebCite Cache ID 6FdFEKvuW]
126. David I, Poissant L, Rochette A. Clinicians' expectations of web 2.0 as a mechanism for knowledge transfer of stroke best practices. J Med Internet Res 2012 Sep;14(5):e121 [FREE Full text] [doi: 10.2196/jmir.2016] [Medline: 23195753]
127. Stewart SA, Abidi SS. Applying social network analysis to understand the knowledge sharing behaviour of practitioners in a clinical online discussion forum. J Med Internet Res 2012 Dec;14(6):e170 [FREE Full text] [doi: 10.2196/jmir.1982] [Medline: 23211783]
128. Gilbert E, Karahalios K. Predicting tie strength with social media. In: Proceedings of the SIGHI Conference on Human Factors in Computing Systems. 2009 Presented at: SIGHI Conference on Human Factors in Computing Systems; April 4-9, 2009; Boston, MA p. 211-220.

## Abbreviations

**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**RCT:** randomized controlled trial

*Edited by G Eysenbach; submitted 13.09.11; peer-reviewed by P Schulz, AM Auvinen, B Crotty; comments to author 04.10.11; revised version received 12.07.12; accepted 26.03.13; published 23.04.13.*

*Please cite as:*
*Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C*
*A New Dimension of Health Care: Systematic Review of the Uses, Benefits, and Limitations of Social Media for Health Communication*
*J Med Internet Res 2013;15(4):e85*
*URL: http://www.jmir.org/2013/4/e85/*
*doi:10.2196/jmir.1933*
*PMID:23615206*

Original Paper

# Web 2.0-Based Crowdsourcing for High-Quality Gold Standard Development in Clinical Natural Language Processing

Haijun Zhai[1*], PhD; Todd Lingren[1*], MA; Louise Deleger[1], PhD; Qi Li[1], PhD; Megan Kaiser[1], BA; Laura Stoutenborough[1], BSN; Imre Solti[1], MD, PhD

Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

[*]these authors contributed equally

**Corresponding Author:**
Imre Solti, MD, PhD
Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center
3333 Burnet Avenue
Cincinnati, OH, 45229
United States
Phone: 1 513 636 1020
Fax: 1 513 636 1020
Email: imre.solti@cchmc.org

## Abstract

**Background:** A high-quality gold standard is vital for supervised, machine learning-based, clinical natural language processing (NLP) systems. In clinical NLP projects, expert annotators traditionally create the gold standard. However, traditional annotation is expensive and time-consuming. To reduce the cost of annotation, general NLP projects have turned to crowdsourcing based on Web 2.0 technology, which involves submitting smaller subtasks to a coordinated marketplace of workers on the Internet. Many studies have been conducted in the area of crowdsourcing, but only a few have focused on tasks in the general NLP field and only a handful in the biomedical domain, usually based upon very small pilot sample sizes. In addition, the quality of the crowdsourced biomedical NLP corpora were never exceptional when compared to traditionally-developed gold standards. The previously reported results on medical named entity annotation task showed a 0.68 F-measure based agreement between crowdsourced and traditionally-developed corpora.

**Objective:** Building upon previous work from the general crowdsourcing research, this study investigated the usability of crowdsourcing in the clinical NLP domain with special emphasis on achieving high agreement between crowdsourced and traditionally-developed corpora.

**Methods:** To build the gold standard for evaluating the crowdsourcing workers' performance, 1042 clinical trial announcements (CTAs) from the ClinicalTrials.gov website were randomly selected and double annotated for medication names, medication types, and linked attributes. For the experiments, we used CrowdFlower, an Amazon Mechanical Turk-based crowdsourcing platform. We calculated sensitivity, precision, and F-measure to evaluate the quality of the crowd's work and tested the statistical significance ($P$<.001, chi-square test) to detect differences between the crowdsourced and traditionally-developed annotations.

**Results:** The agreement between the crowd's annotations and the traditionally-generated corpora was high for: (1) annotations (0.87, F-measure for medication names; 0.73, medication types), (2) correction of previous annotations (0.90, medication names; 0.76, medication types), and excellent for (3) linking medications with their attributes (0.96). Simple voting provided the best judgment aggregation approach. There was no statistically significant difference between the crowd and traditionally-generated corpora. Our results showed a 27.9% improvement over previously reported results on medication named entity annotation task.

**Conclusions:** This study offers three contributions. First, we proved that crowdsourcing is a feasible, inexpensive, fast, and practical approach to collect high-quality annotations for clinical text (when protected health information was excluded). We believe that well-designed user interfaces and rigorous quality control strategy for entity annotation and linking were critical to the success of this work. Second, as a further contribution to the Internet-based crowdsourcing field, we will publicly release the JavaScript and CrowdFlower Markup Language infrastructure code that is necessary to utilize CrowdFlower's quality control and crowdsourcing interfaces for named entity annotations. Finally, to spur future research, we will release the CTA annotations that were generated by traditional and crowdsourced approaches.

***(J Med Internet Res 2013;15(4):e73)*** doi:10.2196/jmir.2426

**KEYWORDS**

clinical informatics; natural language processing; named entity; reference standards; crowdsourcing; user computer interface; quality control

## Introduction

One of the key components of supervised machine learning-based clinical natural language processing (NLP) systems is the high-quality gold standard used for training and testing. In clinical NLP projects, expert annotators are traditionally asked to double annotate the text for the purposes of the gold standard. Expert annotators could be clinicians or extensively trained laypeople [1]. Unless the expert annotators are volunteers, they are very costly to pay and it is usually not easy to build a sufficiently large group of expert annotators locally and, consequently, fast contingent of annotators. To reduce the cost of expert human annotation, many projects in general NLP have turned to crowdsourcing, which involves submitting a large number of smaller subtasks to a coordinated marketplace of workers on the Internet. These workers (called turkers) are paid small amounts (usually a few cents) for each task, sometimes resulting in considerable overall savings over the traditional expert annotator model. The trade-off is usually between the accuracy of the annotation result and the cost savings. Because anonymous turkers from all over the world have different levels of proficiency in the task and are not trained to accomplish the task, efficient quality control and judgment voting methods are required to generate good results.

Many studies have been conducted in the area of crowdsourcing tasks. As early as 2008, Snow et al [2] were the first to explore the feasibility of crowdsourcing in NLP. Five NLP tasks were published on Amazon Mechanical Turk (AMT, [3]) to turkers. Their results indicated that non-expert labellers could obtain high-quality annotations. Since then, data created by crowdsourcing has been widely studied for different research areas. Lawson et al [4] described how using a competitive payment system and interannotator agreement improved the quality of named entity annotations on AMT. Unlike traditional named entity experiments, Finin et al [5] presented their experience by leveraging AMT and CrowdFlower [6] to annotate named entities in Twitter data. It was the first work of named entity recognition in the new domains of Facebook and Twitter. Meanwhile, several studies attempted to use crowdsourcing to create data for machine translation systems. Ambati and Vogel [7] explored the effectiveness of using AMT to do sentence translation for creating parallel corpora. Denkowski et al [8-10] attempted to generate annotated data in a variety of languages. In addition, crowdsourcing was also applied to transcription [11-13], part-of-speech tagging [14], and other tasks [15,16].

In a recent publication of the Journal of Medical Internet Research, Turner et al [17] reported on the use of crowdsourcing to collect feedback on the design of health promotion messages for oral health. Luengo-Oroz et al [18] evaluated the feasibility of crowdsourcing to conduct malaria image analysis. Gathering a large number of high quality annotations is a critical challenge in biomedical NLP, which was presented in detail in the editorial of Chapman et al [1]. As demonstrated by studies in the general NLP field, crowdsourcing is a decidedly promising solution to this research area. However, in contrast to the general NLP domain, there are only a few studies involving crowdsourcing in biomedical NLP and almost none for clinical NLP. Most recently, Burger et al [19] performed a task of extracting the gene-mutation relations in Medical Literature Analysis and Retrieval System Online (MEDLINE) abstracts on AMT. In their work, candidate mutations were extracted from 250 MEDLINE abstracts using the Extractor of Mutations (EMU) presented together with the curated gene lists from the National Center for Biotechnology Information (NCBI). Using a customized interface, it was feasible for turkers to apply their judgments. They reported that the weighted accuracy was 82%. This work was somewhat similar to our linking of medications and their attributes, but it focused on a very specific gene-mutation domain. Norman et al [20] investigated leveraging crowdsourcing to facilitate the discovery of new medicines.

Yetisgen-Yildiz et al [21,22] explored the task of using AMT to annotate biomedical text. Clinical trial announcements annotated 3 entity types (medical conditions, medications, and laboratory tests). The authors indicated that AMT was a very promising tool for annotating clinical text and a well-designed interface and annotation guidelines were helpful to further improve the performance. Building on these earlier works, we designed our medical named entity annotation experiment to include a large-scale data set, easy-to-use graphical user interface and strict quality control. Specifically, a corpus ten times of earlier works was used in our experiments. CrowdFlower Markup Language (CML) and JavaScript were leveraged to implement the interface. We implemented a 4-component quality control strategy to improve the crowd-generated annotation.

Improving the quality of judgments is one of the most important issues in crowdsourcing, especially for the tasks without strong quality control. A variety of methods have been proposed to assess the quality of judgments from turkers. Kumar and Lease [23,24] presented a weighted voting method based on turkers' accuracies, which can be estimated by taking the full set of labels into account. Jung and Lease [25] conducted a large-scale consensus study on relevant judgements between query/document pairs for Web search on the ClueWeb09 dataset [26]. In their work, approximately 20,000 labels were collected from 766 Mechanical Turk workers. They reported that computing the Z-score could filter noisy labels and achieve a significant improvement, in comparison to a majority vote baseline. Based on the previous work, a semi-supervised approach was proposed to maximize the benefit from consensus [27] with consensus labels from both labelled and unlabeled examples. As these studies indicated, though much progress has been made, quality control and aggregating judgments are still the major challenges of crowdsourcing. The highest reported performance of medication name entity annotation from earlier crowdsourcing attempts in the biomedical NLP domain was

0.68 (F-measure) for agreement between traditional and crowd-generated corpora [21,22].

In our research, we applied strict quality control to select qualified turkers and investigated multiple approaches to aggregating judgments. The goals of our study were to build upon previous work from the general crowdsourcing research and to evaluate the usability of crowdsourcing approach in the clinical NLP domain. This will help us automate clinical trial eligibility screening. The clinical NLP tasks that we used for the purpose of evaluation were medical named entity recognition and entity linking in a clinical trial announcement (CTA) corpus. The entities involved were medication names and medication types, as well as their attributes. During our research, we first studied the turkers' performance to annotate medical named entities on a large-scale data set. Second, we proposed to use crowdsourcing to link named entities and their attributes, in which the entities and attributes were pre-annotated in the text and the crowdsourcing task was to identify entity/attribute pairs that are associated in the text. Third, we attempted to find a new solution to produce a more robust, manually-created gold standard (ie, correction) by investigating whether an iterative model of crowdsourcing tasks can correct errors from previous generations of tasks. Finally, we studied 3 methods to aggregate multiple annotations of the same text to generate a better gold standard.

Our research contributed to the field of clinical NLP by: (1) evaluating the usability of crowdsourcing in the clinical NLP domain, (2) publicly releasing the user interface software that is necessary for crowdsourced, named-entity annotation, and (3) implementing a 4-component quality control strategy to improve the crowd generated annotation, including an introductory quiz to filter the automated scripts, a geographical constraints for turkers, training turkers for the task, and continuous performance monitoring. We will release the annotated corpora in December 2013 when our NIH grant funding concludes.

## Methods

### Definition of Annotated Named Entities and Linkages

This section presents the definitions and examples of medication entities (medication names and medication types) and medication-attribute linkages annotated in this work.

#### *Medication Name*

Medication names are specific names of drugs, biological substances, and treatments. Some examples of mediation names are ibuprophen, phosphonoacetic acid, vancomycin, and ganciclovir.

#### *Medication Type*

Medication types refer to classes of drugs (eg, antibiotics, anti-inflammatory drugs, benzodiazapines), types of drug therapy (eg, chemotherapy), and general references to medications (eg, "study drug", "other drugs", "medication").

#### *Attribute*

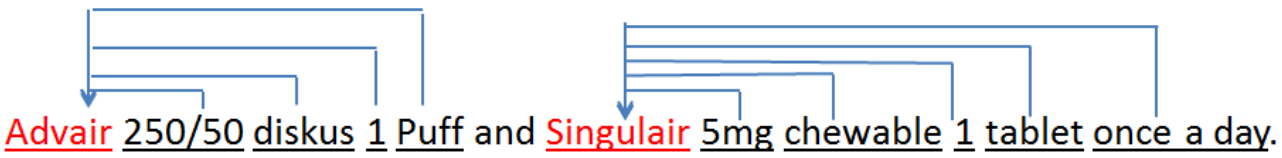
Attributes define how much, how often, and in what form medications or medication types are taken. We distinguished between the following categories of attributes (based on the schema of the SHARPn project [28]):

- Date: indicating all dates associated with the medication (eg, start dates, concluding dates)
- Strength: indicating the strength number and units of the prescribed drug
- Dosage: indicating the amount of each medication used by the patient and type of dose it is (eg, high dose, low dose, stable dose)
- Form: indicating the shape or configuration of the medication (eg, tablet, capsule, liquid, injection, infusion)
- Frequency: indicating how often each dose of the medication should be taken
- Duration: indicating how long the patient is expected to take the drug
- Route: indicating route or method of the medication (eg, intravenous, oral, chew, topical)
- Status change: indicating whether the medication is currently being taken or not (eg, active, inactive, hold, incomplete, started, discontinued, increased, decreased, no change)
- Modifier: indicating mentions that could exist under certain circumstances (eg, conditional modifier), develop or alter a mention (eg, course modifier), or generic modifier (eg, conventional)

### Linking

The linking task associates attributes to their corresponding medication entities, assuming medications and attributes have already been pre-annotated. The following sentence demonstrates the linking task: "Advair 250/50 diskus 1 puff and Singulair 5mg chewable 1 tablet once a day". In this sentence, "Advair and Singulair" are the medication names and "250/50, diskus, 1, puff, 5mg, chewable, 1, tablet, and once a day" are the attributes. In this example, "250/50, diskus, 1, and puff" are the attributes of Advair, "5mg, chewable, 1, tablet and once a day" are the attributes of Singulair, as shown in Figure 1.

**Figure 1.** Example of linkages between medications and their attributes.



## Gold Standard to Evaluate Turker Performance

In one of our previous projects, CTA were annotated for medication extraction. In this paper, we present the most important features of the gold standard used in the study. Details of the corpora and the process of the gold standard development were thoroughly described in a separate manuscript that was published in the 2012 AMIA Annual Conference Proceedings [29]. The corpus was double annotated for medication names, types, and attributes by two annotators (college graduates with bachelor degrees) to create a gold standard, at a cost of 20 days per annotator for annotation of medication names and types and an additional 20 days per annotator for the attributes. Additionally, each attribute was linked to its respective medication name or medication type.

The CTA corpus was composed of 3000 CTAs randomly selected from the ClinicalTrials website (105,598 documents as of March 2011). We annotated only the eligibility criteria sections of the trial announcements. Table 1 shows the descriptive statistics of the corpus (number of documents and number of annotations in the traditional gold standard). In this study we used crowdsourcing to annotate only medication names and medication types. The linking crowdsourcing experiment utilized pre-annotated text: medication names, medication types, and attributes.

**Table 1.** CTA corpus statistics.

| Corpus statistics | |
|---|---|
| Documents | 3000 |
| Tokens | 635,003 |
| **Annotations** | |
| Medication name | 9968 |
| Medication type | 11,789 |
| Date | 16 |
| Dosage | 645 |
| Duration | 644 |
| Form | 482 |
| Frequency | 381 |
| Route | 894 |
| Status change | 598 |
| Strength | 409 |
| Modifier | 5827 |
| All classes | 31,653 |

Because the CTAs were longer than the text of many crowdsourcing tasks, and considering the difficulty of clinical NLP annotations, we decided to break up the CTAs into smaller paragraph-length sizes for the tasks. Based on a tokenizer we wrote to count discrete basic units, the average token count in a CTA document was 212. In the paragraph-size tasks, we split the CTAs into paragraphs with at least 50 tokens, preserving the original format and the integrity of the CTA file (no paragraphs spanned into different CTAs). This resulted in 9773 paragraphs or “units”.

## Crowdsourcing User Interface

We selected CrowdFlower (CF) as the platform through which we would access AMT because CF’s self-service product met our needs for flexibility in graphical user interface (GUI) modification and offered means for strict and continuous quality control for the annotations. We wrote a custom JavaScript program that was loaded into a CF job, allowing the turkers to highlight and classify entities in a similar fashion to the Knowtator plug-in for Protégé [30] that was used in our traditional annotation methods.

In addition to the customizable GUI, another key benefit of using the CF crowdsourcing platform over directly accessing AMT is that it has strict quality control measures. CF provides an interface for creating and editing “gold standard answers” for quality control. “Gold standard answers” are randomly included (without the turkers being aware of their presence) in the submitted data and a turker is required to meet a minimum threshold of accuracy in these “gold” examples in order to continue submitting tasks. When a turker meets this threshold, he/she is deemed “trusted”. Only the “trusted” turkers’ data are collected to establish final judgments. If an example has 3 medications in the unit and the turker annotates only two correctly, the system will score the judgment as incorrect as there are no partial scores in determining a turker’s trust status within a particular unit of annotated text.

In pilots, we experimented with different thresholds. Lower thresholds resulted in lower agreement of the turker-annotated corpora with the gold standard corpora. Higher thresholds prevented the successful completion of the task by eliminating too many turkers. Because of the complexity of the task, we experimented with a trust-based threshold and found 50% (on unit level) to be the most feasible threshold number. A turker

presented with "gold standard" examples must accurately annotate 50% of the unit-based responses. That is, if the turker annotated 4 units of "gold" examples, at least 2 (of the 4) had to be exact matches for him/her to establish trustworthiness. The 50% threshold was evaluated on the unit's level and not on the named entity level. That is, all of a unit's annotations, or judgments, had to be matched exactly with the "gold standard" answers, irrespective the number of named entities per unit.

We also found that the training mode of CF was very helpful in winnowing the pool of turker candidates to only the highest quality annotators. In training mode, the turkers were directed to several training examples first. All of the training examples were gold standard examples and the turker must complete 4 examples correctly to proceed to the production annotation task. Based on these interfaces, we implemented our quality control strategy. Of all our tasks, 20% of the total number of units submitted for judgment were uploaded and setup as "gold" units. That is, 20% of the annotated units were gold standard units where the CF system could continuously gauge the trustworthiness of the turker. If a turker's trustworthiness slipped below 50%, the turker was warned. If his/her performance did not improve during the next two gold tests, then the turker's entire output was excluded from the collected data and the system subsequently blocked the turker from submitting any further judgements.

In the in-house experts' generated gold standard, approximately 30% of the CTAs had no medications or medication types. Due to the splitting of the CTAs into smaller units, however, the empty percentage grew to 42%. Several initial pilot experiments were conducted regarding the study's design features, including training mode, trusted-turker accuracy threshold, and whether empty tasks were included or not. We tested the performance of excluding empty units (where the data included at least one entity from the in-house gold standard in every unit and a turker had to mark at least one entity to submit) and including empty units (ie, units that have no entities from the in-house gold standard). To mirror the original task given to the traditional annotators and to keep the annotated sample representative of the full CTA corpus, we kept the empty units at 30% of the crowdsourced units.

During the pilot annotations, we had difficulty with a large numbers of untrusted turkers and judgments coming from Asia so we restricted the project to turkers from Australia, Canada, the United Kingdom, and the United States. We also requested 5 judgments per unit (from 5 different turkers) in order to allow flexibility with voting measures and methods. In addition to the training mode, a qualification quiz was presented to each turker the first time they signed up for our tasks. They had to read and understand the instructions, and answer a short quiz (3 multiple choice questions) in order to gain access to the job. The quiz blocked "robot scripts" from participating in our tasks.

Figure 2 shows snippets of the corresponding CML code, common style sheet (CSS) code, and custom JavaScript. As seen in Figure 2, CF provides the interface for users to edit CML, CSS, and JavaScript in corresponding text areas named CML, CSS, and JavaScript, respectively. The complete GUI code can be found in the Multimedia Appendix 1. This interface was primarily used for creating our annotation program. The main restriction for the custom JavaScript code is that it runs only once, when CF randomly selects a unit and presents it to a turker to perform a judgment. In order to deal with this restriction, we created the program based on event-driven programming, in which each user's annotation operation was captured and processed by a designated function. This dynamic JavaScript program worked for each unit submitted for judgment. For the medication and medication type named-entity annotation tasks, our program displayed the unit, allowing for an offset (a single word or group of words) to be selected with a left-click and drag of the mouse or a double-click and a subsequent right click event, in which the turker would select the class associated with the highlighted text. The program kept a sorted record of the offsets, classified for both entity classes, and submitted these offsets as named entities when the turker clicked "submit". The program also handled discontinuous entities, as described below. The performance evaluation described in the experiments section involves comparing the offsets submitted between the turker's judgments and the gold standard. Figure 3 shows the interface of the medication and medication type named-entity annotation task. In this interface, 4 of the major functions were provided, which were "word selection", "annotation selection", "annotation information display", and "discontinuous highlighting".

The "word selection" function supported double-clicking to select a single word, automatic word-extending and invalid character-shrinking to improve the accuracy and efficiency of the turkers' operations. Two buttons ("extend highlight" and "shrink highlight") were provided to extend and shrink the highlighted (annotated) area on the right hand side by one character at a time. After selecting one word or more, a menu with two options ("Medication Name" and "Medication Type") popped up for the turkers to select the target annotation type. After the turkers clicked the option, the selected word(s) was highlighted by a corresponding color (eg, green was for "Medication Type," as shown in Figure 3) and all the current annotation information was displayed in the table named "Annotated Entity List". If the turkers needed to remove annotations already highlighted with a label or if they wanted to change the label of the highlighted word, they had to left-click on the highlighted word and click "OK" to confirm their choice to remove the annotation from the table at the bottom of the page. It should be noted that entities comprised of discontinuous tokens could be highlighted as a single entity by concurrently pressing down CTRL.

The interface for the correction task (correcting previously annotated entities) was similar to the annotation task with the only difference being that some words were pre-annotated (highlighted). The offsets associated with these highlighted words were prefilled into the unit judgment table.

Figure 4 shows the interface for the linking task. In this interface, all medications (medication names and medication types) and attributes were highlighted with their respective colors. Medications were highlighted with yellow and attributes were highlighted with light gray. Turkers had to left click on the medications and attributes in corresponding pairs. The

selected medications and attributes were displayed in the corresponding textbox to link them together and the linked pairs were shown in the table named “Linking Information List”. If the turkers wanted to remove an entity-attribute pair, they could click the "Remove" button to the left of the pair in the “Linking Information List.” In all of these tasks, the results were internally represented by offsets instead of the original text, which was necessary to address the problem of words occurring more than once in the same annotation unit. The GUI worked in Firefox and Chrome browsers. The JavaScript checked the browser type when a turker signed up for our tasks and, if the turker did not use one of the two browsers, it would instruct the turker to download and install a correct browser.

**Figure 2.** Snippets of CML, CSS and JavaScript.

CML Show title & instructions

```
<p>{{sentences}}</p>
<cml:text label="medication or type" class=""></cml:text>
  <cml:text label="attribute" class=""></cml:text>
  <cml:hidden label="linkinginfo" gold="true" class=""></cml:hidden>
  <cml:text label="requirearea" class="" validates="required" default="---"></cml:text>
```

CSS

```
.highlightmedicationname {
background-color: yellow;
 }

 .highlightmedicationtype {
background-color: limegreen;
 }
.cmenuAll
{
 width:165px;
  height:65px;
  position:absolute;
```

JavaScript *(runs once on page load)*

```
* author: Haijun Zhai
* date: 5/15/2012
* verson: 1.0
*/
function HightlightRange(selectRange, varID, varStyle)
{
  var varStartOffset = selectRange.startOffset;
  var varEndOffset = selectRange.endOffset;
  var varParentNodeTemp = selectRange.startContainer.parentNode;
  //debug to check the exact mean of start/end offset of Range, [checked that, the endoffset is
  the start of the continous next string]
  if(selectRange.startContainer == selectRange.endContainer){
    var varBeginStr = selectRange.startContainer.textContent.substring(0, varStartOffset);
    var varEndStr = selectRange.startContainer.textContent.substring(varEndOffset);
    var varSelectStr = selectRange.startContainer.textContent.substring(varStartOffset,
```

**Figure 3.** Medication named entity recognition task interface.

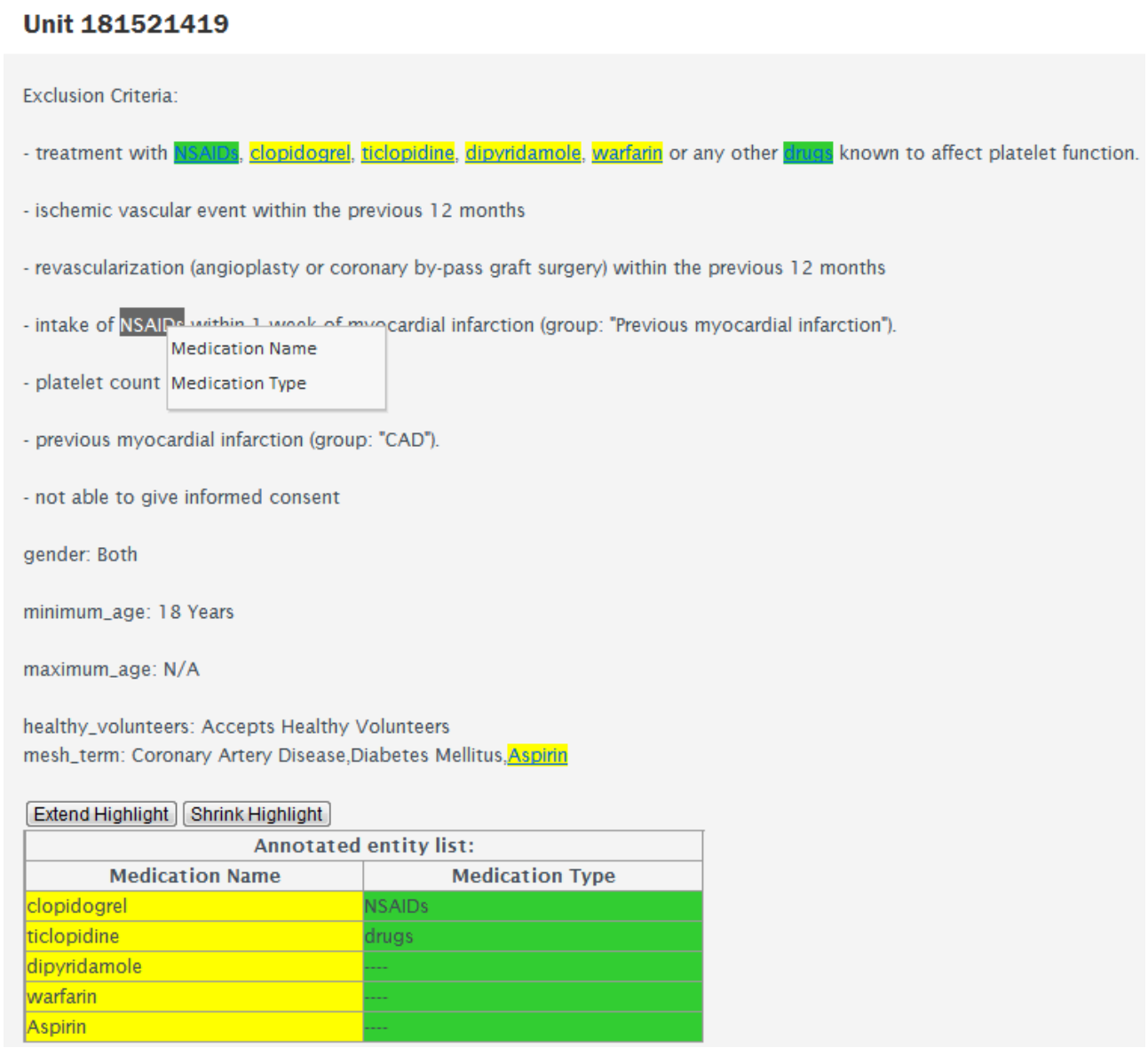
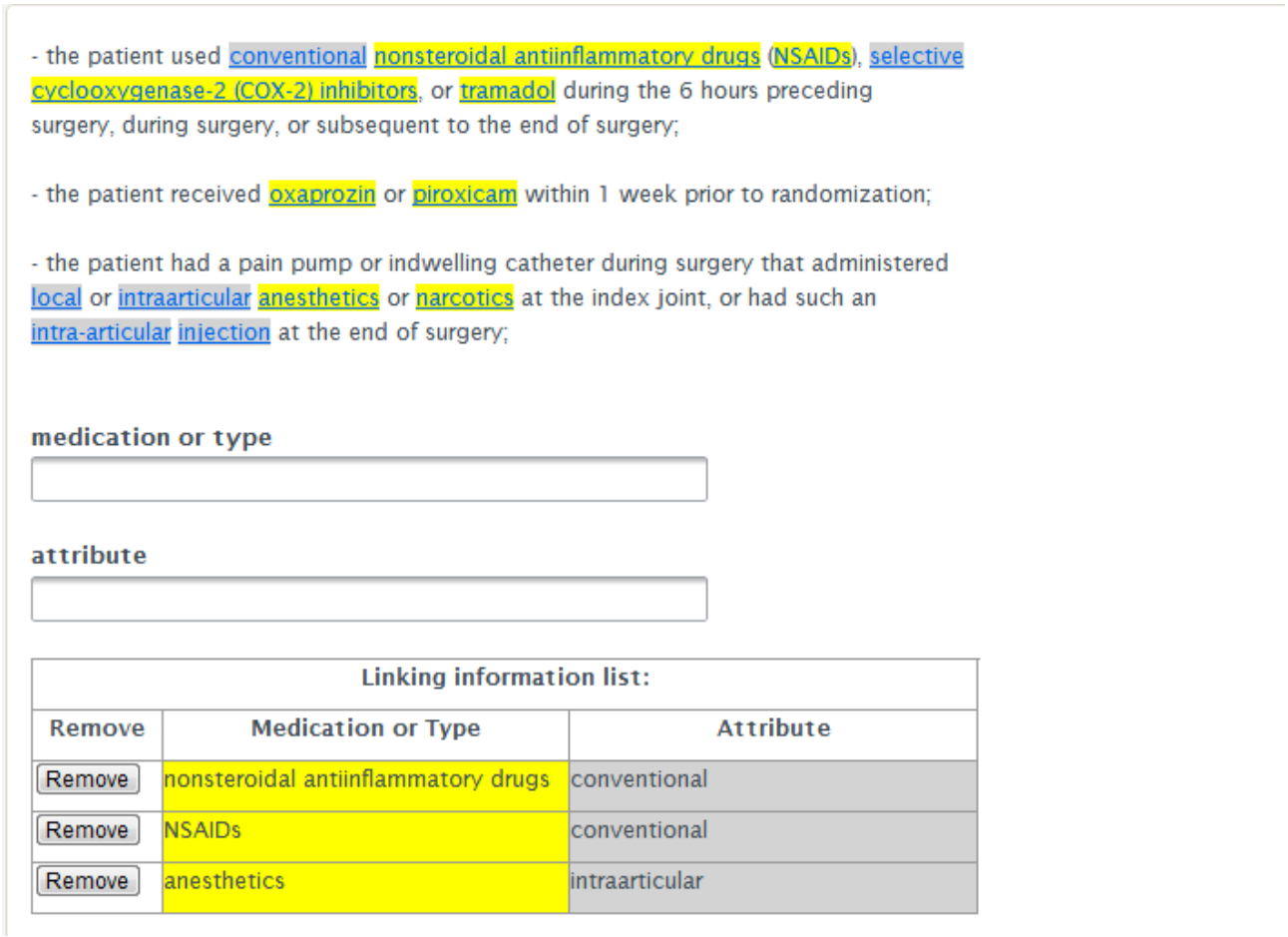


**Figure 4.** Linking task interface.




## Experiments

After the initial, smaller pilot experiments, we selected a larger number of units for the complete named-entity recognition task. In an earlier unpublished project to develop a machine learning-based medication entity-extraction pipeline, we determined that 1042 CTAs were necessary for the training set to achieve higher than 0.80 F-measures (0.86 for medication name and 0.82 for medication type, using Conditional Random Fields algorithm for information extraction). We used this empirically determined corpus size of 1042 CTAs, corresponding to 3400 units as mentioned in previous section for both the medical name-entity recognition and entity-linking jobs. Several samples annotated by turkers and their corresponding gold standard are presented in the Multimedia Appendix 2.

Based on the pilot medication named-entity annotation experiments, the correction experiment was performed by taking a smaller data set with 200 units and its corresponding 1000 judgments (5 judgments for each unit) and submitting the unique judgments to another crowdsourcing job. The previous experiment had 735 unique judgments (out of 1000). If a particular unit had 3 unique judgments and two additional duplicate judgments, we resubmitted only the 3 unique judgments for the correction job. A judgment was defined by the response of a turker to a unit, covering all of the entities annotated for that unit. In this example, the original job had 5 judgments for the unit and the correction job had 15 judgments for that same unit (3 unique judgments submitted for 5 correction judgments each). For each correction judgment, a turker had the opportunity to remove annotations, add additional annotations, or provide no change to that unit.

## Measurements

In this paper, standard named-entity recognition and classification measurements were adopted to evaluate the performance of the experiments, including Precision (P), Recall (R), and F-measure (F), which are defined in the Multimedia Appendix 3.

## Voting Methods

One of the aims of this study was to evaluate different methods of voting on judgments from crowdsourced outputs. Because these are named-entity and linking tasks, the calculation is on the entity and linking level. We experimented with 3 voting methods for the medication name and medication-type entity recognition and the medication attribute linking tasks.

We investigated 3 voting methods: simple voting (simple), trusted score weighted voting (trust), and turker experience weighted voting (experience). All voting was performed at the entity and linking level (micro average), regardless of the number of entities and linkages in a given task unit. Equations (1), (2), and (3) shown in Multimedia Appendix 4 describe the formulas we used for the 3 voting methods. Let $e$ be the number of votes for a particular named entity and let $J$ be the number of judgments (number of turkers who submitted responses) in

this unit. Let $t_i$ be the trust score of turker $i$ who annotated the entity. Let $u_i$ be the total number of judgments user $i$ performed and let $m$ be the maximum number of judgments the most prolific turker performed. For simple voting presented in Equation 1, if there were 2 or more annotations (out of 5 judgments/responses) for a particular entity, it was selected for the adjudicated judgment.

Equation 2 gives the trusted score voting, which weighs a particular turker's entity vote with their trust score (a trust score of 75% provided a 0.75 vote per entity and the max trust score of 100% provided a single simple vote). As presented in Equation 3, turker experience voting weighted each entity vote by the experience score of the turker. The experience score is the number of judgments performed by a turker relative to the maximum number of judgments the most prolific turker performed in that experiment. For example, in one job, a turker submitted 163 judgments, which was the most of any turker in that job. That turker's weight for all of his entity votes became 1 and the experience score for all other turkers became u/163. Note that the intention of division in 3 equations was to normalize the scores to the range of 0 to 1. As presented in Figure 5, there was high variance in the accomplished number of jobs between turkers. The point of the logarithm in Equation 3 was to scale the difference.

The F-measures were calculated using the 3 voting methods on the original judgments (with each unit having 5 judgments) as correction baselines presented in Table 5. These were then compared to the subsequent correction results computed by 3 voting methods of all correction judgments presented in Table 6. In order to further show the impact of correction, another measure, which we described as a response-level entity vote, is presented in Figure 6. We counted whether the F-measure of the correction judgments improved upon the F-measure of the original judgment.

**Figure 5.** The distributions of turkers' experience for medical named-entity task, correction task and linking task (X axis denotes number of jobs, Y axis indicates number of turkers).

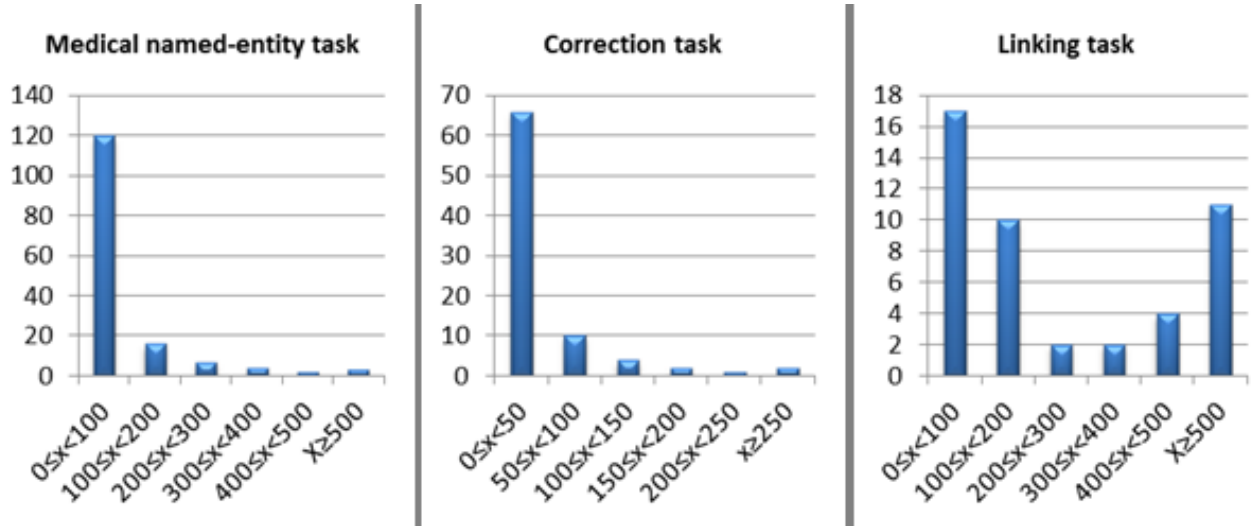
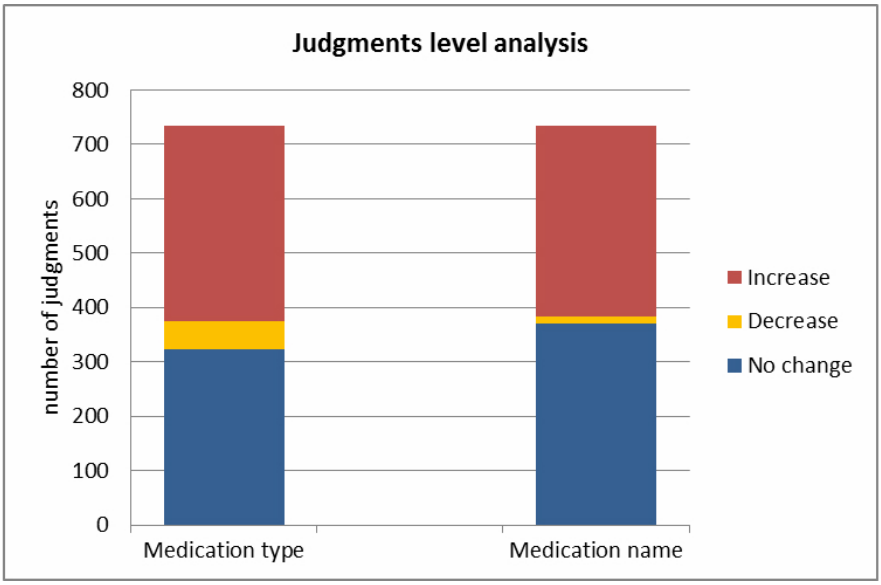



**Figure 6.** Improvement chart for correction task.



### Statistical Significance Test of Turker Performance

In order to analyze the differences between the corpus created by the turkers and the corpus created by in-house expert annotators, a statistical assessment method (named "pooling chi-square test") was proposed to calculate the *P* values. In this method, the voting results from turkers were pooled together with the corpus that was annotated by experts. These pooled results were then tested against the original voting results. The hypothesis was that the turkers with sufficient training and aggregating multiple results could perform as well as experts. If this hypothesis was true, then pooling the results was not expected to change the original CF voting results. Specifically, the hypothesis $H_0$ was that the experts' output did not change the quality of the turkers' annotations (reflected by the number of unique entities annotated correctly and incorrectly). If the *P* value was less than the designated threshold (0.05), it meant that the experts' output significantly affected the quality of the turkers' results. In other words, the turkers did not perform as well as experts. If the *P* value was higher than the predetermined threshold, then we could not reject the hypothesis. Therefore, we could infer that there was no evidence for statistically significant differences between the turkers' and experts' annotations.

## Results

### Information on Turkers

Table 2 shows information on turkers participating in our 3 tasks. We had 156 turkers, 86 turkers, and 46 turkers to complete medical named-entities task, correction task and linking task, respectively. Figure 5 presents the distribution of turkers by the number of performed jobs for the 3 tasks. We found that the top 5 most prolific turkers completed 39.9% (6778/17,000) medical named-entities jobs, 44.0% (1616/3675) correction jobs, and 45.4% (7716/17,000) linking jobs. Figure 7 shows the distribution of F-measure of turkers for the 3 tasks. We can see that F-measures of greater than 0.6 were achieved by over 83% turkers for the medical named-entities task, over 88% of turkers for the correction task and 100% for the linking task. Table 3 presents the cost and completion time of the 3 tasks. The payment of 3.84 cents per judgment included 3 cents paying for turkers and 0.84 cents charged by CF. Table 3 also presents the time required for the in-house annotators to complete the same tasks. Additionally, the time to receive results from in-house annotation is around 5 times longer than crowdsourcing due to the parallel nature of the crowdsourcing task and the traditional work hours (eg, Monday to Friday, 9am-5pm). The 133 hours represented by the total in-house annotation were the total work hours. The total elapsed time was 10 days (8 work days plus 2 weekend days).

### Results of Medical Named-Entities Annotation Task

Table 4 shows the results of the turkers' medical named-entity annotation with the 3 voting methods that were implemented. It shows the turkers' generated corpus' agreement with the in-house experts' generated gold standard at various threshold levels.

**Table 2.** Information on turkers participating in the 3 tasks.

| Task name | Participating turkers | Turkers passing the test |
|---|---|---|
| Medical named-entities | 1144 | 156 |
| Correction | 678 | 86 |
| Linking | 644 | 46 |

**Table 3.** Cost and time of the 3 tasks.

| | | Crowdsourcing | | | In-house | |
|---|---|---|---|---|---|---|
| Task name | Total units | Total judgments | Total cost (per judgment) | Total time (per judgment) | Total judgments | Total time (per judgment) |
| Medical named-entities | 3400 | 17,000 | $652.85 (3.84 cents) | 57 hours (12.07 seconds) | 6800 | 128 hours (67.76seconds) |
| Correction | 735 | 3675 | $141.13 (3.84 cents) | 38 hours (37.22 seconds) | N/A | N/A |
| Linking | 3400 | 17,000 | $652.85 (3.84 cents) | 27 hours (5.72 seconds) | 6800 | 44 hours (23.29 seconds) |

**Figure 7.** The distribution of turkers' F-measure for medical named-entity task, correction task and linking task (X axis denotes F-measure, Y axis indicates number of turkers).

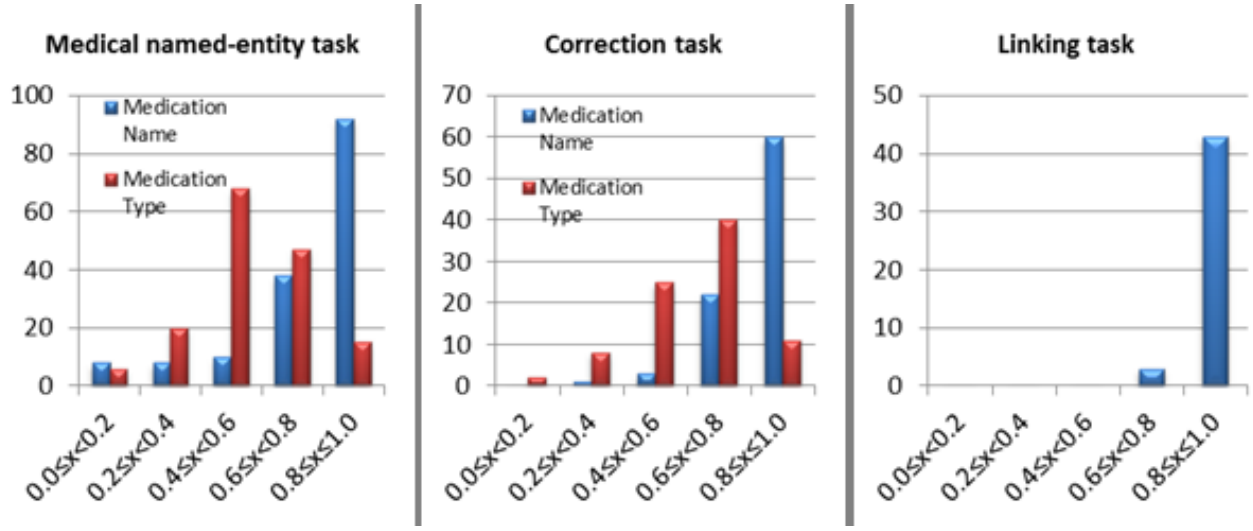



**Table 4.** Results of medical named entity annotation (the pre-determined threshold and its corresponding P[a], R[b], and F[c] for each column are italicized).

| Simple | | | | Trust | | | | Experience | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Th[d] | P | R | F | Th | P | R | F | Th | P | R | F |
| **Medication name** | | | | | | | | | | | |
| 0.20 | 0.694 | 0.931 | 0.796 | 0.18 | 0.835 | 0.887 | 0.860 | 0.18 | 0.807 | 0.895 | 0.849 |
| *0.40* | *0.864* | *0.879* | *0.871* | *0.24* | *0.864* | *0.879* | *0.871* | *0.24* | *0.869* | *0.874* | *0.872* |
| 0.60 | 0.920 | 0.815 | 0.865 | 0.30 | 0.869 | 0.874 | 0.871 | 0.30 | 0.885 | 0.854 | 0.870 |
| 0.80 | 0.955 | 0.696 | 0.805 | 0.36 | 0.916 | 0.819 | 0.865 | 0.36 | 0.910 | 0.820 | 0.863 |
| **Medication type** | | | | | | | | | | | |
| 0.20 | 0.431 | 0.879 | 0.579 | 0.18 | 0.632 | 0.781 | 0.699 | 0.18 | 0.583 | 0.800 | 0.675 |
| *0.40* | *0.698* | *0.763* | *0.729* | *0.24* | *0.698* | *0.763* | *0.729* | *0.24* | *0.711* | *0.751* | *0.731* |
| 0.60 | 0.831 | 0.598 | 0.696 | 0.30 | 0.709 | 0.745 | 0.727 | 0.30 | 0.756 | 0.703 | 0.729 |
| 0.80 | 0.911 | 0.396 | 0.552 | 0.36 | 0.819 | 0.608 | 0.698 | 0.36 | 0.816 | 0.614 | 0.700 |

[a]precision
[b]recall
[c]F-measure
[d]threshold

## Results of Correction Task

The results of the correction task, its corresponding correction baseline, and the results of combined judgments are presented in Tables 5 and 6, respectively. In the correction task, the turkers and experts agreement F-measure of medication name and medication type achieved 0.900 and 0.760 by simple vote, respectively. With comparison to the F-measure of its corresponding correction baseline, relative improvements of 2.62% (medication name Baseline F-measure = 0.877, After_Correction_F-measure = 0.900; computed by (After_Correction_F-measure - Baseline_F-measure)/ Baseline_F-measure * 100) and 10.79% (n/N; medication type name Baseline_F-measure = 0.686, After_Correction_F-measure = 0.760; computed by (After_Correction_F-measure - Baseline_F-measure)/ Baseline_F-measure * 100) were gained (Tables 4 and 5).

**Table 5.** Results of correction task with 200 units and 1000 judgments (the pre-determined threshold and its corresponding P[a], R[b], and F[c] for each column are italicized).

| Simple | | | | Trust | | | | Experience | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Th[d] | P | R | F | Th | P | R | F | Th | P | R | F |
| **Medication name** | | | | | | | | | | | |
| 0.20 | 0.796 | 0.938 | 0.861 | 0.18 | 0.825 | 0.927 | 0.873 | 0.18 | 0.845 | 0.921 | 0.881 |
| *0.40* | *0.896* | *0.904* | *0.900* | *0.24* | *0.861* | *0.916* | *0.888* | *0.24* | *0.876* | *0.908* | *0.892* |
| 0.60 | 0.950 | 0.812 | 0.875 | 0.30 | 0.898 | 0.906 | 0.902 | 0.30 | 0.900 | 0.891 | 0.896 |
| 0.80 | 0.972 | 0.732 | 0.835 | 0.36 | 0.908 | 0.887 | 0.897 | 0.36 | 0.933 | 0.867 | 0.899 |
| **Medication type** | | | | | | | | | | | |
| 0.20 | 0.610 | 0.916 | 0.733 | 0.18 | 0.655 | 0.892 | 0.755 | 0.18 | 0.662 | 0.872 | 0.752 |
| *0.40* | *0.732* | *0.790* | *0.760* | *0.24* | *0.691* | *0.843* | *0.759* | *0.24* | *0.703* | *0.817* | *0.756* |
| 0.60 | 0.851 | 0.541 | 0.661 | 0.30 | 0.736 | 0.792 | 0.763 | 0.30 | 0.756 | 0.773 | 0.764 |
| 0.80 | 0.945 | 0.382 | 0.544 | 0.36 | 0.776 | 0.744 | 0.760 | 0.36 | 0.783 | 0.684 | 0.730 |

[a]precision
[b]recall
[c]F-measure
[d]threshold

**Table 6.** Baseline Results of medical named entity annotation corresponding to the correction task (the pre-determined threshold and its corresponding P[a], R[b], and F[c] for each column are italicized).

| Simple | | | | Trust | | | | Experience | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Th[d] | P | R | F | Th | P | R | F | Th | P | R | F |
| **Medication name** | | | | | | | | | | | |
| 0.20 | 0.712 | 0.934 | 0.808 | 0.18 | 0.839 | 0.891 | 0.864 | 0.18 | 0.774 | 0.908 | 0.835 |
| *0.40* | *0.868* | *0.887* | *0.877* | *0.24* | *0.868* | *0.887* | *0.877* | *0.24* | *0.872* | *0.887* | *0.879* |
| 0.60 | 0.909 | 0.788 | 0.844 | 0.30 | 0.870 | 0.887 | 0.878 | 0.30 | 0.881 | 0.876 | 0.879 |
| 0.80 | 0.956 | 0.655 | 0.778 | 0.36 | 0.899 | 0.803 | 0.848 | 0.36 | 0.894 | 0.809 | 0.849 |
| **Medication type** | | | | | | | | | | | |
| 0.20 | 0.473 | 0.879 | 0.615 | 0.18 | 0.627 | 0.724 | 0.672 | 0.18 | 0.594 | 0.779 | 0.674 |
| *0.40* | *0.669* | *0.704* | *0.686* | *0.24* | *0.669* | *0.704* | *0.686* | *0.24* | *0.668* | *0.698* | *0.683* |
| 0.60 | 0.737 | 0.519 | 0.609 | 0.30 | 0.670 | 0.700 | 0.685 | 0.30 | 0.681 | 0.664 | 0.673 |
| 0.80 | 0.890 | 0.358 | 0.510 | 0.36 | 0.726 | 0.550 | 0.626 | 0.36 | 0.717 | 0.558 | 0.628 |

[a]precision
[b]recall
[c]F-measure
[d]threshold

Furthermore, we analyzed the practical significance of these improvements by calculating the F-measure of medication name and medication type for each unique judgment (the total number of unique judgments was 735) and its corresponding 5 correction judgments. Based on empirical evidence acquired in previous experiments, the F-measure was computed based on a simple vote with the threshold of 0.4. The results are shown in Figure 6. Improvement was seen for 50.5% (370/735) and 44.1% (324/735) of the judgments for medication name and medication type after the turkers' correction, respectively. In contrast, 1.9% (14/735) and 6.9% (51/735) judgments became worse.

## Result of Linking Task

Table 7 shows the results of the linking experiment. Non-expert annotators (turkers) did an excellent job, in which the F-measure achieved 0.962. Meanwhile, as previous results indicated, the simple method could obtain very good results in case of strict quality control.

**Table 7.** Results of linking task (the pre-determined threshold and its corresponding P[a], R[b], and F[c] for each column italicized).

| Simple | | | | Trust | | | | Experience | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Th[d] | P | R | F | Th | P | R | F | Th | P | R | F |
| 0.20 | 0.845 | 0.984 | 0.910 | 0.18 | 0.927 | 0.982 | 0.954 | 0.18 | 0.927 | 0.979 | 0.952 |
| *0.40* | *0.949* | *0.975* | *0.962* | *0.24* | *0.949* | *0.975* | *0.962* | *0.24* | *0.950* | *0.974* | *0.962* |
| 0.60 | 0.981 | 0.959 | 0.970 | 0.30 | 0.949 | 0.975 | 0.962 | 0.30 | 0.955 | 0.973 | 0.964 |
| 0.80 | 0.990 | 0.925 | 0.956 | 0.36 | 0.975 | 0.967 | 0.971 | 0.36 | 0.977 | 0.965 | 0.971 |

[a]precision

[b]recall

[c]F-measure

[d]threshold

### Results of Statistical Significance Analysis

For all the results above, Chi-square statistical significance tests were conducted between the corpora created by Crowdflower's and the gold-standard generated by the in-house annotators. The *P* values (at $P<.001$) showed no statistically significant difference between the best CrowdFlower generated corpora and corresponding in-house generated gold-standard sets.

## Discussion

### Overview

To our knowledge, the medical named-entity annotation task described in this work is the largest scale crowdsourcing experiment in the clinical NLP research field. The results demonstrated that crowdsourcing is a feasible solution for creating a gold standard for medical named-entities. Many works were described in the introduction section, but only one performed a similar medical named entity crowdsourced annotation and is directly comparable to our current study. All other works focused on different corpora and entity types and cannot be compared directly with these works. We improved upon the previously reported results on medical named entity annotation task [21,22] with more than 27.9% of the F-measure (F-measure_Current_Study = 0.87 vs F-measure_Earlier_Work = 0.68 for agreement between the crowdsourced and traditionally developed corpora; computed by (F-measure_Current_Study vs F-measure_Earlier_Work)/ F-measure_Earlier_Work * 100) for named-entity annotation. This experiment also showed that the crowdsourcing performance for medication name annotations is much better than those of medication type. This is a similar finding to the in-house results with trained, expert annotators. We attribute this phenomenon to the clarity of the task for medication name annotation. In other words, the definition and the gold standard answers of medication names are easier to understand and to capture than those of medication type. In the future, we plan to use a more easily interpretable definition of medication types to improve performance. We also plan to use crowdsourcing to annotate attributes, such as date, dosage, as listed in Table 1.

Based on our experiments, we found that it was easy to find a large number of turkers by crowdsourcing. Around 10% (156/1144, 86/678, 46/644 for medication name entity, linking and correction tasks, respectively) of the turkers passed our quality control test (see Table 2). Among those turkers, around 10% (14/156, 11/86, 7/46) of them contributed over 50% (10,521/17,000, 10,900/17,000, 1907/3675) of the jobs.

As shown in Table 4, the non-expert annotators performed at a very high quality and the results indicated that the simple method could obtain very good results, provided the quality control is strict. In our previous work [29], we reported inter-annotator agreement (IAA) F-measures for medication names and medication types, 94.2% and 88.2% respectively. Additionally, what could have conceivably been weeks' worth of in-house annotation work was achieved in less than a day of crowdsourcing effort.

Our previous study conducted experiments by implementing a rule-based linking system [31]. The result (around 0.72 F-measure) showed that manual annotation is definitely needed to develop an effective training set for a machine learning-based linking system. The presented linking experiment is the first work known to us that attempted to link medications to their corresponding attributes with crowdsourcing. The results indicated that linking is not a difficult task and the data created can be sufficiently applied to real applications. Based upon this experiment, we plan to create a larger scale data set using crowdsourcing and to apply it to clinical NLP tasks. We will further evaluate the performance of linking by implementing our linking strategy for other clinical named entities. The results of the linking task are excellent, with a near 100% (N=3400) agreement between crowd and traditionally developed corpora.

As shown in Table 3, the linking task took much less time than the other two tasks, most likely because the linking task is much easier than the other two tasks. The time per judgment for medical named entity annotation task is much less than that of the correction task (12.07 vs 37.22 seconds respectively). The reason is that the medical named entity annotation task has more participating turkers (156 vs 86). We can conclude that the difficulty of tasks and the number of participating turkers strongly affect the completion time of the tasks. In contrast to traditional annotation, crowdsourcing achieved 55.5% time (71/128 hours) and 75.0% cost ($1958/$2611) savings for medical named entity annotation. For the linking task, 38.6% time (17/44 hours) and 27.2% cost ($244/$897) savings were seen when using crowdsourcing.

To our knowledge, we were the first to conduct clinical NLP correction experiments. The results of that experiment are quite encouraging. Our correction F-measure was 0.90 (medication names) and even the worst final F-measure improved by more than 10% after the corresponding voting (medication types). We believe that this experiment showed another feasible and efficient way to improve the output of crowdsourcing. We designed an efficient strategy to perform correction. Future work will focus on determining the number of iterative cycles to achieve the best results.

As was mentioned in the previous sections, creating a smaller batch of gold standard data (in-house with expert annotators) is a critically important step for crowdsourcing quality control. This in-house gold standard can be used later to: (1) train turkers, (2) perform quality control, and (3) determine thresholds to aggregate judgments. In this study, we also modified the gold standard management interfaces of CF to perform turker training and quality control by setting gold standard answers. There is room for further research in different methods to train turkers and to experiment with quality control thresholds.

Finally, 3 different voting methods were investigated to aggregate judgments. The results showed that it is quite possible to acquire a high-quality annotated corpus by implementing simple voting under the condition of strict quality control. In pilots, we experimented with different voting thresholds. Lower thresholds resulted in lower agreement of the turker-annotated corpora with the gold standard corpora. Higher thresholds prevented the successful completion of the task by eliminating too many turkers. The thresholds used in the paper (eg, 2 judgments out of 5 or 0.4) were set empirically based on our pilot experiments and earlier related work [32,33]. For the judgment-based voting (eg, trust-based and experience-based voting) more complicated voting methods could be implemented and compared.

A potential limitation of this study was that, the proportion of empty units in our experimental corpus was less (30%) than that in the general population of CTA documents (42%). On the other hand, our pilot experiments show that the proportion of empty units did not influence the performance of the turkers. A second potential limitation was that we included only 3 voting methods among the tested voting schemas. We plan to address this limitation in our future works.

## Conclusions

In this study, we evaluated the feasibility of crowdsourcing for creating gold standard data for clinical NLP tasks. Although direct comparison with all related work in the literature was not possible because of corpora and entity type differences, by implementing strict quality control for turker selection and by continuously monitoring the turkers' performance, we improved upon the directly comparable results in the literature with more than 27.9% for the named-entity annotation task. 3 major experiments were conducted: (1) named-entity annotation, (2) entity linking, and (3) annotation correction. In addition, 3 voting methods were studied. To our knowledge we were the first to investigate the feasibility of crowdsourcing for clinical named-entity annotation on a large-scale corpus. Similarly, we are not aware of a competing work in the clinical NLP domain that proposed to use crowdsourcing to create an entity-linking gold standard for information extraction, on our experiments' scale. Furthermore, we proposed a successful correction strategy that applied crowdsourcing to crowdsourcing results to improve the quality of the annotated corpus. We found that a high-quality, clinical NLP gold standard data could be obtained by a simple voting method, if a strict quality control is implemented.

Based upon the results of our experiments, we conclude that crowdsourcing is a feasible, inexpensive, fast, and practical approach to annotate clinical text (when protected health information is not included) on large scale for medical named-entities. We believe that well-designed user interfaces for entity annotation and linking were critical to the success of this work. As a further contribution to the Web 2.0-based crowdsourcing field, we publicly released the JavaScript and CML infrastructure code that is necessary to utilize CrowdFlower's quality control and crowdsourcing interfaces for named entity annotations [34].

## Acknowledgments

The work presented was partially supported by NIH grants 5R00LM010227-04, 1R21HD072883-01, and 1U01HG006828-01.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

GUI source code.

[PDF File (Adobe PDF File), 528KB - jmir_v15i4e73_app1.pdf ]

## Multimedia Appendix 2

Samples annotated by turkers and their corresponding gold standard.

[XLS File (Microsoft Excel File), 6MB - jmir_v15i4e73_app2.xls ]

### Multimedia Appendix 3

Definitions of precision, recall, and F-measure.

[PDF File (Adobe PDF File), 97KB - jmir_v15i4e73_app3.pdf ]

### Multimedia Appendix 4

Voting method equations.

[PDF File (Adobe PDF File), 99KB - jmir_v15i4e73_app4.pdf ]

## References


1. Chapman WW, Nadkarni PM, Hirschman L, D'Avolio LW, Savova GK, Uzuner O. Overcoming barriers to NLP for clinical text: the role of shared tasks and the need for additional creative solutions. J Am Med Inform Assoc 2011;18(5):540-543 [FREE Full text] [doi: 10.1136/amiajnl-2011-000465] [Medline: 21846785]
2. Snow R, O'Connor B, Jurafsky D, Ng AY. Cheap and fast-but is it good?: evaluating non-expert annotations for natural language tasks. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics; 2008 Presented at: Empirical Methods in Natural Language Processing; 2008; Honolulu, Hawaii p. 254-263.
3. Amazon Mechanical Turk. Seattle, WA: Amazon URL: https://www.mturk.com:443/mturk/welcome [accessed 2012-10-10] [WebCite Cache ID 6BJam8G4X]
4. Lawson N, Eustice K, Perkowitz M, Yildiz M. Annotating large email datasets for named entity recognition with Mechanical Turk. Stroudsburg, PA: Association for Computational Linguistics; 2010 Presented at: NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk; 2010; Los Angeles, CA p. 71-79.
5. Finin T, Murnane W, Karandikar A, Keller N, Martineau J. Annotating Named Entities in Twitter Data with Crowdsourcing. Stroudsburg, PA: Association for Computational Linguistics; 2010 Presented at: NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk; 2010; Los Angeles, CA p. 80-88.
6. Crowdflower. URL: http://crowdflower.com/ [accessed 2012-10-10] [WebCite Cache ID 6BJaZqlrF]
7. Ambati V, Vogel S. Can crowds build parallel corpora for machine translation systems? Stroudsburg, PA: Association for Computational Linguistics; 2010 Presented at: NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk; 2010; Los Angeles, CA p. 62-65.
8. Denkowski M, Al-Haj H, Lavie A. Turker-assisted paraphrasing for English-Arabic machine translation. Stroudsburg, PA: Association for Computational Linguistics; 2010 Presented at: NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk; 2010; Los Angeles, CA p. 66-70.
9. Gao Q, Vogel S. Semi-supervised word alignment with mechanical turk. Stroudsburg, PA: Association for Computational Linguistics; 2010 Presented at: NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk; 2010; Los Angeles, CA p. 30-34.
10. Bloodgood M, Callison-Burch C. Using mechanical turk to build machine translation evaluation sets. Stroudsburg, PA: Association for Computational Linguistics; 2010 Presented at: NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk; 2010; Los Angeles, CA p. 208-211.
11. Audhkhasi K, Georgiou PG, Narayanan SS. Analyzing quality of crowd-sourced speech transcriptions of noisy audio for acoustic model adaptation. : IEEE; 2012 Mar 25 Presented at: International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2012; Kyoto, Japan p. 4137-4170. [doi: 10.1109/ICASSP.2012.6288829]
12. Evanini K, Higgins D, Zechner K. Using amazon mechanical turk for transcription of nonnative speech. Stroudsburg, PA: Association for Computational Linguistics; 2010 Presented at: NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk; 2010; Los Angeles, CA p. 53-56.
13. Lee CY, Glass J. A transcription task for crowdsourcing with automatic quality control. : ISCA; 2011 Presented at: Interspeech; 2011; Florence p. 3041.
14. Gimpel K, Schneider N, O'Connor B, Das D, Mills D, Eisenstein J, et al. Part-of-speech tagging for Twitter: annotation, features, and experiments. Stroudsburg, PA: Association for Computational Linguistics; 2011 Presented at: 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies; 2011; Portland, Oregon p. 42-47.
15. Yano T, Resnik P, Smith NA. Shedding (a thousand points of) light on biased language. Stroudsburg, PA: Association for Computational Linguistics; 2010 Presented at: NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk; 2010; Los Angeles, CA p. 152-158.
16. Jha M, Andreas J, Thadani K, Rosenthal S, McKeown K. Corpus creation for new genres: a crowdsourced approach to PP attachment. Stroudsburg, PA: Association for Computational Linguistics; 2010 Presented at: NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk; 2010; Los Angeles, California p. 13-20.
17. Turner AM, Kirchhoff K, Capurro D. Using crowdsourcing technology for testing multilingual public health promotion materials. J Med Internet Res 2012;14(3):e79 [FREE Full text] [doi: 10.2196/jmir.2063] [Medline: 22664384]

18. Luengo-Oroz MA, Arranz A, Frean J. Crowdsourcing malaria parasite quantification: an online game for analyzing images of infected thick blood smears. J Med Internet Res 2012;14(6):e167 [FREE Full text] [doi: 10.2196/jmir.2338] [Medline: 23196001]
19. Burger JD, Doughty E, Bayer S, Tresner-Kirsch D, Wellner B, Aberdeen J, et al. Validating candidate gene-mutation relations in MEDLINE abstracts via crowdsourcing. In: Bodenreider O, Rance B, editors. Data Integration in the Life Sciences. Lecture Notes in Computer Science, Volume 7348/2012. Berlin: Springer; 2012:83-91.
20. Norman TC, Bountra C, Edwards AM, Yamamoto KR, Friend SH. Leveraging crowdsourcing to facilitate the discovery of new medicines. Sci Transl Med 2011 Jun 22;3(88):88mr1. [doi: 10.1126/scitranslmed.3002678] [Medline: 21697527]
21. Yetisgen-Yildiz M, Solti I, Xia F. Using Amazon's mechanical turk for annotating medical named entities. In: AMIA Annu Symp Proc. 2010 Presented at: AMIA 2010 Annual Symposium; 2010; Washington, DC p. 1316.
22. Yetisgen-Yildiz M, Solti I, Xia F, Halgrim SR. Preliminary experiments with Amazon's mechanical turk for annotating medical named entities. Stroudsburg, PA: Association for Computational Linguistics; 2010 Presented at: NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk; 2010; Los Angeles, California p. 180-183.
23. Kumar A, Lease M. Learning to rank from a noisy crowd. New York, NY: ACM; 2011 Presented at: 34th International ACM SIGIR Conference on Research and Development in Information Retrieval; 2011; Beijing, China p. 1221-1222. [doi: 10.1145/2009916.2010129]
24. Kumar A, Lease M. Modelling annotator accuracies for supervised learning. New York, NY: Association for Computing Machinery; 2011 Presented at: WSDM Workshop on Crowdsourcing for Search and Data Mining (WSDM); 2011; Hong Kong p. 19-22.
25. Jung HJ, Lease M. Improving consensus accuracy via Z-score and weighted voting. In: Human Computation: Papers from the 2011 AAAI Workshop (WS-11-11). Menlo Park, California: The AAAI Press; 2011 Presented at: 3rd Human Computation Workshop (HCOMP); 2011; San Francisco, CA p. 88-90.
26. Callan J. The ClueWeb09 Dataset. 2009. URL: http://lemurproject.org/clueweb09/ [accessed 2013-03-21] [WebCite Cache ID 6FHijsLsq]
27. Tang W, Lease M. Semi-supervised consensus labeling for crowdsourcing. 2011 Presented at: SIGIR Workshop on Crowdsourcing for Information Retrieval; 2011; Beijing, China p. 66-75.
28. Albright D, Lanfranchi A, Fredriksen A, Styler WF, Warner C, Hwang JD, et al. Towards comprehensive syntactic and semantic annotations of the clinical narrative. J Am Med Inform Assoc 2013 Jan 25:1-9. [doi: 10.1136/amiajnl-2012-001317] [Medline: 23355458]
29. Deleger L, Li Q, Lingren T, Kaiser M, Molnar K, Stoutenborough L, et al. Building Gold Standard Corpora for Medical Natural Language Processing Tasks. 2012. Presented at: American Medical Informatics Association 2012 Annual Symposium; 2012; Chicago, IL p. 144-153.
30. Ogren P. Knowtator: a protégé plug-in for annotated corpus construction. In: Proceedings of the 2006 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: companion volume: demonstrations. Stroudsburg, PA: Association for Computational Linguistics; 2006 Presented at: NAACL-Demonstrations '06; 2006; New York, New York p. 273-275. [doi: 10.3115/1225785.1225791]
31. Li Q, Zhai H, Deleger L, Lingren T, Kaiser M, Stoutenborough L, et al. A sequence labeling approach to link medications and their attributes in clinical notes and clinical trial announcements for information extraction. Journal of the American Medical Informatics Association 2012:1-7.
32. Lawson N, Eustice K, Perkowitz M, Yestisgen-Yildiz M. Annotating large email datasets for named entity recognition with mechanical turk. Stroudsburg, PA: Association for Computational Linguistics; 2010 Presented at: NAACL HLT Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk; 2010; Los Angeles p. 71-79.
33. Demartini G, Difallahmm DE, Cudré-Mauroux P. ZenCrowd: leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking. New York, NY: ACM; 2012 Presented at: 21st International Conference on World Wide Web (WWW '12); 2012; Lyon, France p. 469-478.
34. Solti lab code page. URL: https://code.google.com/p/soltilab/ [accessed 2013-03-21] [WebCite Cache ID 6FI2XbM4k]


## Abbreviations

**AMT:** Amazon Mechanical Turk
**CF:** CrowdFlower
**CML:** CrowdFlower Markup Language
**CSS:** common style sheet
**CTA:** clinical trial announcement
**EMU:** Extractor of Mutations
**GUI:** graphical user interface
**IAA:** inter-annotator agreement
**MEDLINE:** Medical Literature Analysis and Retrieval System Online

**NCBI:** National Center for Biotechnology Information
**NLP:** natural language processing

*Edited by G Eysenbach; submitted 06.11.12; peer-reviewed by M Luengo-Oroz, H Xu, L Hirschman; comments to author 25.11.12; revised version received 12.12.12; accepted 08.03.13; published 02.04.13.*

*Please cite as:*
*Zhai H, Lingren T, Deleger L, Li Q, Kaiser M, Stoutenborough L, Solti I*
*Web 2.0-Based Crowdsourcing for High-Quality Gold Standard Development in Clinical Natural Language Processing*
*J Med Internet Res 2013;15(4):e73*
*URL: http://www.jmir.org/2013/4/e73/*
*doi:10.2196/jmir.2426*
*PMID:23548263*

Original Paper

# Tweaking and Tweeting: Exploring Twitter for Nonmedical Use of a Psychostimulant Drug (Adderall) Among College Students

Carl L Hanson[1], PhD; Scott H Burton[2]; Christophe Giraud-Carrier[2], PhD; Josh H West[1], MPH, PhD; Michael D Barnes[1], PhD; Bret Hansen[1], BS(Current)

[1]Computational Health Science Research Group, Department of Health Science, Brigham Young University, Provo, UT, United States

[2]Computational Health Science Research Group, Department of Computer Science, Brigham Young University, Provo, UT, United States

**Corresponding Author:**
Carl L Hanson, PhD
Computational Health Science Research Group, Department of Health Science
Brigham Young University
213 Richards Building
Provo, UT, 84602
United States
Phone: 1 (801) 422 9103
Fax: 1 (801) 422 0273
Email: Carl_Hanson@byu.edu

## Abstract

**Background:** Adderall is the most commonly abused prescription stimulant among college students. Social media provides a real-time avenue for monitoring public health, specifically for this population.

**Objective:** This study explores discussion of Adderall on Twitter to identify variations in volume around college exam periods, differences across sets of colleges and universities, and commonly mentioned side effects and co-ingested substances.

**Methods:** Public-facing Twitter status messages containing the term “Adderall” were monitored from November 2011 to May 2012. Tweets were examined for mention of side effects and other commonly abused substances. Tweets from likely students containing GPS data were identified with clusters of nearby colleges and universities for regional comparison.

**Results:** 213,633 tweets from 132,099 unique user accounts mentioned “Adderall.” The number of Adderall tweets peaked during traditional college and university final exam periods. Rates of Adderall tweeters were highest among college and university clusters in the northeast and south regions of the United States. 27,473 (12.9%) mentioned an alternative motive (eg, study aid) in the same tweet. The most common substances mentioned with Adderall were alcohol (4.8%) and stimulants (4.7%), and the most common side effects were sleep deprivation (5.0%) and loss of appetite (2.6%).

**Conclusions:** Twitter posts confirm the use of Adderall as a study aid among college students. Adderall discussions through social media such as Twitter may contribute to normative behavior regarding its abuse.

***(J Med Internet Res 2013;15(4):e62)*** doi:10.2196/jmir.2503

**KEYWORDS**

Adderall; Twitter; social media; prescription drug abuse

## Introduction

The mixed salt amphetamine Adderall, commonly prescribed as a treatment of Attention Deficit Hyperactivity Disorder (ADHD), is the most commonly abused prescription stimulant among college students [1]. Colleges, as well as medical and dental schools, report abuse rates of stimulant ADHD medications [2,3] ranging from a low of 8.1% to a high of 43% [4,5]. According to the National Survey on Drug Use and Health, 6.4% of college students aged 18-22 abused Adderall in the past year [6]. Given high academic expectations and competition in college settings, some students turn to prescription stimulants like Adderall as a study aid to improve concentration and increase mental alertness [7-9]. Rates of nonmedical use or abuse of ADHD drugs tend to be higher at colleges and universities where admission standards are higher [10]. A contributing factor to abuse of ADHD drugs is attention difficulties and the notion that these drugs can help with academic success [11]. DeSantis confirmed this finding and

reported a higher tendency toward abuse among fraternity members during periods of high academic stress [12].

Other studies have affirmed racial and gender discrepancies in stimulant drug abuse as well as a correlation between prescription drug abuse and other illicit drug use among college students [4,7,12,13]. Nonacademic motivations are also common and include, but are not limited to, counteracting the effects of other drugs, feeling a high, or as an appetite suppressor [7] as well as self-diagnosis of ADHD [11,14]. A contributing factor for illicit drug abuse and prescription stimulant abuse among college students is the misperception that the vast majority of their peers use drugs [15,16]. Elevated misperceptions about the prevalence of drug use among peers are attributed to the traditional media's (eg, popular television depictions of college students using Adderall to gain academic advantages) portrayal of abuse. Misperception of reality is believed to be a leading contributor to increased levels of acceptance of abusive drug behavior, community norms for abuse, and higher levels of abuse [16]. Additional misperceptions such as the lack of danger of abusing prescription stimulants have also been found to contribute to justifications for illicit use [12].

Social media provides a relatively new and untapped resource for monitoring and understanding public health problems. As a surveillance tool, real-time data obtained through social media can be collected and analyzed quicker than traditional public health assessment tools such as questionnaires. In addition, research using social media provides an avenue for observing discussion between people in their natural interactions with one another, eliminating the Hawthorne Effect, where the presence of the researchers biases the response. Likewise, because people make statements as they occur, memory recall biases common with cross-sectional surveys or questionnaires are reduced. With the expansion of the Internet and social media, new fields of study such as infodemiology and infoveillance have emerged and represent "the science of distribution and determinants of information in an electronic medium, specifically the Internet, or in a population, with the ultimate aim to inform public health and public policy" (page 3) [17].

Studies have demonstrated the utility of online information for understanding public health problems and their determinants. Using information obtained on trends in Internet searches, researchers have predicted outbreaks of influenza [18-20], listeriosis from contaminated foods [21], and gastroenteritis and chickenpox [22]. The feasibility of using online information for epidemiological intelligence purposes has led to the creation of proprietary systems such as Google Flu Trends, which is an online search query system that has demonstrated the ability to track regional outbreaks of influenza 7-10 days in advance of conventional Centers for Disease Control and Prevention (CDC) mechanisms for reporting [23]. In addition, Healthmap represents a public system for aggregating large amounts of online information (eg, news sources) for the purpose of monitoring global disease activity [24].

Recognizing the wealth of user-generated information produced by people through their participation with social media, researchers have begun tapping or mining this information to better understand health outcomes and even health behavior. For example, Corley, Cook, Mikler, and Singh [25] mined text data in the blogosphere for "influenza" and "flu." Their findings revealed trends in posts about the flu that were consistent with CDC report data. Several studies mined YouTube content for information relative to immunizations [26], H1N1 influenza pandemic [27], smoking cessation [28], cardiopulmonary resuscitation [29], kidney stones [30], and prostate cancer [31]. To date, no identified study has analyzed user-generated content in social media to describe the nonmedical use of Adderall.

The purpose of this study was to leverage the power of social media (ie, Twitter) to better understand Adderall abuse as a study aid among college and university students. More specifically, the following research questions were examined: (1) When do Twitter users typically tweet about Adderall?, (2) To what extent do tweets about Adderall abuse differ among various college and university clusters in the United States?, (3) What, if any, substances do Twitter users tweet about commonly abusing in combination with Adderall?, and (4) What common side effects are mentioned? Twitter was selected as the social media application for data collection because of its appeal with young adults including the ubiquitous research design advantages identified above. Twenty-six percent of all Internet users age 18-29 and 31% of all Internet users age 18-24 are also Twitter users [32]. Finally, using Twitter as a data source affords the ability to observe nationwide (and even international) behaviors simultaneously, as opposed to arbitrarily restricting a study to only a few regions. The use of social media data, in particular tweets, remains largely a novel concept for public health researchers. Questions surrounding the validity and utility of the data exist. Furthermore, little is known about the extent to which Twitter users might actually tweet about potentially sensitive health topics, such as Adderall abuse. Studies like the current one contribute to a type of validity testing process whereby researchers can determine the extent to which trends in Twitter content coincide with documented patterns of behavior.

## Methods

### Procedures

Twitter is a popular online social media website in which users post status updates, or "tweets," that are limited to 140 characters. Public tweets are available and given without expectation of privacy. In addition, Twitter provides an Application Programing Interface (API), enabling programmatic consumption of the data. Specifically, the Twitter streaming API supplies tweets in real-time matching any given filter criteria. For example, using the keyword filter of "Adderall," all tweets mentioning the substance are collected.

In addition to the content of tweets, many users also provide location indicators [33]. Specifically in Twitter, users can potentially supply exact global positioning system (GPS) coordinates (eg, from a smart phone or other GPS-enabled device) or a GPS specified place (such as a neighborhood or city). Note that users providing only state or country level GPS were not included. Furthermore, tweets were excluded if they did not originate in the United States, based on GPS location. This GPS data can be used to associate a twitter user with a

nearby college/university. However, because many college campuses are within close geographic proximity, Twitter users may not necessarily have association with the campus to which they are physically nearest. Because of this proximity issue, rather than try to determine which of two nearby colleges should be used, the colleges were instead grouped into a cluster and treated as a single entity. Colleges and university with a student population of 10,000 or more were identified using the National Center for Education Statistics database. Clusters were determined using hierarchical agglomerative clustering (HAC) [34,35] with complete linkage with a cutoff distance of 150 miles. HAC produces a dendrogram of the complete sequence of nested clusterings, as follows. HAC starts by assigning each college to its own cluster. Then, the two closest clusters are merged into a single new cluster. This pairwise merging process is repeated until a single cluster containing all of the colleges is obtained. Although we have a distance defined over colleges, HAC also needs a distance over clusters. Several distance measures may be considered, the most common of which are complete linkage, which uses the maximum distance between all pairs of objects across clusters, single linkage, which takes the minimum distance, and average linkage, which computes the average of all intercluster distances. We chose complete linkage here as it tends to create more compact, clique-like clusters [36]. Given the fully nested sequence of clusterings, the choice of a specific final grouping is typically made by selecting a level at which to cut through the dendrogram, and defining the clusters as the groups of elements hanging from the subtrees whose top branches intersect with the horizontal line corresponding to the chosen level. Our cut point of 150 miles means that pairs of colleges in a cluster were no more than 150 miles apart.

The student body population of a cluster was determined by summing the populations of each included college. Twitter users are then associated with the nearest college cluster if a college in that cluster is within 100 miles of the user's GPS location.

## Measures

Keywords related to co-ingestion with other drugs, alternative motives, and possible side effects are shown in Table 1. A case-insensitive comparison was performed to count the number of tweets containing the keywords specified. Where multiple words are given as a single term, it was considered an exact phrase.

## Data Analysis

After the tweets were obtained from the Twitter API, the data were imported to Microsoft Excel spreadsheets and then into SPSS version 20 for analysis. Frequencies, percentages, means, medians, and standard deviations were used to describe the Adderall abuse. ArcGIS 10 was used to create maps of rates for GPS Adderall tweeters.

The Institutional Review Board (IRB) at Brigham Young University approved this study.

**Table 1.** Search terms for alternative motive, co-ingestion, and side effects.

| Topic | Subtopic | Search Terms |
|---|---|---|
| **Alternative Motive** (study aid) | | test, final, finals[a], study[a], studi[a], college, class, midterm, exam, homework, paper, essay, project, school, cram[a], quiz, assignment, all-night[a], allnight[a] |
| **Co-ingestion** | | |
| | Alcohol-related | alcohol, wine, vodka, shots, patron, booz[a], margarita, mimosa, beer, drink[a], bud |
| | Stimulants | coffee, caffeine, red bull, monster, no dose, no doze, 5 hour energy, five hour energy, rockstar |
| | Cocaine-related | cocaine, coke, crack, rock, freebase |
| | Marijuana | marijuana, MJ, pot, weed, grass, reefer, Mary Jane |
| | Anti-anxiety | xanax, tranquilizer, valium, beanies, ativan, benzo[a] |
| | Meth-related | crystal, meth, methamphetamine, amphetamine |
| **Side Effects** | | |
| | Sleep deprivation | tired, awake, sleep[a], slept, insomnia, restless, asleep, trouble sleeping, |
| | Anxiety | anxiety, anxious, antsy, jitter[a], shak[a], nerv[a], nervous, uneas[a], worry, tense, tension, dread, restless[a] |
| | Teeth grinding | teeth, tooth, grind[a], file, grat[a], grit[a], clench[a], grit[a], gnash[a], scrap[a] |
| | Diarrhea | diarrhea, diarrea, diarhea, the runs, squirts |
| | Weakness | weak[a], feeble, puny, scrawny |
| | Dizziness | dizz[a], faint, wobbly, shaky, lightheaded, light-headed, woozy, dazed |
| | Headache | headache, migraine, migrain, migrane |
| | Sweating | sweat[a], perspir[a], drip[a] |
| | Nausea/vomiting | nausea, vomit[a], throw up, stomach pain, stomach ache, upset stomach, puke, barf, heave |
| | Loss of appetite | hungry, food, eat[a], ate, weight, appetite, meal, thin, skinny, starv[a], slim[a], slender |
| | Obsessive compulsive behavior | can't stop, clean[a], brush[a] teeth, wash[a] hands, nails, nail-biting |

[a] A wildcard matching zero or more additional letters.

## Results

Using the Twitter Streaming API with the keyword filter “Adderall,” all tweets mentioning Adderall for the dates of November 29, 2011, to May 31, 2012, were collected. There were 14,282 tweets from users whose screen-names included “Adderall” or “pharm” that were removed from the sample because they were not representative of typical users, but rather those that were pushing or promoting Adderall or other pharmaceuticals. The resulting sample consisted of 213,633 tweets mentioning the term Adderall, from 132,099 unique user accounts.

The vast majority of tweets discussed Adderall use in a joking, sarcastic, or casual manner. Observed tweets included (original spelling and punctuation preserved): “I need adderall. Can't focus on studying or finishing these reviews”, “this whole no adderall for the past 3 days is really getting to me #StillDoingWork #DontKnoHowTho”, “Does anyone have adderall? #desperate”, “adderall + school = winning”, “wish i had adderall to get my room cleaned faster”, “Adderall stockpile for finals”, “We would all graduate with a 4.0 if adderall was sold over the counter”, “Running on coffee and Adderall”, “yay for adderall-induced optimism #givemeaprescription”, and “Adderall, Coffee, Red Bull. Epic focus. Or a heart attack.” Note that words beginning with “#” are hashtags, or user-defined topics that are often used in Twitter as a means of self-classification.

Table 2 lists the number of tweets matching each of the categories defined in Table 1. It should be noted that the results shown in Table 2 capture words that occur in the same tweet as

the term Adderall. In this sense, they may be a conservative underestimate of actual events because it is possible that a user may tweet about Adderall but mention a side effect, motive, or another substance in another tweet. Because subtopics are not mutually exclusive, some tweets match multiple subtopics and are counted for each. Thus, the total number of *unique* tweets for a topic is not a sum of the subtopic values.

## Adderall Use by Hour, Day, and Week

Figure 1 illustrates the average number of Adderall-related tweets per day of the week, over the course of the study. Tweets tend to peak on Wednesday and reach a low on Saturday. As shown in Figure 2, the number of Adderall tweets per day varied significantly throughout the year, with consistently more tweets on the weekdays than the weekends. Large spikes in Twitter conversations were observed during the months of December and May—during traditional final exam times. The one-way ANOVA results indicate a significant difference between Adderall mentions between weeks ($P<.001$). Tweets regarding Adderall peaked December 13th at 2813 and April 30th at 2207 and dropped to a low of 292 on December 25th and 440 on May 27th. Over the course of 6 months while data were collected, the mean number of Adderall tweets per day was 930 with a median of 855. The large spike on May 30-31 was attributed to a US Food and Drug Administration news release warning consumers of counterfeit versions of Adderall being sold on the Internet in response to its' being on the FDA's drug shortage list [37]. This FDA news release was reported by news agencies, and links to the subsequent stories were tweeted by many users. The 10 days in the middle of April when no tweets were observed is the result of a failure of the investigators' servers.

**Table 2.** Frequency distribution of Adderall tweets for search terms.

| Topic | Subtopic | n | % |
|---|---|---|---|
| **Alternative Motive** (study aid) | | 27,473 | 12.9 |
| **Co-ingestion** | | | |
| | Alcohol-related | 10,229 | 4.8 |
| | Stimulants | 10,043 | 4.7 |
| | Cocaine-related | 1993 | 0.9 |
| | Marijuana | 1696 | 0.8 |
| | Anti-anxiety | 881 | 0.4 |
| | Meth-related | 788 | 0.4 |
| | Total Unique Co-ingestion Tweets | 24,167 | 11.3 |
| **Side Effects** | | | |
| | Sleep Deprivation | 10,687 | 5.0 |
| | Anxiety | 1204 | 0.6 |
| | Teeth Grinding | 605 | 0.3 |
| | Diarrhea | 11 | 0.01 |
| | Weakness | 140 | 0.07 |
| | Dizziness | 77 | 0.04 |
| | Headache | 223 | 0.1 |
| | Sweating | 381 | 0.2 |
| | Nausea/vomiting | 154 | 0.07 |
| | Loss of Appetite | 5562 | 2.6 |
| | Obsessive Compulsive Behavior | 1937 | 0.9 |
| | Total Unique Side Effect Tweets | 19,539 | 9.1 |

**Figure 1.** Adderall-related tweets by day of the week.

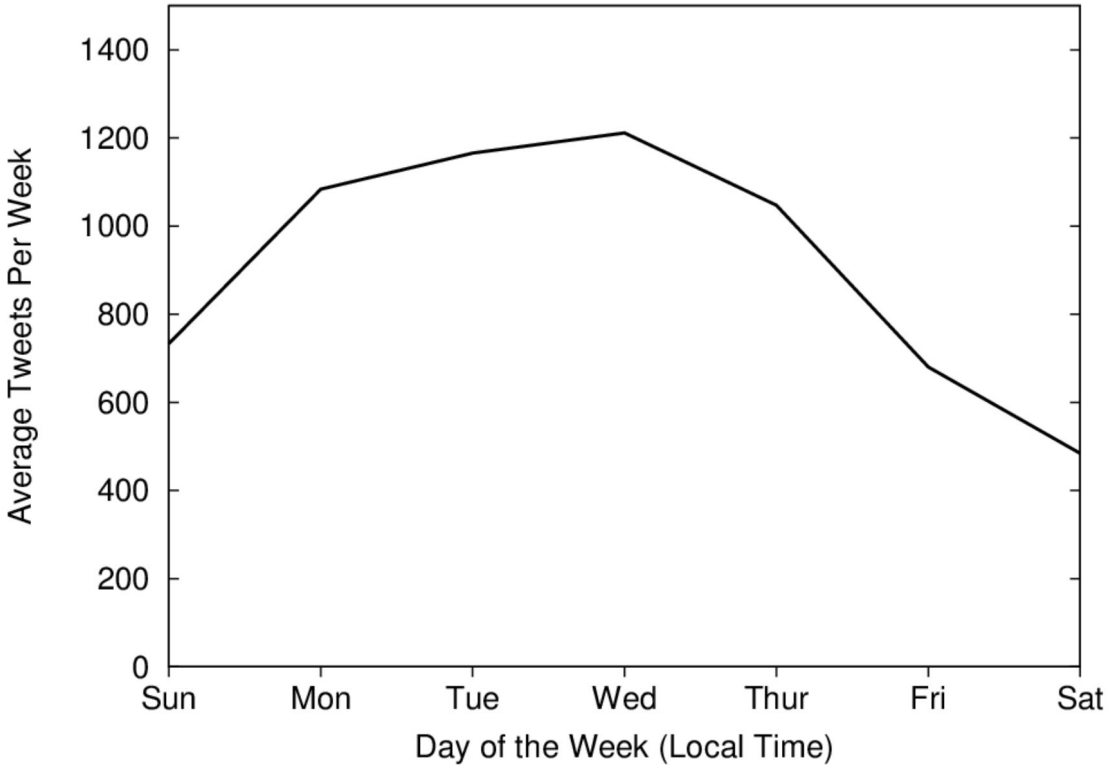
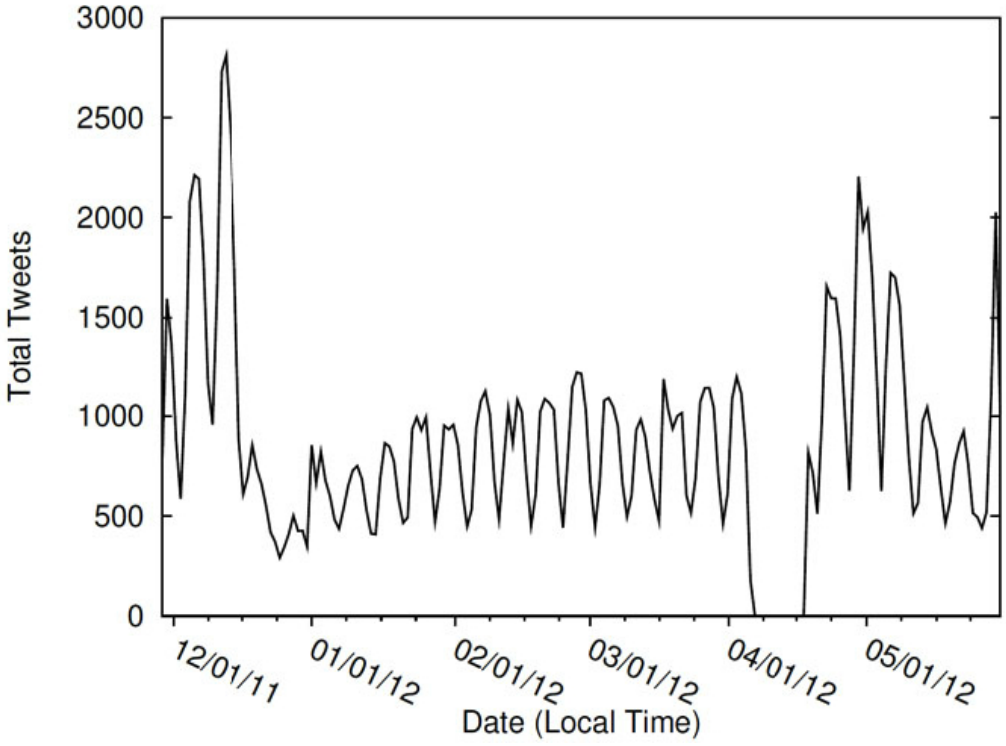



**Figure 2.** Distribution of Adderall-related tweets over 6 months.




## College and University Clusters

Of the 213,633 tweets referencing Adderall, 27,473 (12.9%) also included reference to an alternative motive for use (eg, finals, studying, project, all-nighter), as shown in Table 2. Several of these alternative motives seem to be indicative of misuse among college-age students. To focus the analysis on college-age students, Adderall tweets were analyzed in clusters of colleges and universities that were within 150 miles of each other. A total of 586 colleges and universities in the United States were identified with a student body population of at least 10,000. Colleges and universities within 150 miles of each other

resulted in a total of 87 clusters ranging in size from 1 to 48 colleges and universities in each cluster. The mean size of student-body population per cluster was 131,562, and the median was 93,281.

Of the 132,099 unique users in the sample, 3698 (2.8%) provided GPS data. In order to restrict this set of GPS-enabled users to include only those users who are likely to be students, we obtained the 3200 most recent tweets (the maximum provided by Twitter) from each user with GPS data and searched these tweets for the following student-related terms: “homework”, “teacher”, “professor”, “class”, “final”, “test”, “exam”, and “study.” Of the 3698 users with GPS information, 2335 (60.7%) included one of these student-related terms in their tweets and are referred to as GPS Adderall Tweeters.

Figure 3 illustrates the 150-mile college clusters in the contiguous 48 states of the United States according to the rate of GPS Adderall Tweeters per 100,000 students, where the center of the circle is the average of the locations of the colleges in the cluster, and the size of the circle corresponds to the rate. Table 3 lists the ten clusters with the highest rates, and Table 4 lists the ten clusters with the lowest rates. Cluster identifications (ID) represent the state(s) to which the majority of colleges and universities in the cluster belong. As shown in these tables, the amount of GPS Adderall Tweeters per 100,000 students ranges from a high of 66.4 in the Vermont cluster and 54.6 in the Massachusetts cluster to a low of 1.4 in the South-Eastern Texas cluster and 2.1 in the Central Illinois cluster. Rates reveal a greater rate of GPS Adderall Tweeter in the northeast and south regions of the United States.

**Table 3.** Top 10 rates of Adderall tweets for 150-mile college and university clusters in the United States.

| Rank | ID[a] | Rate | GPS Adderall Tweeters | Total Cluster Population | Number in the Cluster |
|---|---|---|---|---|---|
| 1 | Vermont | 66.4 | 9 | 13,554 | 1 |
| 2 | Massachusetts | 54.6 | 162 | 296,704 | 16 |
| 3 | Alabama | 52.2 | 38 | 72,748 | 3 |
| 4 | South Carolina, Southern Georgia | 48.8 | 57 | 116,891 | 6 |
| 5 | Central Georgia | 44.2 | 52 | 117,765 | 6 |
| 6 | North Georgia, Southern South Carolina | 44.0 | 36 | 81,773 | 4 |
| 7 | Northern Florida | 44.0 | 18 | 40,921 | 3 |
| 8 | Southern Pennsylvania, Northern West Virginia | 42.9 | 43 | 100,336 | 5 |
| 9 | Ohio | 37.3 | 54 | 144,659 | 7 |
| 10 | Western North Carolina, Eastern Tennessee | 37.0 | 41 | 110,718 | 6 |

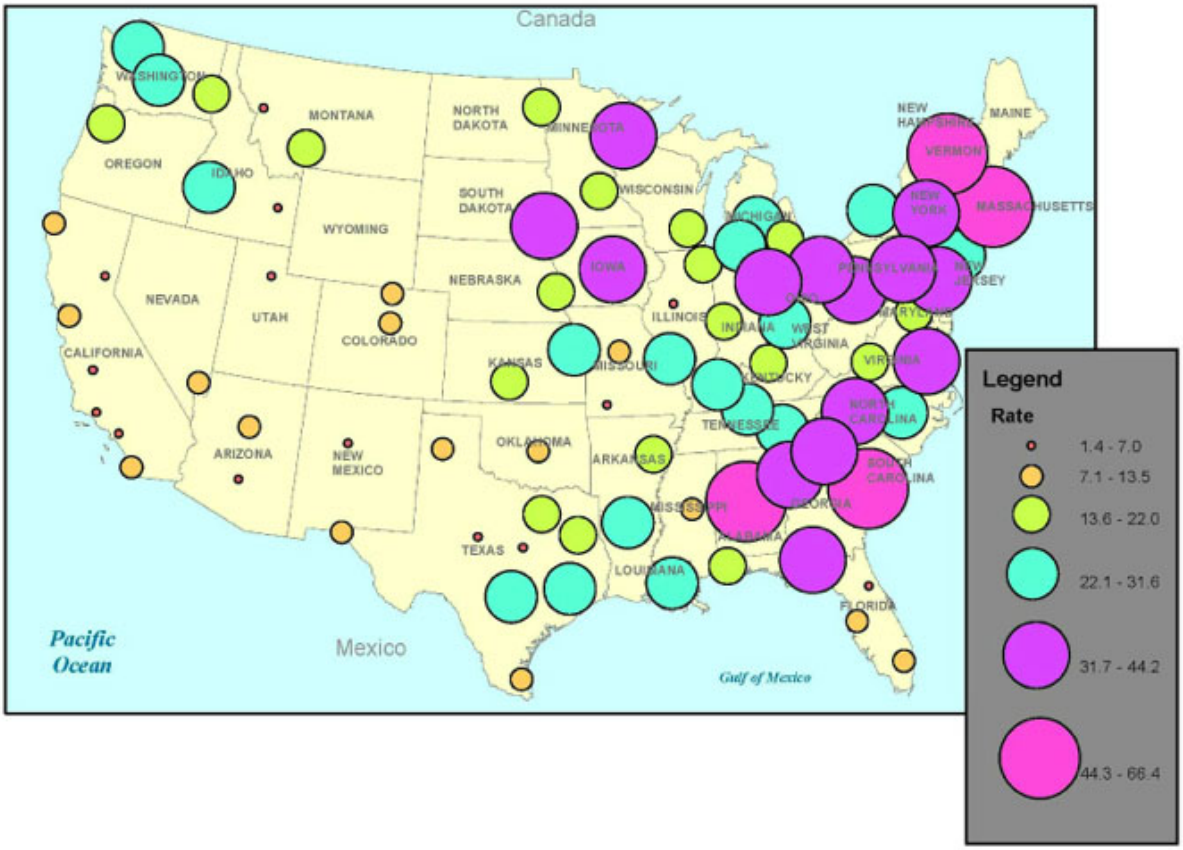
[a] ID represents the state(s) to which the majority of colleges and universities in the cluster belong.

**Table 4.** Bottom 10 rates of Adderall tweets for 150-mile college and university clusters in the United States.

| Rank | ID[a] | Rate | GPS Adderall Tweeters | Total Cluster Population | Number in the Cluster |
|---|---|---|---|---|---|
| 77 | Central Texas | 5.8 | 3 | 52,076 | 3 |
| 78 | Alaska | 5.5 | 1 | 18,154 | 1 |
| 79 | Southern California | 5.5 | 55 | 1,008,210 | 48 |
| 80 | Puerto Rico | 5.0 | 3 | 60,579 | 4 |
| 81 | Northern Nevada | 4.5 | 3 | 66,242 | 4 |
| 82 | New Mexico | 3.9 | 3 | 77,236 | 3 |
| 83 | Northern Utah, Southern Idaho | 3.6 | 1 | 27,476 | 1 |
| 84 | Northern California | 3.5 | 4 | 115,026 | 7 |
| 85 | Central Illinois | 2.1 | 1 | 46,797 | 3 |
| 86 | South-Eastern Texas | 1.4 | 1 | 69,949 | 3 |

[a] ID represents the state(s) to which the majority of colleges and universities in the cluster belong.

**Figure 3.** Rates of Adderall tweets by 150 mile college clusters in the United States (rate per 100,000 students).




### Co-ingestion and Side Effects

A total of 19,125 (8.9%) tweets also mentioned another substance along with Adderall in their tweet (see Table 2). Analysis revealed that the most common substance terms were alcohol-related (4.8%, n=10,229) and stimulants, such as coffee or Red Bull (4.7%, n=10,043). Other substances were cocaine-related (0.9%, n=1993), marijuana (0.8%, n=1696), methamphetamine-related (0.4%, n=788), and depressants, such as Xanax and painkillers (0.3%, n=728).

Sleep deprivation (5.0%, n=10,687) and loss of appetite (2.6%, n=5562) were the most common side effects associated with Adderall tweets (see Table 2). Diarrhea (0.01%) was the least common side effect mentioned followed by weakness (0.01%, n=140) and nausea/vomiting (0.07%, n=154).

## Discussion

This study demonstrated the use of Twitter posts (ie, tweets) as a way to examine Adderall abuse among a sample of college students in the United States. More specifically, the study sought to determine: (1) When do Twitter users typically tweet about Adderall?, (2) To what extent do tweets about Adderall abuse differ among various college clusters in the United States?, (3) What, if any, substances do Twitter users tweet about commonly abusing in combination with Adderall?, and (4) What common side effects are mentioned?

Findings indicate that Twitter posts regarding Adderall vary across day of the week and week of the month. Consistent with traditional college final exams schedules, tweets regarding Adderall peaked during December and May. Similarly, tweets regarding Adderall peaked during the middle of the academic week and declined to fewer mentions over the weekend. These findings are consistent with previous research that has suggested that college students who abuse prescription ADHD stimulants do so primarily during times of high academic stress [12]. In addition, preexisting attention difficulties have been shown to be a predictor of nonmedical use of prescription ADHD medication in order for college students to experience greater academic success [11].

Grouping colleges within 150-mile clusters ultimately provided a mechanism for comparing geographic regions within the United States. Analysis of these college clusters revealed a concentration of GPS Adderall tweeters along the northeastern portion of the United States and in some of the southern states. The rates of GPS Adderall twitters per 100,000 students in the east and south clearly indicated greater Twitter conversations related to the use and abuse of Adderall. These findings are consistent with previous studies that examined the nonmedical use of prescription stimulants. McCabe [4] observed

geographical patterns of nonmedical use of prescription stimulants with higher rates of use among college students in the north-eastern region of the country. Additionally, these findings are consistent with the Monitoring the Future study where higher rates of nonmedical methylphenidate use were found among college-age young adults in the northeastern region of the United States [38]. Other studies at select colleges in the east have shown high rates of nonmedical use of prescription stimulants [8,9].

Additional research is needed to better understand the reasons for geographical variations in use. One possible explanation includes the fact that the U.S. fraternity/sorority system has deep historical roots at northeastern colleges and universities, and prevalence of nonmedical use of prescription stimulants is higher among fraternity/sorority members [4]. Future research might explore the link between nonmedical use of prescription stimulants and the geographical distribution of colleges and universities and their admission standards, student/family income, as well as the distribution of prescription drug monitoring program in the United States. Research has associated nonmedical use of prescription stimulants with competitive admission standards [4] and students coming from families with higher incomes [39].

Geographical findings can provide practitioners with evidence necessary for prioritizing intervention resources for targeting priority populations. This study has demonstrated how grouping can occur; however, and more importantly, it provides a social media solution for segmenting a broader population into more meaningful and manageable groups for intervention purposes. Colleges can be clustered in numerous different ways as needed and defined by researchers.

Because social media is, by its very nature, a *social* endeavor, the users' postings can have a great impact on the social norms of others. This is particularly relevant in the context of drug abuse, where drug abuse behavior can be represented. Social norms theory suggests that individual behavior (eg, drug use) is influenced by individual perceptions of what is perceived as "normal" or "typical." This theory is rooted in Social Cognitive Theory [40] as well as the Theory of Planned Behavior [41]. In this light, the data that 8.9% of Adderall tweets mention another substance in the same tweet is significant because it may influence others to think that co-ingestion is normal and not dangerous. This is particularly troubling because it is through poly drug use or co-ingestion that morbidity and mortality risk increases. Poly drug use occurs among college Adderall abusers and combining Adderall with other stimulants like cocaine increases risk of heart attack and stroke [6]. Also in this regard, even tweets that are sarcastic, joking, or simply restating song lyrics, are relevant in their misrepresentations because of their impact on social norms.

Nearly 1 in 10 tweets included in this sample referenced a side effect of Adderall use/abuse. Effects relative to sleep deprivation and loss of appetite were discussed the most. Whereas more tweeters discussed an alternative motive for use (ie, study aid), individual tweeter perception of the benefits of Adderall use (eg, study aid) may outweigh the costs of use (eg, side effect such as irritability). Future research might further explore individual perceptions of Adderall side effects among college students to gain a better understanding of why some college students abuse, while others do not.

## Limitations

These findings should be interpreted based on the following limitations. First, not every Adderall tweet is related to actual use. For example, we observed song lyrics that impact these counts, such as the two often quoted lines "College hoes love alcohol and popping adderall" and "I've been up for 3 days… adderall and redbull." In our sample, there were 4275 tweets that have the words "college hoes love" and 894 that have the words "been up for three/3 days". These numbers likely inflate the number of matches for "college", "alcohol", and "redbull" above the number of people tweeting about actually using these substances. However, as discussed, even sarcastic mentions, or the quotation of song lyrics, are pertinent because of the impact they may have on social norms. Second, our study did not consider misspellings of the word "Adderall" or other ADHD medications, such as Ritalin. While our sample would have been increased by these inclusions, it is not likely that their absence resulted in any particular sampling bias. Third, our analysis focused exclusively on public tweets. It is unclear, and indeed difficult to assess, what the impact of other tweets (eg, direct messages) may have on our results. Fourth, our analysis focused only on colleges and universities with a student population of 10,000 or more. No attempt was made to designate whether the colleges and universities in this sample were on a quarter or semester system. Finally, the keyword approach to identifying college students may have included other students (eg, high school) or others that simply mentioned academic-related terms. While these additional users could inflate our overall values, we have no reason to believe they would be substantially biased toward different areas of the nation.

## Conclusions

The twitter-based surveillance methodology in this study produced similar findings to traditional survey designs. In response to the noted research questions, Twitter posts regarding Adderall vary across day of the week and week of the month among users. Consistent with college traditional final exams schedules, tweets regarding Adderall peaked during December and May. Similarly, tweets regarding Adderall peaked during the middle of the academic week and declined to fewer mentions over the weekend, which suggests that college students who abuse prescription ADHD stimulants do so primarily during times of high academic stress.

Additionally, tweets about Adderall abuse differ among various college clusters in the United States. Using 150-mile college clusters, regional comparisons identified a concentration of GPS Adderall tweeters along the northeastern portion of the United States and in some of the southern states, and thus indicate greater Twitter conversations related to the use and abuse of Adderall. Further, co-ingestion of other substances, notably alcohol, stimulants (such as coffee or Red Bull), cocaine-related, marijuana, methamphetamine-related, and depressants, (such as Xanax and painkillers) are the substances most commonly mentioned with Adderall. Such poly drug use or co-ingestion

is known to increase morbidity and mortality risk. Finally, the most common side effects associated with Adderall tweets include sleep deprivation and loss of appetite. Thus, Adderall abuse is associated with college or university life. Given the risks and trends for Adderall acceptance among college-age students, there is a need to renew interest and priorities to influence college campus norms, promote the safe and legal use of these substances, and promote stronger student wellbeing and study habits to better manage the academic demands and pressures that are typical on college campuses in the United States.

## Conflicts of Interest

None declared.

## References


1. Ford JA, Arrastia MC. Pill-poppers and dopers: a comparison of non-medical prescription drug use and illicit/street drug use among college students. Addict Behav 2008 Jul;33(7):934-941. [doi: 10.1016/j.addbeh.2008.02.016] [Medline: 18403131]
2. McNiel AD, Muzzin KB, DeWald JP, McCann AL, Schneiderman ED, Scofield J, et al. The nonmedical use of prescription stimulants among dental and dental hygiene students. J Dent Educ 2011 Mar;75(3):365-376 [FREE Full text] [Medline: 21368261]
3. Herman L, Shtayermman O, Aksnes B, Anzalone M, Cormerais A, Liodice C. The use of prescription stimulants to enhance academic performance among college students in health care programs. J Physician Assist Educ 2011;22(4):15-22. [Medline: 22308929]
4. McCabe SE, Teter CJ, Boyd CJ. Medical use, illicit use and diversion of prescription stimulant medication. J Psychoactive Drugs 2006 Mar;38(1):43-56 [FREE Full text] [Medline: 16681175]
5. Advokat CD, Guidry D, Martino L. Licit and illicit use of medications for Attention-Deficit Hyperactivity Disorder in undergraduate college students. J Am Coll Health 2008;56(6):601-606. [doi: 10.3200/JACH.56.6.601-606] [Medline: 18477513]
6. Substance Abuse and Mental Health Services Adminstration. The NSDUH Report: Nonmedical use of Adderall among full-time college students URL: http://oas.samhsa.gov/2k9/adderall/adderall.htm [accessed 2012-12-17] [WebCite Cache ID 6CzLAcN9e]
7. Teter CJ, McCabe SE, Cranford JA, Boyd CJ, Guthrie SK. Prevalence and motives for illicit use of prescription stimulants in an undergraduate student sample. J Am Coll Health 2005;53(6):253-262. [doi: 10.3200/JACH.53.6.253-262] [Medline: 15900989]
8. Babcock Q, Byrne T. Student perceptions of methylphenidate abuse at a public liberal arts college. J Am Coll Health 2000 Nov;49(3):143-145. [doi: 10.1080/07448480009596296] [Medline: 11125642]
9. Low KG, Gendaszek AE. Illicit use of psychostimulants among college students: A preliminary study. Psychology, Health & Medicine 2002 Aug 2002;7(3):283-287. [doi: 10.1080/13548500220139386]
10. McCabe SE, Knight JR, Teter CJ, Wechsler H. Non-medical use of prescription stimulants among US college students: prevalence and correlates from a national survey. Addiction 2005 Jan;100(1):96-106. [doi: 10.1111/j.1360-0443.2005.00944.x] [Medline: 15598197]
11. Rabiner DL, Anastopoulos AD, Costello EJ, Hoyle RH, McCabe SE, Swartzwelder HS. Motives and perceived consequences of nonmedical ADHD medication use by college students: are students treating themselves for attention problems? J Atten Disord 2009 Nov;13(3):259-270. [doi: 10.1177/1087054708320399] [Medline: 18664714]
12. DeSantis AD, Webb EM, Noar SM. Illicit use of prescription ADHD medications on a college campus: a multimethodological approach. J Am Coll Health 2008;57(3):315-324. [doi: 10.3200/JACH.57.3.315-324] [Medline: 18980888]
13. White BP, Becker-Blease KA, Grace-Bishop K. Stimulant medication use, misuse, and abuse in an undergraduate and graduate student sample. J Am Coll Health 2006;54(5):261-268. [doi: 10.3200/JACH.54.5.261-268] [Medline: 16539218]
14. Judson R, Langdon SW. Illicit use of prescription stimulants among college students: prescription status, motives, theory of planned behaviour, knowledge and self-diagnostic tendencies. Psychol Health Med 2009 Jan;14(1):97-104. [doi: 10.1080/13548500802126723] [Medline: 19085316]
15. Lewis MA, Neighbors C. Gender-specific misperceptions of college student drinking norms. Psychol Addict Behav 2004 Dec;18(4):334-339. [doi: 10.1037/0893-164X.18.4.334] [Medline: 15631605]
16. Perkins HW, Meilman PW, Leichliter JS, Cashin JR, Presley CA. Misperceptions of the norms for the frequency of alcohol and other drug use on college campuses. J Am Coll Health 1999 May;47(6):253-258. [doi: 10.1080/07448489909595656] [Medline: 10368559]
17. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
18. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006:244-248 [FREE Full text] [Medline: 17238340]

19. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
20. Polgreen PM, Chen Y, Pennock DM, Nelson FD. Using internet searches for influenza surveillance. Clin Infect Dis 2008 Dec 1;47(11):1443-1448 [FREE Full text] [doi: 10.1086/593098] [Medline: 18954267]
21. Wilson K, Brownstein JS. Early detection of disease outbreaks using the Internet. CMAJ 2009 Apr 14;180(8):829-831 [FREE Full text] [doi: 10.1503/cmaj.090215] [Medline: 19364791]
22. Pelat C, Turbelin C, Bar-Hen A, Flahault A, Valleron A. More diseases tracked by using Google Trends. Emerg Infect Dis 2009 Aug;15(8):1327-1328 [FREE Full text] [doi: 10.3201/eid1508.090299] [Medline: 19751610]
23. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis 2009 Nov 15;49(10):1557-1564 [FREE Full text] [doi: 10.1086/630200] [Medline: 19845471]
24. Brownstein JS, Freifeld CC, Chan EH, Keller M, Sonricker AL, Mekaru SR, et al. Information technology and global surveillance of cases of 2009 H1N1 influenza. N Engl J Med 2010 May 6;362(18):1731-1735 [FREE Full text] [doi: 10.1056/NEJMsr1002707] [Medline: 20445186]
25. Corley CD, Cook DJ, Mikler AR, Singh KP. Text and structural data mining of influenza mentions in Web and social media. Int J Environ Res Public Health 2010 Feb;7(2):596-615 [FREE Full text] [doi: 10.3390/ijerph7020596] [Medline: 20616993]
26. Keelan J, Pavri-Garcia V, Tomlinson G, Wilson K. YouTube as a source of information on immunization: a content analysis. JAMA 2007 Dec 5;298(21):2482-2484. [doi: 10.1001/jama.298.21.2482] [Medline: 18056901]
27. Pandey A, Patni N, Singh M, Sood A, Singh G. YouTube as a source of information on the H1N1 influenza pandemic. Am J Prev Med 2010 Mar;38(3):e1-e3. [doi: 10.1016/j.amepre.2009.11.007] [Medline: 20171526]
28. Backinger CL, Pilsner AM, Augustson EM, Frydl A, Phillips T, Rowden J. YouTube as a source of quitting smoking information. Tob Control 2011 Mar;20(2):119-122. [doi: 10.1136/tc.2009.035550] [Medline: 20966132]
29. Murugiah K, Vallakati A, Rajput K, Sood A, Challa NR. YouTube as a source of information on cardiopulmonary resuscitation. Resuscitation 2011 Mar;82(3):332-334. [doi: 10.1016/j.resuscitation.2010.11.015] [Medline: 21185643]
30. Sood A, Sarangi S, Pandey A, Murugiah K. YouTube as a source of information on kidney stone disease. Urology 2011 Mar;77(3):558-562. [doi: 10.1016/j.urology.2010.07.536] [Medline: 21131027]
31. Steinberg PL, Wason S, Stern JM, Deters L, Kowal B, Seigne J. YouTube as source of prostate cancer information. Urology 2010 Mar;75(3):619-622. [doi: 10.1016/j.urology.2008.07.059] [Medline: 19815255]
32. Smith A, Brenner J. Twitter use 2012. 2012. URL: http://pewinternet.org/Reports/2012/Twitter-Use-2012/Findings.aspx [accessed 2012-12-17] [WebCite Cache ID 6CzHsIlpU]
33. Burton SH, Tanner KW, Giraud-Carrier CG, West JH, Barnes MD. "Right time, right place" health communication on Twitter: value and accuracy of location information. J Med Internet Res 2012;14(6):e156 [FREE Full text] [doi: 10.2196/jmir.2121] [Medline: 23154246]
34. Johnson SC. Hierarchical clustering schemes. Psychometrika 1967 Sep 1967;32(3):241-254. [doi: 10.1007/BF02289588]
35. Jain AK, Murty MN, Flynn PJ. Data clustering: a review. ACM Comput Surv 2000;31(3):264-323. [doi: 10.1145/331499.331504]
36. Jain AK, Dubes RC. Algorithms for clustering data. Englewood Cliffs, N.J: Prentice Hall; 1988.
37. U.S. Food and Drug and Aministration. FDA warns consumers about counterfeit version of Teva's Adderall. 2012. URL: http://www.webcitation.org/6CzHsIlpU [accessed 2012-12-17]
38. Johnston LD, O'Malley PM, Bachman JG. Monitoring the future national survey results on drug use, 1975-2002. Bethesda, Maryland: National Institute on Drug Abuse; 2003. Volume II: College students and adults ages 19-40 URL: http://monitoringthefuture.org/pubs/monographs/vol2_2002.pdf [WebCite Cache ID 6F2nFZ7hv]
39. Teter CJ, McCabe SE, Boyd CJ, Guthrie SK. Illicit methylphenidate use in an undergraduate student sample: prevalence and risk factors. Pharmacotherapy 2003 May;23(5):609-617. [Medline: 12741435]
40. Bandura A. Social learning theory. Englewood Cliffs, N.J: Prentice Hall; 1977.
41. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec 1991;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]


## Abbreviations

**API:** Application Programming Interface
**CDC:** Centers for Disease Control and Prevention
**GPS:** Global Positioning System
**HAC:** Hierarchical Agglomerative Clustering
**ID:** Identification
**IRB:** Institutional Review Board

*Edited by G Eysenbach; submitted 18.12.12; peer-reviewed by J Giménez, T Mackey; comments to author 18.01.13; revised version received 01.02.13; accepted 13.02.13; published 17.04.13.*

*Please cite as:*
*Hanson CL, Burton SH, Giraud-Carrier C, West JH, Barnes MD, Hansen B*
*Tweaking and Tweeting: Exploring Twitter for Nonmedical Use of a Psychostimulant Drug (Adderall) Among College Students*
*J Med Internet Res 2013;15(4):e62*
*URL: http://www.jmir.org/2013/4/e62/*
*doi:10.2196/jmir.2503*
*PMID:23594933*

Original Paper

# Tailored System to Deliver Behavioral Intervention and Manage Data in Randomized Trials

Hua Zheng[1], PhD; Milagros C Rosal[2], PhD; Carol A Oatis[3], PT, PhD; Wenjun Li[2], PhD; Patricia D Franklin[1], MD, MPH, MBA

[1]Department of Orthopedics and Physical Rehabilitation, University of Massachusetts Medical School, Worcester, MA, United States

[2]Division of Preventive and Behavioral Medicine, University of Massachusetts Medical School, Worcester, MA, United States

[3]Department of Physical Therapy, Arcadia University, Glenside, PA, United States

**Corresponding Author:**
Patricia D Franklin, MD, MPH, MBA
Department of Orthopedics and Physical Rehabilitation
University of Massachusetts Medical School
ACC Building, AC-7037
55 Lake Ave North
Worcester, MA, 01655
United States
Phone: 1 508 856 5748
Fax: 1 508 856 2581
Email: Patricia.Franklin@umassmed.edu

## Abstract

**Background:** The integrity of behavioral intervention trials depends on consistent intervention delivery, and uniform, comprehensive process data collection. It can be challenging in practice due to complex human interactions involved.

**Objective:** We sought to design a system to support the fidelity of intervention delivery and efficient capture of qualitative and quantitative process data for a telephone-delivered behavioral counseling intervention to increase physical activity and function after total knee replacement surgery.

**Methods:** A tailored system was designed to prompt the intervention coach in the delivery of a 5 step counseling protocol to support intervention fidelity across patients. System features included structured data components, automated data exchange functions, user-friendly data capture screens, and real-time surveillance reporting. The system structured the capture of patient goals and open-ended conversation.

**Results:** The system recorded intervention process data from each of 12 sessions held with the 92 intervention patients. During the trial, 992 telephone sessions were conducted, and more than 97% (4816/4960) of intervention process data fields were completed in the system. The coach spent 5-10 minutes preparing for each counseling call using system-generated summaries of historical data and 10-15 minutes entering intervention process data following each telephone session.

**Conclusions:** This intervention delivery system successfully supported the delivery of a structured behavioral counseling intervention and collection of intervention process data. It addressed the unique needs of clinical behavioral intervention trials, and had promising potential to facilitate high-fidelity translation of the intervention to broad clinical practice and Web-based multicenter clinical trials in the future.

***(J Med Internet Res 2013;15(4):e58)*** doi:10.2196/jmir.2375

**KEYWORDS**

structured behavioral intervention; intervention fidelity; clinical translational research; intervention delivery support system

## Introduction

Clinical settings are transitioning to computerized systems from traditional pen-and-paper methods to structure data collection and enhance data access and retrieval. Electronic data capture systems can increase data collection efficiency, improve recording reliability and accuracy, and reduce medical costs, thereby having potential to promote quality of care and evidence-based medicine [1-3]. These systems have been developed for, and applied to, a wide range of clinical fields to

enhance access to the information through Web communication [4-6].

Behavioral intervention studies require consistent intervention delivery and methods to capture process data. Process data is the sum of data collected to document the implementation of intervention and processes used in modeling the intervention process. With regard to behavioral interventions, telephone is a common delivery method that allows researchers to reach a larger population than face-to-face sessions while supporting one-to-one relationships and complete data capture. Telephone-based counseling interventions have been successfully used to promote behavior change, such as increased physical activity, weight management, and smoking cessation, and to improve palliative care in advanced cancer patients [7-11]. Of note, telephone is still the preferred technology for interventions within the elderly population, compared with Internet and computers [12,13]. Historically, telephone interventions were accomplished by training, ongoing supervision of the interventionist, and use of comprehensive data collection forms. This process was usually labor-intensive, time-consuming, and possibly data-redundant. Computer-assisted telephone interviewing (CATI) can deliver health behavior surveys and patient counseling [14,15] through pre-defined scripts, with branching logic to anticipate varied participant responses, or through pre-recorded telephone surveys. CATI can facilitate structured data collection and direct computer entry, but it is mainly used for collecting short, closed response data and prevents the interventionist from addressing open-ended topics [16]. Capturing intervention process data in practice is often challenging due to the complex human interactions involved (eg, a tailored counseling to help an individual set personal goals and strategies based on his specific motivations and problems). Therefore we aimed to design a "semi-structured" information system that *consistently* delivers a counseling intervention while simultaneously collecting *complete* process data; we achieved this by collecting both quantitative and qualitative study data on all participant interviews while supporting a range of open responses. Internet deployment of the system supports consistent interventions across geographic settings, and can facilitate multicenter behavioral intervention trials.

Recent telephone intervention studies have focused on the effectiveness of the systems, but very few have reported the design of systems that effectively capture telephone conversations. Most intervention data capture systems are viewed as electronic tools for data collection and storage, and do not consider efficient data recording and counseling fidelity. The latter is a particularly significant issue for behavioral interventions that are centered on participant adherence and adoption to specific behaviors [17]. The published literature lack empiric information to guide the design and development of this kind of system. Well-designed user-centered screens are needed that parallel the counseling interaction and structure to capture process data that will inform future counseling sessions. The system should be easy to learn, increase user productivity, decrease user errors, and not interfere with the counseling exchange [18]. User-centered intervention support systems have the potential to enhance the consistent delivery of research interventions and facilitate the translation of research interventions to clinical practice at study completion.

In this paper, we present the design and implementation of a system to support a behavioral counseling intervention and to collect process data to broadly support intervention delivery and evaluation. We present a general-purpose template for software to support behavioral interventions to facilitate translation of successful interventions to broad clinical settings and Web-based multicenter clinical trials.

## Methods

### Study Overview

The Joint Action Study was a National Institute of Health (NIH) funded randomized clinical trial (RCT) to test a self-care support intervention that aimed to increase activity and physical function among patients with advanced knee arthritis after total knee replacement (TKR) surgery. The surgeon-blinded RCT allocated a stratified random sample of TKR patients to one of two study conditions: behavioral intervention versus control. The control group received usual peri-operative care, including written educational information to guide rehabilitation exercises. Patients assigned to the intervention group received a newly designed telephone-delivered behavioral counseling intervention which was implemented by a trained coach and complemented with a website and written materials to distribute patient education components. The intervention used the 5A framework, namely discussion of the patient's *A*genda for the session, *A*ssessment of progress and goal achievement, delivery of brief personalized *A*dvice, *A*ssistance with goal setting, problem solving and action planning, and *A*rrangements for follow up. A total of 12 intervention sessions were delivered: 3 prior to surgery, 1 call the evening before surgery, a hospital visit at 2-3 days after surgery, and 7 additional sessions post-discharge between weeks 2 and 9 post-surgery over a period of 10 weeks. All participants completed assessments before TKR, and at 8 weeks, 6 months, and 12 months after TKR. The details on the theoretical rationale and design of this randomized clinical trial can be found in our published article [19].

To support the successful conduct and analysis of this RCT, the system design explicitly considered the continuum of study procedures, including patient enrollment, counseling intervention delivery, process data collection and management, statistical evaluation, and future translation to multisite clinical practice. Content experts, including the study behavioral psychologist, physical therapist, intervention coach, study recruiter, and biostatistician, were actively involved in the system design to assure that the system served each aspect of the study from patient recruitment and enrollment through data analytics, in an effort to enhance the functionality and efficiency of the system. Enrollment and follow up for control participants was also captured in this single system but is not the focus of this paper.

We developed this intervention support system using Microsoft Access 2007, and designed the functionality under the Visual Basic for Applications (VBA) development environment, to store and administer the collected intervention process data.

## Research Intervention and Process Data Collection

Each of the 12 counseling sessions had a dedicated screen to structure the call and present and capture relevant data. The data organization for each intervention session paralleled the 5 counseling components according to the 5As described above, and the components defined the data capture sequence during each session. At each of the study intervention calls, the coach systematically recorded process data (ie, call status, patients' short and long-term goals, challenges to activity, strengths and resources, and action plans at each session) in structured domains. These data were carried forward and displayed on screens for subsequent calls to assure prompt access for reference by the coach. The system prompted the coach to mail relevant intervention educational information based on the patient's stage in recovery and topics discussed. Finally, the system captured open-ended patient goals, barriers to physical activity adherence, and key patient-reported concerns or successes.

## Process Data Structure Design

### *Intervention Session Structure*

The intervention process data structure was designed to logically connect relational tracking data and transfer corresponding results across the telephone sessions over time. We organized each display screen to replicate the specific order of process data collection during the counseling session. Figure 1 summarizes the detailed structure of the intervention (orange boxes) as captured in the system. First, the intervention coach reviewed any items the patient would like to discuss during the call (patient's "Agenda"). In the "Assess" portion of the initial patient session, motivations, knowledge of the TKR procedure and subsequent recovery process, outcome expectations, and history of knee problems were captured. In the subsequent calls, the coach and patient assessed progress toward achieving post-operative rehabilitation goals established at the prior session; concrete accomplishments were documented. Next, the coach responded to patient's questions, to "Advise" on specific behaviors that supported progress toward functional goals. Free text was an option to document these details. Fourth, new goals were established, both short- and long-term, potential challenges identified, and problem-solving strategies were used, which culminated in an action plan. The coach guided the patient to obtain phase-specific information through an educational website specially designed for this study and other online sources, and selected tailored items from a "Toolkit" of print materials, that either addressed particular questions and concerns of the patient, or supported his or her goals, and mailed them to the patient. The system tracked which Toolkit materials were delivered to each patient at which session. Finally, the coach and the patient scheduled a date and time for the next intervention call that was documented in the "Arrange" section of the system. The documentation for each of the 5 components of the call was captured on a separate page that was navigated with the use of tabs at the top of the screen. Guided by the system, intervention fidelity across counselors was assured by consistent movement across the 5 counseling phases while allowing open-ended, relationship-based conversations.

Two additional tabs per counseling session included (1) call records, and (2) the coach's assessment of the call, which are illustrated in green boxes in Figure 1. Each call and its status (eg, no answer, busy signal, or leave message) were recorded. The pattern of successful calls suggested preferred days and times to reach each patient. After each call, the coach summarized the overall assessment and documented any comments and possible additional issues.

In addition, a "quick view" section displayed key historical variables from the first call and long-term goals for each patient and was available to the coach at every session for use in tailoring the intervention (highlighted in the blue box at the top of Figure 1).

### *Automated Process Data Exchange*

When preparing for a call, the intervention coach reviewed information from the previous session, such as the planned activity goals and reported challenges and successes from prior weeks. To avoid repeated "look-back" operations or duplicate manual typing work, the system functioned to automatically carry forward key topics from prior weeks to the current call screen.

New short-term goals set by the patient were recorded in the "Assist" screen for the current session. The goal-setting information from the previous session could also be displayed immediately on the screen when "same as last session" was checked. This function addressed the common situation in which the patient was still working on goals from the previous week. For each item included in short-term goals, the status "compared with last session" provided options to choose from new, same, advanced, simpler, attained, or stopped statuses. The status was carried-forward from the prior session (with the exception of attained or stopped) and the counselor was prompted to update the status at the current session. The information delivery process is demonstrated by dotted black arrows in Figure 1. In addition, the short-term goals set in the last session, including challenges and strategies, were automatically transmitted to the current "Assess" screen. Thus, the intervention coach could easily review the patient's progress toward his or her last goals during each call, and worked only within the current session screens (see red arrows in Figure 1). Similarly, the scheduled date and time for the next call populated the current "Arrange" to the next "Call Records" pages, shown using yellow arrows in Figure 1.

Initial and updated long-term goals were displayed in the "quick view" portion of the call screen for quick reference. If the long-term goal was changed, the latest content was transmitted to the "quick view" section for historical review. In this way, the most recent long-term goals were displayed in the current intervention screen using the live data delivery function, shown using blue arrows in Figure 1.

Two shared data sections were designed to dynamically display aggregated results: one for the successful call list and the other for previously distributed toolkit information. This function is shown in Figure 1 as the two areas filled with green and orange, respectively. In this way, the total number of successful calls

and mailed educational information, along with the mailed dates, were available for review by the coach at any call.

## User Interface Development

### *Overview*

User interface design addressed the sequence and content of the intervention calls, while employing user-friendly data capture tools. The interface development made use of the graphical user interface (GUI) components based on the application design pattern.

### *Process Data Entry Screens*

Recording of intervention process data is an important, but time consuming activity. Access to longitudinal process data by the intervention coach supports a longitudinal record of counseling activity, patient goals and motivations, and facilitates process evaluation of the intervention. Creating functional, user-friendly, high-fidelity data entry screens is critical for efficient collection of process data. The system's organization leads the coach systematically through the call while collecting information in pre-defined data fields.

We used a hierarchical structure to organize the intervention display screens, shown in Figure 2. The Intervention Participants List screen presented a list of all the intervention patients with their progress information (eg, next session number and scheduled call date) and informed the coach of the patient's intervention status. The Intervention Participant screen listed all 12 sessions and the associated information, such as status (eg, Need Contact, Completed, or Missed), call target date range, scheduled call date, and actual call date. Clicking on a specific session number navigated the coach to the Intervention Session screens. The tab user interface (UI) assured optimal use of space and made it easy to access the screens for each of the 5 counseling components. The information in the data entry interface was displayed in the same sequence that the coach used during the call. The high-fidelity data entry structure facilitated real-time recording.

The first intervention call was unique because the coach assessed and recorded the baseline patient knowledge about TKR, attitudes, and expectations. Information on these factors was important for the coach's reference during subsequent sessions. Thus, an area located on the left side of the data entry screens in sessions 2 to 12 carried the "quick view" data forward. See an example of intervention data entry screens illustrated in Figure 3.

## Surveillance and Analytics Reports

The system's basic reporting function produced interval reports at the individual patient level and in aggregate across all intervention patients. For individual patients, three types of reports were generated for the coach's review: key variables of the first session, goals across all sessions to date, and toolkit topics (advice) sent previously. In addition, monthly surveillance reports summarized all enrolled participants and described participant demographic attributes, adherence to intervention delivery intervals (eg, rates of completed and missed calls within target windows), and patient progress (eg, numbers of goals attained). All data collected in the system directly populated a database to support statistical analyses at the conclusion of the trial. While these reports were valuable to monitoring the multisite study, in clinical practice, these reports could be central to quality monitoring or documenting requirements for reimbursement.

**Figure 1.** 5A intervention session structure.

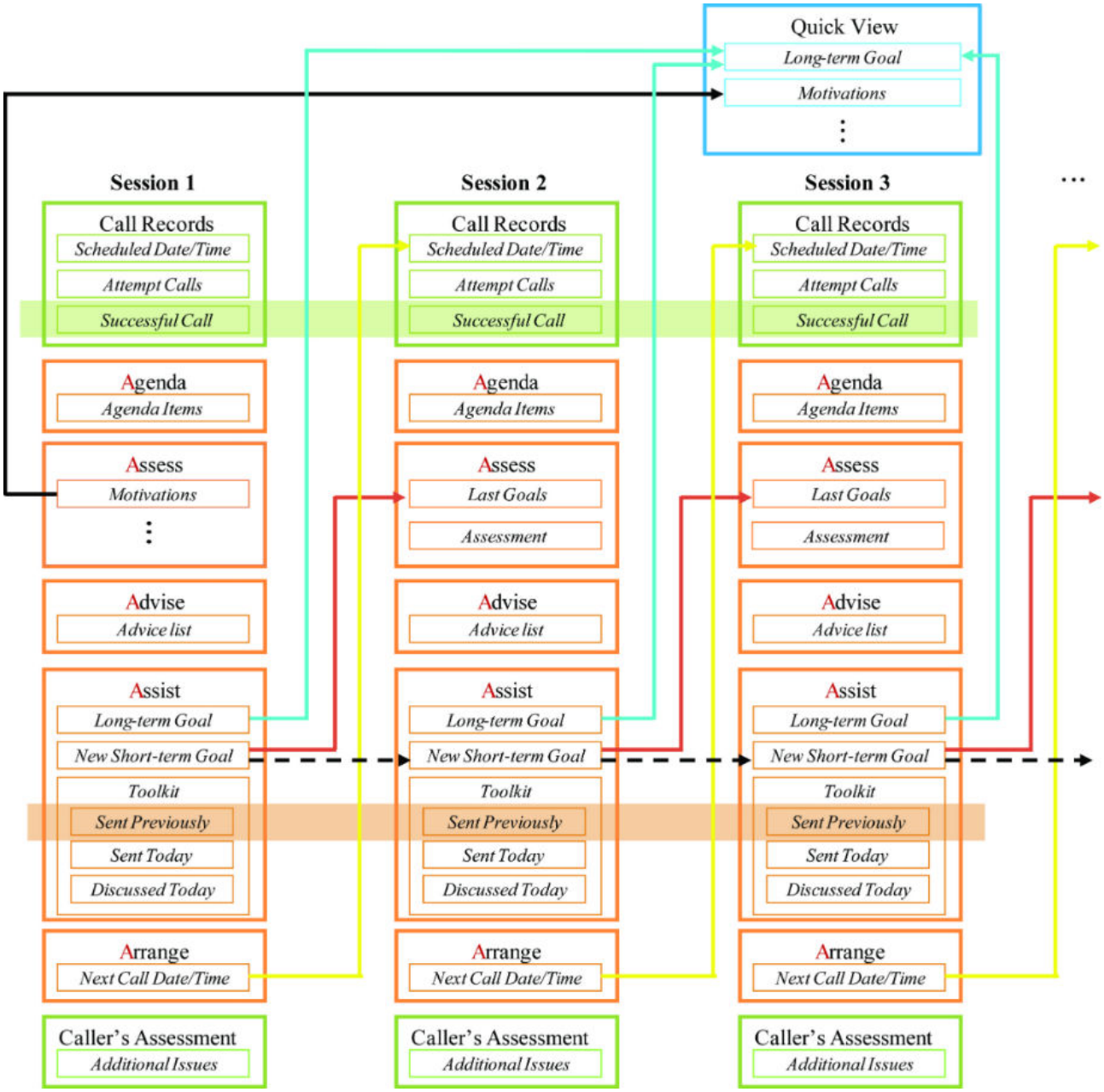



**Figure 2.** Hierarchical structure of intervention data entry.

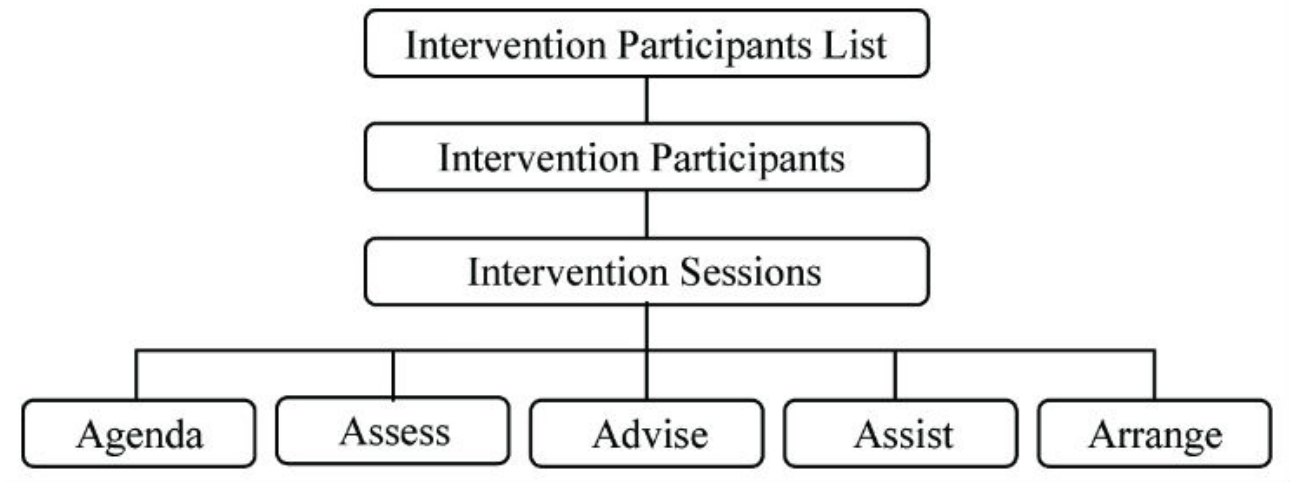
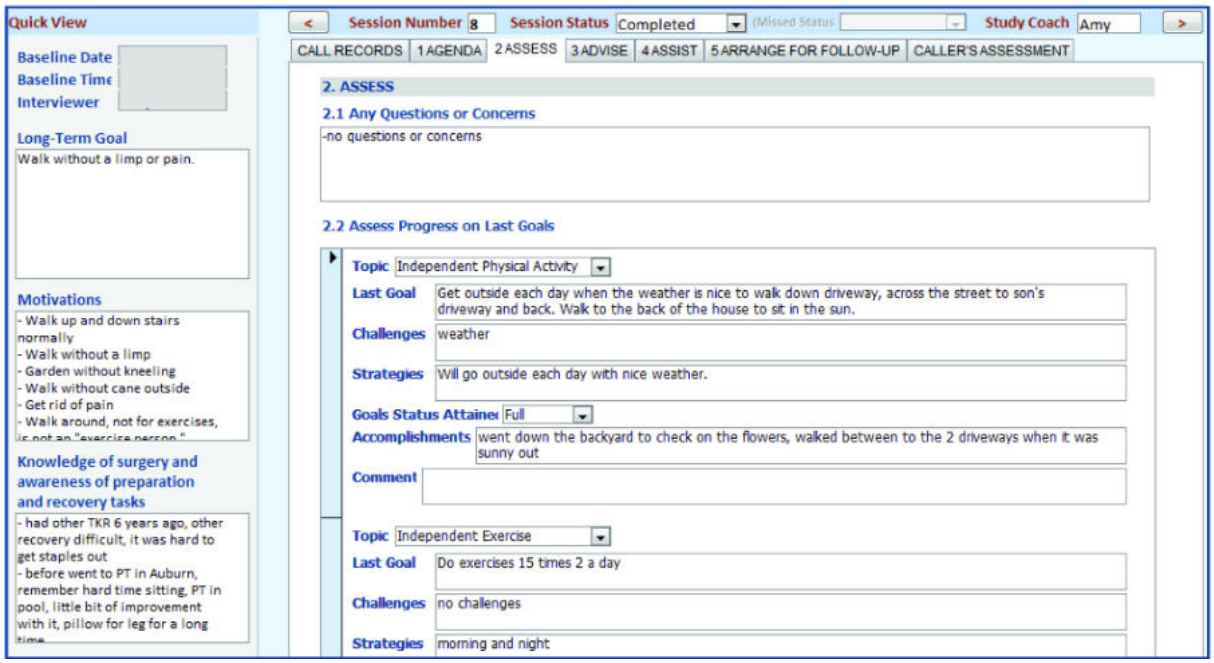


**Figure 3.** Example of intervention data entry screen.




## Results

During the period from September 2008 to March 2011, data were recorded in the system for 94 intervention patients; 92 completed the telephone sessions and the other 2 postponed their surgery. Table 1 shows the aggregated call information per session. 992 telephone sessions were conducted successfully with only 2.2% (24/1104) of calls unable to reach or refused. The average percentage of completed calls for Sessions 4-12 (post-surgery) exceeded 98% (818/828). The lower rate of call completion for Sessions 2 and 3 was related to short time frames between study enrollment and surgery date, and unrelated to the intervention and this system.

The intervention process data captured from the successful 992 calls were aggregated based on the 5As (Table 2). The call summary data for Agenda, Assess, and Assist sections were all captured without any missing information. The Assist portion included six summary categories for the session goal status: new (compared to last session), same, advanced, simpler, attained, and stopped. Goal-setting patterns could be observed in these data. For example, during the post-surgery phase, on average patients discussed three or four short-term goals during each call with the coach. These goals include one or two *new* goals, a goal that was the *same* as the week before, and one *attained* goal. The call arrangement for the next session was completed 86.4% (857/992) of the time. Missing data were due to patients who were not sure of a convenient time to have a phone call. In total, more than 97% (4816/4960) of intervention data fields were completed in the system.

A real-time surveillance report shows that the mean time spent for the first intake call was 47 (SD 15) minutes, and the mean time spent for subsequent sessions was 15 (SD 9) minutes.

With this structured system, the intervention coach spent 5-10 minutes to prepare for each counseling call and 10-15 minutes to complete data entry for each telephone session. At the completion of all 12 sessions, the coach spent 5-10 minutes for final data quality review. Total average time for system maintenance and tuning was less than 30 minutes per week. This intervention support system was well accepted by the intervention coach and other study staff.

**Table 1.** Intervention calls per session.

| | | Completed calls | | No calls due to Interval too short | | Missed calls due to Unable to reach | | Refused | |
|---|---|---|---|---|---|---|---|---|---|
| | | n | % | n | % | n | % | n | % |
| **Prior surgery** | | | | | | | | | |
| | Session 1 | 92 | 100.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 |
| | Session 2 | 32 | 34.8 | 58 | 63.0 | 1 | 1.1 | 1 | 1.1 |
| | Session 3 | 55 | 59.8 | 30 | 32.6 | 5 | 5.4 | 2 | 2.2 |
| **The evening prior surgery** | | | | | | | | | |
| | Session 4 | 85 | 92.4 | 0 | 0.0 | 6 | 6.5 | 1 | 1.1 |
| **Hospital visit** | | | | | | | | | |
| | Session 5 | 92 | 100.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 |
| **Post-surgery** | | | | | | | | | |
| | Session 6 | 92 | 100.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 |
| | Session 7 | 89 | 96.7 | 0 | 0.0 | 3 | 3.3 | 0 | 0.0 |
| | Session 8 | 91 | 98.9 | 0 | 0.0 | 1 | 1.1 | 0 | 0.0 |
| | Session 9 | 91 | 98.9 | 0 | 0.0 | 1 | 1.1 | 0 | 0.0 |
| | Session 10 | 92 | 100.0 | 0 | 0.0 | 0 | 0 | 0 | 0.0 |
| | Session 11 | 91 | 98.9 | 0 | 0.0 | 1 | 1.1 | 0 | 0.0 |
| | Session 12 | 90 | 97.8 | 0 | 0.0 | 1 | 1.1 | 1 | 1.1 |
| **Total** | | | | | | | | | |
| | | 992 | 89.9 | 88 | 8.0 | 19 | 1.7 | 5 | 0.5 |

**Table 2.** Intervention process data captured for 5As (N=992).

| Agenda | | Assess | | Advise | | Assist | | Arrange for follow-up | |
|---|---|---|---|---|---|---|---|---|---|
| n | % | n | % | n | % | n | % | n | % |
| 992 | 100.0 | 992 | 100.0 | 983 | 99.1 | 992 | 100.0 | 857 | 86.4 |

## Discussion

To date, electronic clinical trial systems have been used primarily for data entry and storage. We demonstrated that such systems could be expanded to support the *implementation* of behavioral interventions and the capture of intervention process data. Our system also facilitates translation of efficacious interventions to "real world" clinical settings as it structures the intervention for future coaches.

While telephone counseling is an effective approach to deliver behavioral change interventions, there is no consensus on how to capture themes of coach-patient telephone conversations. CATI is efficient and has become the data collection method of choice for an increasing number of research surveys. CATI permits direct data entry in an electronic format, reducing processing time and costs, and supports complicated question structure, allowing for better data quality. However, this technique is most appropriate for gathering relatively simple information requiring a short closed response, such as Yes/No or a simple quantitative response, and is not suitable for collecting open-ended qualitative data. We propose a compromise between open-ended sessions and CATI counseling mechanisms. While we structured and sequenced key intervention components, our system offers flexibility to the coach to tailor the delivery of such components and document the process. Therefore, all the counseling data are systematically collected and organized into a structured dataset supplemented by free text responses. The structured data summarize key process metrics, while the free text enhances qualitative evaluation.

This system was developed as a stand-alone application because it served a singe-site trial. It is housed on a network drive hosted by the hospital and research university campus. The coach can access the system anywhere by mapping the network drive when in the campus or by using the virtual private network (VPN) when off campus. In the second phase, we plan to use a Web-based intervention support system, which will serve multicenter clinical trials. The Web-based platform will ensure easy connection through the Internet and appropriate distribution of information on trial progress or protocols. The system guides structured intervention calls to assure that behavioral interventions and data collection are consistent across sites.

Trial interventionists may enter data to a centralized database via the Internet. Data monitoring, management, and transfer security are among the key considerations. The uniform system structure fosters easy data sharing in a standard format across sites, and reduces the burden of obtaining and standardizing a much broader set of data [20,21].

This system was also designed to overcome two key barriers to make clinical and translational research more efficient [22]—fragmented intervention data infrastructure and incompatibility between clinical practice and clinical research databases. To address the first issue, we considered the entire data flow and structure for intervention support and systematic data collection when designing the study system. The system was composed of separate but related data parts; each of the 5 intervention components was parallel in structure and communicated among each other and across sessions. Structural integrity of the system assures consistent intervention delivery and complete data collection. To address the second issue, the system design not only served the needs for this NIH-funded clinical trial, but also has the potential to foster translation from clinical research to clinical practice. For example, trained coaches who use an Internet-based version of the system in diverse clinical settings would be guided through each of the 12 intervention calls, the basic 5As within each call, and process data collection, thus allowing consistent intervention delivery across coaches and settings. If the intervention successfully promotes improved physical activity and function after TKR surgery, the intervention database will provide a template for software to distribute with other clinical settings through the Internet to replicate the intervention. It is also possible to apply this model to face-to-face counseling interventions or adopt it for other chronic diseases. It is not limited to a simple research tool.

One limitation of this system was the manual entry of patients' medical history information. Before intervention sessions, our research coordinator had to obtain patients' existing medical information from hospital databases, and manually enter these into the intervention database. In the future, we plan to automate the input of medication lists and co-morbid conditions as well as hospital operating room data from the electronic medical record to enhance research efficiency and reduce entry errors.

In conclusion, this innovative system to support intervention delivery and collection of intervention process data facilitated the successful delivery of a 12-session peri-operative behavioral intervention. It was also structured to support future translation to an Internet-based clinical practice tool, should the study outcome warrant broad implementation. The template and experience with this system design can inform additional behavioral interventions to support chronic disease management across diverse disease conditions.

## Acknowledgments

This research was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIH/NIAMS 1R01AR054479). The authors thank the study staff, including Betsy Costello, Amy Borg, and Wenyun Yang for their contribution to the database design, and Sylvie Puig for her editorial assistance.

## Conflicts of Interest

None declared.

## References

1. Buntin MB, Burke MF, Hoaglin MC, Blumenthal D. The benefits of health information technology: a review of the recent literature shows predominantly positive results. Health Aff (Millwood) 2011 Mar;30(3):464-471. [doi: 10.1377/hlthaff.2011.0178] [Medline: 21383365]
2. Kush RD, Helton E, Rockhold FW, Hardison CD. Electronic health records, medical research, and the Tower of Babel. N Engl J Med 2008 Apr 17;358(16):1738-1740. [doi: 10.1056/NEJMsb0800209] [Medline: 18420507]
3. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet 2012 Jun;13(6):395-405. [doi: 10.1038/nrg3208] [Medline: 22549152]
4. 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 Aug;11(3):e32 [FREE Full text] [doi: 10.2196/jmir.1221] [Medline: 19674959]
5. Roshanov PS, Misra S, Gerstein HC, Garg AX, Sebaldt RJ, Mackay JA, CCDSS Systematic Review Team. Computerized clinical decision support systems for chronic disease management: a decision-maker-researcher partnership systematic review. Implement Sci 2011 Aug;6:92 [FREE Full text] [doi: 10.1186/1748-5908-6-92] [Medline: 21824386]
6. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. JAMA 2008 Sep 10;300(10):1181-1196. [doi: 10.1001/jama.300.10.1181] [Medline: 18780847]
7. Goode AD, Reeves MM, Eakin EG. Telephone-delivered interventions for physical activity and dietary behavior change: an updated systematic review. Am J Prev Med 2012 Jan;42(1):81-88. [doi: 10.1016/j.amepre.2011.08.025] [Medline: 22176852]
8. Jepson RG, Harris FM, Platt S, Tannahill C. The effectiveness of interventions to change six health behaviours: a review of reviews. BMC Public Health 2010;10:538 [FREE Full text] [doi: 10.1186/1471-2458-10-538] [Medline: 20825660]

9. Sangster J, Furber S, Allman-Farinelli M, Haas M, Phongsavan P, Mark A, et al. A population-based lifestyle intervention to promote healthy weight and physical activity in people with cardiac disease: the PANACHE (Physical Activity, Nutrition And Cardiac HEalth) study protocol. BMC Cardiovasc Disord 2010 Apr;10:17 [FREE Full text] [doi: 10.1186/1471-2261-10-17] [Medline: 20374661]
10. Stead LF, Perera R, Lancaster T. Telephone counselling for smoking cessation. Cochrane Database Syst Rev 2006 Jul(3):CD002850. [doi: 10.1002/14651858.CD002850.pub2] [Medline: 16855992]
11. Bakitas M, Lyons KD, Hegel MT, Balan S, Brokaw FC, Seville J, et al. Effects of a palliative care intervention on clinical outcomes in patients with advanced cancer: the Project ENABLE II randomized controlled trial. JAMA 2009 Aug 19;302(7):741-749. [doi: 10.1001/jama.2009.1198] [Medline: 19690306]
12. Olson KE, O'Brien MA, Rogers WA, Charness N. Diffusion of technology: frequency of use for younger and older adults. Ageing Int 2011 Mar;36(1):123-145 [FREE Full text] [doi: 10.1007/s12126-010-9077-9] [Medline: 22685360]
13. Dickinson A, Hill RL. Keeping in touch: talking to older people about computers and communication. Educ Gerontol 2007 Aug;33(8):613-630. [doi: 10.1080/03601270701363877]
14. White MJ, Stark JR, Luckmann R, Rosal MC, Clemow L, Costanza ME. Implementing a computer-assisted telephone interview (CATI) system to increase colorectal cancer screening: a process evaluation. Patient Educ Couns 2006 Jun;61(3):419-428. [doi: 10.1016/j.pec.2005.05.008] [Medline: 15993558]
15. Dal Grande E, Fullerton S, Taylor AW. Reliability of self-reported health risk factors and chronic conditions questions collected using the telephone in South Australia, Australia. BMC Med Res Methodol 2012 Jul;12:108 [FREE Full text] [doi: 10.1186/1471-2288-12-108] [Medline: 22834889]
16. Curtin R, Presser S, Singer E. The effects of response rate changes on the index of consumer sentiment. Public Opin Q 2000;64(4):413-428. [Medline: 11171024]
17. Dane AV, Schneider BH. Program integrity in primary and early secondary prevention: are implementation effects out of control? Clin Psychol Rev 1998 Jan;18(1):23-45. [Medline: 9455622]
18. Johnson CM, Johnson TR, Zhang J. A user-centered framework for redesigning health care interfaces. J Biomed Inform 2005 Feb;38(1):75-87. [doi: 10.1016/j.jbi.2004.11.005] [Medline: 15694887]
19. Rosal MC, Ayers D, Li W, Oatis C, Borg A, Zheng H, et al. A randomized clinical trial of a peri-operative behavioral intervention to improve physical activity adherence and functional outcomes following total knee replacement. BMC Musculoskelet Disord 2011 Oct;12:226 [FREE Full text] [doi: 10.1186/1471-2474-12-226] [Medline: 21981909]
20. Thwin SS, Clough-Gorr KM, McCarty MC, Lash TL, Alford SH, Buist DS, et al. Automated inter-rater reliability assessment and electronic data collection in a multi-center breast cancer study. BMC Med Res Methodol 2007 Jun;7:23 [FREE Full text] [doi: 10.1186/1471-2288-7-23] [Medline: 17577410]
21. Kahn MG, Raebel MA, Glanz JM, Riedlinger K, Steiner JF. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. Med Care 2012 Jul;50 Suppl:S21-S29. [doi: 10.1097/MLR.0b013e318257dd67] [Medline: 22692254]
22. Sung NS, Crowley WFJ, Genel M, Salber P, Sandy L, Sherwood LM, et al. Central challenges facing the national clinical research enterprise. JAMA 2003 Mar 12;289(10):1278-1287. [Medline: 12633190]


## Abbreviations

**CATI:** computer-assisted telephone interviewing
**GUI:** graphical user interface
**NIH:** National Institute of Health
**RCT:** randomized clinical trial
**TKR:** total knee replacement
**UI:** user interface
**VBA:** Visual Basic for Applications
**VPN:** virtual private network


*Edited by G Eysenbach; submitted 03.10.12; peer-reviewed by LS Suggs, P Lam; comments to author 21.01.13; revised version received 06.02.13; accepted 09.02.13; published 11.04.13.*

*Please cite as:*
*Zheng H, Rosal MC, Oatis CA, Li W, Franklin PD*
*Tailored System to Deliver Behavioral Intervention and Manage Data in Randomized Trials*
*J Med Internet Res 2013;15(4):e58*
*URL: http://www.jmir.org/2013/4/e58/*
*doi:10.2196/jmir.2375*
*PMID:23579207*

Original Paper

# Clinical Outcome and Cost-Effectiveness of a Synchronous Telehealth Service for Seniors and Nonseniors with Cardiovascular Diseases: Quasi-Experimental Study

Ying-Hsien Chen[1], MD; Yen-Hung Lin[1], MD, PhD; Chi-Sheng Hung[1], MD; Ching-Chang Huang[1], MD; Deng-Feng Yeih[1], MD, PhD; Pao-Yu Chuang[2], Nursing Supervisor; Yi-Lwun Ho[1,3], MD, PhD; Ming-Fong Chen[1], MD, PhD

[1]National Taiwan University Hospital, Departments of Internal Medicine, Taipei, Taiwan

[2]National Taiwan University Hospital, Departments of Nursing, Taipei, Taiwan

[3]Graduate Institute of Clinical Medicine, National Taiwan University College of Medicine, Taipei, Taiwan

**Corresponding Author:**
Yi-Lwun Ho, MD, PhD
National Taiwan University Hospital
Departments of Internal Medicine
#7, Chung-Shan South Road
Taipei, 100
Taiwan
Phone: 886 2 2312 3456 ext 6552
Fax: 886 2 2353 9346
Email: ylho@ntu.edu.tw

## Abstract

**Background:** Telehealth based on advanced information technology is an emerging health care strategy for managing chronic diseases. However, the cost-effectiveness and clinical effect of synchronous telehealth services in older patients with cardiovascular diseases has not yet been studied. Since 2009, the Telehealth Center at the National Taiwan University Hospital has provided a range of telehealth services (led by a cardiologist and staffed by cardiovascular nursing specialists) for cardiovascular disease patients including (1) instant transmission of blood pressure, pulse rate, electrocardiography, oximetry, and glucometry for analysis, (2) mutual telephone communication and health promotion, and (3) continuous analytical and decision-making support.

**Objective:** To evaluate the impact of a synchronous telehealth service on older patients with cardiovascular diseases.

**Methods:** Between November 2009 and April 2010, patients with cardiovascular disease who received telehealth services at the National Taiwan University Hospital were recruited. We collected data on hospital visits and health expenditures for the 6-month period before and the 6-month period after the opening of the Telehealth Center to assess the clinical impact and cost-effectiveness of telehealth services on cardiovascular patients.

**Results:** A total of 141 consecutive cardiovascular disease patients were recruited, including 93 aged ≥65 years (senior group) and 48 aged <65 years (nonsenior group). The telehealth intervention significantly reduced the all-cause admission rate per month per person in the nonsenior group (pretelehealth: median 0.09, IQR 0-0.14; posttelehealth: median 0, IQR 0-0; $P$=.002) and the duration (days per month per person) of all-cause hospital stay (pretelehealth: median 0.70, IQR 0-1.96; posttelehealth: median 0, IQR 0-0; $P$<.001) with increased all-cause outpatient visits per month per person (pretelehealth: median 0.77, IQR 0.20-1.64; posttelehealth: mean 1.60, IQR 1.06-2.57; $P$=.002). In the senior group, the telehealth intervention also significantly reduced the all-cause admission rate per month per person (pretelehealth: median 0.10, IQR 0-0.18; posttelehealth: median 0, IQR 0-0; $P$<.001) and the duration (days per month per person) of all-cause hospital stay (pretelehealth: median 0.59, IQR 0-2.24; posttelehealth: median 0, IQR 0-0; $P$<.001) with increased all-cause outpatient visits per month per person (pretelehealth: median 1.40, IQR 0.52-2.63; posttelehealth: median 1.76, IQR 1.12-2.75; $P$=.02). In addition, telehealth intervention reduced the inpatient cost in the nonsenior group from $814.93 (SD 1000.40) to US $217.39 (SD 771.01, $P$=.001) and the total cost per month from US $954.78 (SD 998.70) to US $485.06 (SD 952.47, $P$<.001). In the senior group, the inpatient cost per month was reduced from US $768.27 (SD 1148.20) to US $301.14 (SD 926.92, $P$<.001) and the total cost per month from US $928.20 (SD 1194.11) to US $494.87 (SD 1047.08, $P$<.001).

**Conclusions:** Synchronous telehealth intervention may reduce costs, decrease all-cause admission rates, and decrease durations of all-cause hospital stays in cardiovascular disease patients, regardless of age.

*(J Med Internet Res 2013;15(4):e87)* doi:10.2196/jmir.2091

**KEYWORDS**

age factors; telehealth; cost-benefit analysis; cardiovascular diseases

## Introduction

Elderly individuals are often frail and at high risk for requiring health care. According to Medicare data, 82% of elderly individuals living in the United States have chronic conditions that account for an annual average medical expenditure of US $5015 per person during 1999 [1]. From research based on a representative sample of seniors aged over 65 years nationwide in Taiwan (N=114,873), the prevalence of chronic disease for seniors is 70.4%, and medical expenditures for seniors with chronic conditions accounts for 92.7% of the total medical expenditures for seniors. Among these expenditures, cardiovascular diseases account for an annual average medical expenditure as high as US $4291 per person, which ranks it as the most costly of the chronic diseases in Taiwan [2]. Those patients with multiple comorbidities requiring continued medical care tend to be older and they have often been denied access to hospitals or have faced unequal access to health care, referral, treatment, or other social security benefits in developed countries [3,4].

Long-term chronic health conditions are one of the greatest challenges to worldwide health care systems [5]. Nurse-led case management for disease control and prevention has been successful in diabetes [6], chronic obstructive pulmonary disease [7], coronary heart disease [8], and heart failure [9]. This is particularly true for one of the most deadly diseases in cardiology, chronic heart failure, in which effective management is a major problem because of its high prevalence, high treatment costs, mortality, and morbidity [10]. In addition to heart failure, cardiovascular diseases also include diverse diseases, such as coronary artery disease (CAD), arrhythmia, cardiomyopathy, congenital heart disease, aortic disease, and valvular heart disease. To date, comprehensive home-based heart failure management care systems are thought to be effective in reducing costs and the reliance on prescription drugs, improving clinical outcomes and functional levels, bolstering patient knowledge and self-care, prolonging total survival, enhancing drug adherence, and improving quality of life in the United States, Canada, Europe, and Australia [9,11-13]. Home-based interventions have been shown to improve the clinical outcome and cost-effectiveness of treatment throughout the world despite differences in ethnicity, costs of treatment, differing medical practices, and health care systems [14,15]. The weight of existing evidence led the American College of Cardiology (ACC) Foundation and the American Heart Association (AHA) to recommend multidisciplinary disease management programs, including telemonitoring for patients at high risk for heart failure as a class I indication, with a class A level of evidence [16]. However, 2 prospective randomized multicenter clinical trials, the Telemonitoring to Improve Heart Failure Outcomes (Tele-HF) study in 2010 and the Telemedical Interventional Monitoring in Heart Failure (TIM-HF) study in 2011, used telehealth monitoring for heart failure patients, but did not demonstrate a benefit on mortality rates through telemonitoring [17,18]. The lack of measured benefits was explained as being a product of several factors: inadequate health care professional staffing, information overload, and noncompliance to telemonitoring intervention (in the Tele-HF study). In addition, the TIM-HF trial was unable to detect clinically relevant differences in mortality because of low statistical power.

Advances in modern telecommunication technologies, such as Internet-based rather than telephone-based data exchange platforms, have provided telehealth medicine a new approach for chronic disease management [19]. Synchronous telemedicine involves real-time processing of patient data obtained remotely; therefore, the patient and care provider need to be simultaneously available [20]. However, the effect of synchronous telehealth medicine based on advanced information technology in elderly patients with cardiovascular disease has never been studied. The Telehealth Center at the National Taiwan University Hospital (NTUH) began physician-led synchronous telehealth services in 2009. In this paper, we assess the clinical impact and cost-benefits achieved using telehealth services for patients with cardiovascular diseases.

## Methods

### Patients

The study recruitment strategy followed a nonconcurrent prospective design, approved by the Institutional Review Board at NTUH (Taipei, Taiwan), and conducted by the Taiwan ELEctroHEALTH study group (TELEHEALTH study group). Informed consent was obtained from all participating patients. The TELEHEALTH study group was located in the Telehealth Center of NTUH and launched telehealth services for cardiovascular disease patients in 2009. In order to include those patients with diagnosed cardiovascular disease as well as patients at high risk for cardiovascular disease, patients who had the following disease conditions were considered for enrollment in the telehealth service: (1) CAD with or without percutaneous coronary intervention, (2) myocardial infarction, (3) congestive heart failure, (4) arrhythmia, including bradyarrhythmia, tachyarrhythmia, or ventricular arrhythmia with an implantable cardiac defibrillator implant, (5) diabetes mellitus, (6) syncope, (7) ≥2 CAD risk factors with angina, and (8) other surgical or congenital heart conditions. Syncope, which may be noncardiogenic, was also used as an indication because biometric information obtained through telemonitoring can be of value in cases that are caused from orthostatic hypotension or vasovagal reflex. The main exclusion criterion for the telehealth service was patient refusal. Between November 2009 and May 2010, consecutive cardiovascular disease patients were evaluated for eligibility for telehealth monitoring during inpatient and outpatient treatment. After obtaining informed consent, the patient and main caregiver completed a face-to-face tutorial for operating the manometer, oximeter, glucometer, and electrocardiography device. Internet access and biometrics

transmission were confirmed before launching the telehealth service at home. Demographic and clinical data were collected from all participants, including gender, age, occupation, marital status, years of schooling, number of people in the household, main caregiver, clinical comorbidities, biochemistry findings, and echocardiography study results. The estimated glomerular filtration rate (eGFR) was calculated using the simplified Modification of Diet in Renal Disease (MDRD) formula. The clinical expenditure data, including outpatient costs, inpatient costs, emergency department costs, and total cost of all-cause health care and the frequency of health care visits (including emergency department visits, inpatient admissions, inpatient and outpatient procedures, and outpatient physician visits), were collected 6 months before and 6 months after the initiation of the telehealth service.

### Telehealth Service

There were 3 components to the telehealth services provided. First, there were real-time transmission of biometrics from the patients to the health care team, including blood pressure, pulse rate, electrocardiography, oximetry, and glucometry. Glucometry was performed in patients with diabetes mellitus and those with impaired fasting glucose and impaired glucose tolerance. Nondiabetic patients did not have glucometry as a telemonitoring module. Each transmitted clinical biometric was recorded in health record clouds that were under synchronous surveillance by the Telehealth Center at NTUH. Second, there were telephone exchanges between the health care team and patients for communication and health promotion. Third, full-time case managers and cardiologists were in charge of care 24 hours a day. The telehealth services included health education, diet therapy, fluid status evaluation, drug adverse effects evaluation, drug compliance monitoring, mood or emotion care, and patient surveillance through the advanced information technology-based monitor system. Nurse case managers tracked and scrutinized the clinical information carefully and contacted the patient or relatives at least once per day (except in cases in which the patients indicated a preference against daily interactions) or when abnormal data were transmitted back to the service station at the Telehealth Center at NTUH. Patients who were under surveillance by the telehealth service could also contact the service station by telephone when subjective or objective abnormalities were a concern. The clinical information was relayed to the cardiology specialist; the cardiology specialists made the final judgment and suggestions regarding care. The Telehealth Center also provided constant analytical and decision-making support. The management of these heart failure patients was conducted according to the ACC/AHA guidelines for heart failure management [21,22].

### Study Follow-Up and Endpoint

The clinical impact analysis was based on data from outpatient and emergency departments, general wards, and intensive care units. The admission rate, total hospital stay, and hospital costs (including the emergency department, inpatient, and outpatient costs) were collected before and after initiation of the telehealth service. The data were expressed as a monthly average (US $1 equivalent to $31.40 New Taiwan dollars based on the exchange rate on January 14, 2010).

### Statistical Analysis

Enrolled participants were divided according to age into 2 groups: the seniors group included patients aged 65 years and older and the nonseniors group included patients younger than 65 years. Because heart failure is the most-studied cardiovascular disease in the literature and a higher proportion of seniors had heart failure in our study, we further stratified all the participants according to the disease state of heart failure for analysis. Continuous data are presented as mean and standard deviation (SD) and the skewed continuous data are presented as medians and interquartile ranges (IQRs). Discrete data are given as counts (n) and percentages. The Student's *t* test was applied to compare continuous unpaired data between seniors and nonseniors; for skewed continuous data, the Mann-Whitney test was used. The chi-square test was used to compare categorical data between seniors and nonseniors. The Shapiro-Francia W' test for normality revealed that the changes in clinical outcome were not normally distributed. Therefore, after stratifying the patients according to age and the state of heart failure, the pretelehealth and posttelehealth clinical variables as well as the paired continuous data were compared with the Wilcoxon signed-rank test. A $P$ value <.05 was considered statistically significant. Stata/SE 11.0 for Windows (StataCorp LP, College Station, TX, USA) was used for statistical analyses.

## Results

Between November 2009 and May 2010, a total of 141 consecutive cardiovascular disease patients who received telehealth service from the Telehealth Center at NTUH were included in this study. Of these 141 patients, the mean age was 68.5 years (SD 14.4), and the mean duration of telehealth service was 52.7 days (SD 38.2). Among the patients, 56.7% of had preexisting CAD, 18.4% had a myocardial infarction, 39.7% had heart failure, and 44.0% had an arrhythmia. The baseline characteristics are presented in Table 1.

**Table 1.** Baseline demographic data for patients with cardiovascular disease receiving telehealth service.

| Demographic data | Nonsenior (<65 years) n=48 | Senior (≥65 years) n=93 | Total N=141 | *P* value |
|---|---|---|---|---|
| Age (years), median (IQR) | 58.0 (49.1-61.0) | 76.3 (71.1-80.9) | 70.8 (60.8-78.3) | <.001 |
| Female, n (%) | 12 (25.0) | 43 (46.2) | 55 (39.0) | .01 |
| Body weight (kg), mean (SD) | 65.3 (15.3) | 60.2 (11.6) | 61.8 (13.1) | .06 |
| Body height (cm), mean (SD) | 165.0 (7.3) | 160.0 (8.6) | 161.6 (8.5) | .005 |
| BMI($kg/m^2$), mean (SD) | 23.7 (4.6) | 23.6 (3.9) | 23.6 (4.1) | .84 |
| Years of schooling, mean (SD) | 13.1 (3.9) | 10.4 (4.7) | 11.3 (4.6) | .008 |
| Number in household, mean (SD) | 2.0 (0.9) | 2.2 (1.4) | 2.1 (1.3) | .31 |
| Married, n (%) | 41 (85.4) | 84 (90.3) | 125 (88.7) | .39 |
| Telehealth service duration (days), mean (SD) | 55.8 (41.8) | 51.1 (36.3) | 52.7 (38.2) | .49 |
| **Enrollment criteria, n (%)** | | | | |
| Coronary artery disease (CAD) | 20 (41.7) | 44 (47.3) | 64 (45.4) | .52 |
| Myocardial infarction | 1 (2.1) | 5 (5.4) | 6 (4.3) | .36 |
| Heart failure | 8 (16.7) | 12 (12.9) | 20 (14.2) | .54 |
| Arrhythmia | 10 (20.8) | 19 (20.4) | 29 (20.6) | .96 |
| Diabetes mellitus | 2 (4.2) | 3 (3.2) | 5 (3.6) | .78 |
| Syncope | 0 | 0 | 0 | — |
| ≥2 CAD risk factors with angina | 2 (4.2) | 2 (2.2) | 4 (2.8) | .49 |
| Surgical or congenital heart diseases | 5 (10.4) | 8 (8.6) | 13 (9.2) | .72 |
| **Comorbidity, n (%)** [a] | | | | |
| Hypertension | 27 (56.3) | 82 (88.2) | 109 (77.3) | <.001 |
| Diabetes mellitus | 14 (29.2) | 39 (41.9) | 53 (37.6) | .14 |
| Dyslipidemia | 23 (47.9) | 36 (38.7) | 59 (41.8) | .29 |
| CAD | 25 (52.1) | 55 (59.1) | 80 (56.7) | .42 |
| Myocardial infarction | 7 (14.6) | 19 (20.4) | 26 (18.4) | .72 |
| Heart failure | 12 (25.0) | 42 (45.2) | 56 (39.7) | .02 |
| Valvular heart disease | 4 (8.3) | 15 (16.1) | 19 (13.5) | .20 |
| Chronic renal insufficiency | 7 (14.6) | 34 (36.6) | 41 (29.1) | .006 |
| Stroke | 5 (10.4) | 22 (23.7) | 27 (19.1) | .06 |
| Arrhythmia | 17 (35.4) | 45 (48.4) | 62 (44.0) | .14 |
| Atrial fibrillation | 8 (16.7) | 25 (26.9) | 33 (23.4) | .18 |
| VT or VF | 2 (4.2) | 2 (2.2) | 4 (2.8) | .49 |
| Bradyarrhythmia | 3 (6.3) | 14 (15.1) | 17 (12.1) | .12 |
| Pacemaker | 6 (12.5) | 11 (11.8) | 17 (12.1) | .91 |
| **Laboratory results, mean (SD)** [b] | | | | |
| LVEF (%) | 61.4 (15.5) | 58.5 (14.1) | 59.4 (14.5) | .18 |
| Hemoglobin (g/dL) | 13.3 (2.8) | 12.4 (1.8) | 12.7 (2.2) | .04 |
| Creatinine (mg/dL) | 2.0 (2.5) | 1.9 (1.7) | 1.9 (2.0) | .79 |
| Fasting glucose (mg/dL) | 113.7 (44.9) | 111.4 (34.3) | 112.2 (38.1) | .76 |
| Total cholesterol (mg/dL) | 186.7 (52.3) | 183.9 (44.0) | 184.9 (46.7) | .75 |
| HDL (mg/dL) | 41.0 (11.6) | 41.7 (13.8) | 41.4 (13.1) | .82 |

| Demographic data | Nonsenior (<65 years) n=48 | Senior (≥65 years) n=93 | Total N=141 | *P* value |
|---|---|---|---|---|
| Triglycerides (mg/dL) | 142.4 (130.9) | 134.6 (87.8) | 137.2 (103.9) | .72 |
| eGFR (mL/min) | 64.2 (26.4) | 49.7 (22.6) | 54.3 (24.7) | .001 |
| **Medical prescriptions, n (%)** [c] | | | | |
| Various HTN drugs, mean (SD) | 1.6 (1.2) | 2.0 (1.0) | 19 (1.0) | .06 |
| Diuretics | 14 (29.2) | 50 (53.8) | 64 (45.4) | .005 |
| Spironolactone | 4 (8.3) | 11 (11.8) | 15 (10.6) | .52 |
| Calcium channel blocker | 17 (45.4) | 40 (43.0) | 57 (40.4) | .38 |
| Antiplatelet | 24 (50.0) | 54 (58.1) | 78 (55.3) | .36 |
| ACE inhibitor or ARB | 17 (35.4) | 56 (60.2) | 73 (51.8) | .005 |
| Beta-blockers | 25 (52.1) | 31 (33.3) | 56 (39.7) | .03 |
| Statin or fibrates | 15 (31.3) | 24 (25.8) | 39 (27.7) | .49 |
| Antiarrhythmic | 15 (31.3) | 29 (31.2) | 44 (31.2) | .99 |
| OHA or insulin | 5 (10.4) | 24 (25.8) | 29 (20.6) | .03 |

[a] VT: ventricular tachycardia; VF: ventricular fibrillation.

[b] LVEF: left ventricle ejection fraction; HDL: high-density lipoprotein; eGFR: estimated glomerular filtration rate, calculated using simplified Modification of Diet in Renal Disease (MDRD) formula.

[c] HTN: hypertension; ACE: angiotensin-converting enzyme; ARB: angiotensin II receptor blocker; OHA: oral hypoglycemic agent.

Patients were further stratified according to age. Seniors (aged ≥65 years) comprised 66.0% (93/141) of the total enrollment with median age 76.3 years (IQR 71.1-80.9 years), and nonseniors (aged <65 years) accounted for 34.0% (48/141) of the total enrollment with median age 58.0 years (IQR 49.1-61.0 years). The indication for enrollment between the senior and nonsenior groups was comparable. Seniors had a significantly higher proportion of females (46.2% vs 25.0%, *P*=.01), fewer years of schooling (mean 10.4, SD 4.7 vs mean 13.1, SD 3.9, *P*=.008), hypertension (88.2% vs 56.3%, *P*<.001), heart failure (45% vs 25%, *P*=.02), and chronic renal insufficiency (36.6% vs 14.6%, *P*=.006). In addition, seniors had a significantly lower level of hemoglobin (mean 12.4 g/dL, SD 1.8 vs mean 13.3 g/dL, SD 2.8, *P*=.04) and eGFR (mean 49.7 mL/min, SD 22.6 vs mean 64.2 mL/min, SD 26.4, *P*=.001). The number of hypertension drugs was comparable between senior and nonseniors; however, compared to the nonseniors, seniors used a significantly higher proportion of diuretics (53.8% vs 29.2%, *P*=.005) and angiotensin-converting enzyme inhibitors or angiotensin II receptor blockers (60.2% vs 35.4%, *P*=.005).

After receiving telehealth services, patients with cardiovascular diseases had a significantly decreased all-cause admission per month rate (pretelehealth: median 0.10, IQR 0-0.17; posttelehealth: median 0, IQR 0-0; *P*<.001), and decreased average days/month duration of all-cause hospital stay (pretelehealth: median 0.60, IQR 0-2.20; posttelehealth: median 0, IQR 0-0; *P*<.001), and an increased all-cause outpatient visits per month rate (pretelehealth: median 1.17, IQR 0.36-2.27; posttelehealth: median 1.70, IQR 1.15-2.72; *P*<.001). There was no significant difference in the all-cause emergency department visit rate.

When stratified according to age, in nonseniors the average all-cause admission rate per month was significantly decreased (pretelehealth: median 0.09, IQR 0-0.14; posttelehealth: median 0, IQR 0-0; *P*=.002) and the all-cause hospital stay days per month duration was significantly decreased (pretelehealth: median 0.70, IQR 0-1.96; posttelehealth: median 0, IQR 0-0; *P*<.001), whereas the all-cause outpatient visit rate per month was significantly increased (pretelehealth: median 0.77, IQR 0.20-1.64; posttelehealth: median 1.60, IQR 1.06-2.57; *P*=.002). In seniors, the average all-cause admission rate per month was significantly decreased (pretelehealth: median 0.10, IQR 0-0.18; posttelehealth: median 0, IQR 0-0; *P*<.001) and the all-cause hospital stay days per month duration was significantly decreased (pretelehealth: median 0.59, IQR 0-2.24; posttelehealth: median 0, IQR 0-0; *P*<.001), whereas the all-cause outpatient visit rate per month was significantly increased (pretelehealth: median 1.40, IQR 0.52-2.63; posttelehealth: median 1.76, IQR 1.12-2.75; *P*=.02). There was no significant difference in all-cause emergency department visit rates in both seniors and nonseniors (Table 2).

**Table 2.** Differences in admission rates and duration of hospital stays 6 months before (pre) and 6 months after (post) initiation of telehealth services in patients with cardiovascular diseases, stratified according to age.

| Final measure | Age group, median (IQR) | | | | | | Total, median (IQR) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Nonseniors (<65 years) | | | Seniors (≥65 years) | | | | | |
| | n=48 | | | n=93 | | | N=141 | | |
| | Pre | Post | *P value* | Pre | Post | *P value* | Pre | Post | *P value* |
| All-cause admission rate[a] | 0.09 (0-0.14) | 0 (0-0) | .002 | 0.10 (0-0.18) | 0 (0-0) | <.001 | 0.10 (0-0.17) | 0 (0-0) | <.001 |
| Duration of all-cause hospital stay[b] | 0.70 (0-1.96) | 0 (0-0) | <.001 | 0.59 (0-2.24) | 0 (0-0) | <.001 | 0.60 (0-2.2) | 0 (0-0) | <.001 |
| All-cause outpatient visits[a] | 0.77 (0.20-1.64) | 1.60 (1.06-2.57) | .002 | 1.40 (0.53-2.63) | 1.76 (1.12-2.75) | .02 | 1.17 (0.36-2.27) | 1.70 (1.15-2.72) | <.001 |
| All-cause emergency department visits[a] | 0 (0-0.10) | 0 (0-0.20) | .10 | 0 (0-0.14) | 0 (0-0) | .24 | 0 (0-0.13) | 0 (0-0) | .06 |

[a] Visits per month per person.

[b] Day(s) per month per person.

When stratified according to heart failure status, the average all-cause admission rate per month of the non–heart failures was significantly decreased (pretelehealth: median 0.09, IQR 0-0.12; posttelehealth: median 0, IQR 0-0; *P*=.003) and the all-cause hospital stay days per month duration was significantly decreased (pretelehealth: median 0.18, IQR 0-1.03; posttelehealth: median 0, IQR 0-0; *P*<.001), whereas the all-cause outpatient visit rate per month was significantly increased (pretelehealth: median 0.99, IQR 0.33-1.96; posttelehealth: mean 1.48, IQR 0.90-2.40; *P*=.009). The average all-cause admission rate per month of the heart failures was significantly decreased (pretelehealth: median 0.14, IQR 0.10-0.28; posttelehealth: median 0, IQR 0-0; *P*=.005) and the all-cause hospital stay days per month duration was significantly decreased (pretelehealth: median 2.25, IQR 0.97-4.49; posttelehealth: median 0, IQR 0-0; *P*<.001), whereas the all-cause outpatient visit rate per month was significantly increased (pretelehealth: median 1.43, IQR 0.52-2.40; posttelehealth: median 2.20, IQR 1.36-3.53; *P*=.009). There was no significant difference in all-cause emergency department visit rates for heart failure patients and non–heart failure patients (Table 3).

After initiation of the telehealth service, the expenditure during inpatient care decreased per month from US $784.15 (SD 1096.74) to US $272.63 (SD 875.08, *P*<.001), and the expenditure during emergency department care increased per month from US $19.09 (SD 32.82) to US $28.14 (SD 99.82, *P*=.007). Although, the expenditures for outpatient care increased per month from US $134.00 (SD 272.71) to US $190.76 (SD 376.98, *P*=.007), the total cost of all-cause health care decreased per month from US $937.25 (SD 1127.84) to US $491.52 (SD 1012.45, *P*<.001). This pattern of impact remained when patients were stratified according to either age (Table 4) or heart failure status (Table 5).

**Table 3.** Differences in admission rates and duration of hospital stays 6 months before (pre) and 6 months after (post) initiation of telehealth services in patients with cardiovascular diseases, stratified according to heart failure status.

| Final measure | Non–heart failure, median (IQR) | | | Heart failure, median (IQR) | | |
|---|---|---|---|---|---|---|
| | n=87 | | | n=54 | | |
| | Pre | Post | *P* value | Pre | Post | *P* value |
| All-cause admission rate[a] | 0.09 (0-0.12) | 0 (0-0) | .003 | 0.14 (0.10-0.28) | 0 (0-0) | .005 |
| Duration of all-cause hospital stay[b] | 0.18 (0-1.03) | 0 (0-0) | <.001 | 2.25 (0.97-4.49) | 0 (0-0) | <.001 |
| All-cause outpatient visits[a] | 0.99 (0.33-1.94) | 1.48 (0.90-2.40) | .009 | 1.43 (0.52-3.40) | 2.20 (1.36-3.53) | .009 |
| All-cause emergency department visits[a] | 0 (0-0.93) | 0 (0-0) | .13 | 0.10 (0-0.18) | 0 (0-0.23) | .38 |

[a] Time per month per person.

[b] Day(s) per month per person.

**Table 4.** Monthly average cost per patient (in US$) 6 months before (pre) and 6 months after (post) initiation of telehealth services in patients with cardiovascular diseases, stratified according to age.

| Final measure | Nonsenior, mean (SD) n=48 | | | Senior, mean (SD) n=93 | | | Total participants, mean (SD) N=141 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Pre | Post | *P value* | Pre | Post | *P value* | Pre | Post | *P value* |
| Outpatient cost | 127.08 (309.34) | 263.51 (569.44) | .04 | 137.57 (253.47) | 153.21 (215.45) | .08 | 134.00 (272.71) | 190.76 (376.98) | .007 |
| Inpatient cost | 814.93 (1000.40) | 217.39 (771.01) | .001 | 768.27 (1148.20) | 301.14 (926.92) | <.001 | 784.15 (1096.74) | 272.63 (875.08) | <.001 |
| Emergency department cost | 12.76 (26.89) | 4.16 (12.76) | .01 | 22.35 (35.18) | 40.51 (120.93) | .11 | 19.09 (32.82) | 28.14 (99.82) | .007 |
| Total cost of all-cause health care | 954.78 (998.70) | 485.06 (952.47) | <.001 | 928.20 (1194.11) | 494.87 (1047.08) | <.001 | 937.25 (1127.84) | 491.52 (1012.45) | <.001 |

**Table 5.** Monthly average cost per patient (in US$) 6 months before (pre) and 6 months after (post) initiation of telehealth services in patients with cardiovascular diseases, stratified according to heart failure status.

| Final measure | Non–heart failure, mean (SD) n=87 | | | Heart failure, mean (SD) n=54 | | |
|---|---|---|---|---|---|---|
| | Pre | Post | *P* value | Pre | Post | *P* value |
| Outpatient cost | 95.47 (213.71) | 114.49 (174.57) | .05 | 196.09 (340.55) | 313.64 (548.64) | .08 |
| Inpatient cost | 496.94 (748.84) | 195.92 (668.84) | <.001 | 1246.89 (1383.40) | 396.22 (1127.68) | <.001 |
| Emergency department cost | 12.99 (32.03) | 9.52 (32.42) | .06 | 28.92 (31.94) | 58.14 (152.08) | .15 |
| Total cost of all-cause health care | 605.39 (777.18) | 319.93 (728.71) | <.001 | 1471.89 (1381.29) | 768.00 (1311.42) | <.001 |

# Discussion

## Principal Results

In our clinical investigation, a real-time telehealth service based on the combination of information technology and rapid-response hospital services improved clinical outcomes, decreased admission rates, and shortened the duration of all-cause hospital stays for cardiovascular disease patients younger than 65 years as well as patients older than 65 years. The telehealth service also provided cost savings by decreasing the expenses of inpatient and total cost of all-cause health care in patients younger than 65 years and in those patients older than 65 years. To the best of our knowledge, this is the first study to demonstrate age differences in the clinical impact and cost-effectiveness of a real-time telehealth service in patients with cardiovascular diseases.

## Comparison to Prior Work

In contrast to traditional case management for chronic diseases, new methods for health promotion and disease control, including Internet-based interventions, telephonic support, home-based interventions, and telemedicine sessions, have been employed for various disease conditions. Telehealth services have been enthusiastically adopted for patients affected by chronic conditions, including chronic obstructive pulmonary disease [23], diabetes self-management [24], renal failure [25], and heart failure [26]. Telemedicine is an efficient approach and is suggested to be an important feature of heart failure management [9,11-13]. Based on advances in telecommunication technologies and their utilization in managing heart failure, 4 generations of telemedicine were proposed: (1) nonreactive data collection and analysis systems, (2) systems with nonimmediate analytical or decision-making structure, (3) remote patient management systems, and (4) fully integrated remote management systems, including invasive and noninvasive medical devices [20]. From the positive telephone support results and ongoing telemonitoring studies, it seems likely that telemedicine will be an efficient approach and become an important feature of heart failure management. However, 2 recent randomized clinical trials in heart failure did not corroborate these findings for morbidity- and mortality-related endpoints [17,18], and another telehealth study including elderly patients with multiple medical issues reported increased mortality rates when telemonitoring was in use [27]. The contradictory results are likely attributed to several factors.

Remote telemedicine delivery methods, such as synchronous (real-time) or asynchronous (store-and-forward) telehealth care delivery [28], vary and this can affect the clinical impact of telehealth care. After receiving the biometrics data through telemonitoring, the manner in which the telehealth service or telemedical remote management professionals utilize the data is one of the most important factors determining clinical impact.

In asynchronous telehealth care delivery, the biometric data can be transmitted synchronously or asynchronously. The data are often stored first and reviewed later; therefore, data processing and remote patient management can be delayed, particularly outside of office hours. The Tele-HF study is representative of asynchronous telemonitoring [18] in which biometric data were reviewed by the patient's physician every weekday; the Tele-HF study showed a negative result of telemonitoring in heart failure patients. In synchronous telehealth care delivery, biometric data are transmitted and processed on a real-time basis. The patient and health care provider need to be simultaneously available, as does the telemedical remote management professional. The Telemonitoring in the Management of Heart Failure (TEMA-HF) [29] and TIMF-HF [30] studies are representative of synchronous telemonitoring which requires a high degree of health professional participation for processing complex incoming physiological data and making subsequent therapeutic decisions. Although the TIM-HF study failed to reveal a clinical benefit, the TEMA-HF trial showed a reduced mortality rate in the telemonitoring group. Variations in the degree of urgency following biometrics transmission may account for the conflicting results of clinical research [20]. For the biometrics transmission and remote patient management to be synchronous, modalities such as telephone touch pad-based telemonitoring [31], video consultation-based telemonitoring [32], and iPhone-based telemonitoring [33] have been used. With advances in modern telecommunication technologies, telehealth has great potential to improve access to remote health care as an adjunct to traditional medical management. However, its adoption in routine health care has been slow, likely due to lack of clarity regarding the value of telehealth implementation [34].

Differences in selected physiological monitoring parameters (eg, body weight, heart rate, blood pressure, body temperature, electrocardiography, oxygen saturation, and fasting glucose) in noninvasive telemedical systems may influence the clinical outcome. However, the question of which physiological parameter provides the best benefit for cardiovascular disease remains unanswered. Alternatively, the solution must be connected to disease severity and the spectrum of disease.

Different diseases of interest or variations in disease severity may result in differences in the efficacy of telehealth services. Heart failure–related diseases with different New York Heart Association functional classes, different stages of chronic kidney disease, and different types of cardiorenal syndrome or associated arrhythmia would fundamentally govern the clinical course.

Social factors, cultural factors, diet habits, social habits, adherence to medical prescriptions, self-empowerment, consumer price index, medical practice behavior, and ethnic differences may govern the results of telehealth service investigations. Indeed, the key to telehealth service success is that the service should offer patients the opportunity to become actively involved in management of their own health care. Information sharing and communication techniques provide the opportunity for health professionals to improve patient health awareness outside of the hospital and ensure that the patient plays an active role in their own therapeutic process. The incorporation of these techniques also will represent a new approach to the therapeutic relationship, which will emulate the traditional face-to-face relationship between health care providers and patients in a remote capacity. Health professionals that adapt these techniques are able to provide consultations related to health promotion, disease prevention, and facilitation of illness recovery in a more responsive manner.

Age is always a major concern in medical practice that should never be overlooked. Older age is usually an independent prognostic factor for infectious disease, acute coronary syndrome, stroke, malignancy, and chronic kidney disease, among others. Despite not actually being considered a minority group, elderly patients may be discriminated against in several health care-related situations. Health insurance providers tend to exclude elderly patients from obtaining comprehensive insurance coverage. In developed countries, issues of restricted access to hospitals or facing unequal access to health care, referral, and treatment remain [3,4]. Clinical trials tend to exclude elderly patients [35]; hence, the conclusions reached in studies of nonelderly populations cannot be extrapolated to elderly patients. As the elderly population continues to grow, geriatrics is a burgeoning science and a subspecialty of internal and family medicine.

There are several concerns about the application of telehealth services for the elderly. First, elderly patients may have more comorbidities and chronic diseases that make periodic inpatient medical care in this group almost inevitable. Therefore, synchronous telehealth service for senior patients may not work as effectively as in nonsenior patients. Second, seniors might not be able to fulfill the telehealth care model completely due to the decline of their physical and mental condition. Utilizing computers and learning new communication technology may be difficult for elderly patients with impaired cognitive and physical function.

However, our study demonstrated that synchronous telehealth service can provide cost-effectiveness benefits as well as clinical benefits for reducing admission rates and duration of hospital stay regardless of age status. We believe that the success was because of the support from caregivers and families who built up the care and fulfilled the complete model of synchronous telehealth service. Although telehealth services cannot prevent diseases or major adverse cardiac events from occurring, the service may still be helpful in early detection of declining health status and assist in the delivery of timely medical therapy. Therefore, cost savings can still be achieved in elderly patients.

## Limitations

There were several limitations of our study. First, the study design lacked randomization, which resulted in the heterogeneity of the patient population and disease severity. There were also no control groups. Therefore, this study was a quasi-experimental study. Second, the relatively short period of follow-up may underestimate or overestimate the cost-effectiveness and clinical outcomes of the telehealth service for patients with cardiovascular disease. Although the initial 6 months of follow-up provided satisfactory results, this initial benefit may not guarantee a long-standing benefit or reflect mortality reduction. Third, the diversity of medical prescriptions

and baseline demographic data may potentially confound the results.

## Conclusions

The results of our study suggest that synchronous telehealth intervention may provide cost savings and clinical improvement in cardiovascular disease patients regardless of age status. Particularly in patients aged older than 65 years whose application of telehealth service was in doubt, they also benefit from reduced health care expenditures, reduced admission rates, and decreased duration of all-cause hospital stays when their illnesses are managed by using a telehealth service. In conclusion, telehealth medicine may be an effective model for cost savings in patients with cardiovascular diseases. Due to the limitations of the quasi-experimental study design, these findings require confirmation in a large randomized controlled trial in an older adult population with chronic diseases.

## Acknowledgments

We would like to thank Hui-Wen Wu, Tse-Pin Hsu, Shwu-Huey Yang, Jing-Lin Chen, Ting-Han Kung, Pei-Shan Chuang, Shu-Ling Tseng, Lee Jia-Wen, Li-Jung Tseng, Chien-Ling Yeh, Hsueh-I Lan, Wan-Ying Tsai, Ya-Ting Chen, and Wan-Chou Chen for their work in the Telehealth Center. This study was supported by the Taiwan National Science Council (grant numbers NSC 100-2220-E-002-020 and NSC 99-2911-I-008-100). This study was also supported by Taiwan University (grant number 10R71608-1) and the Department of Health, Executive Yuan, ROC (grant number DOH 99-TD-B-111-001).

## Conflicts of Interest

None declared.

## References

1. Wolff JL, Starfield B, Anderson G. Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. Arch Intern Med 2002 Nov 11;162(20):2269-2276. [Medline: 12418941]
2. Chi MJ, Lee CY, Wu SC. The prevalence of chronic conditions and medical expenditures of the elderly by chronic condition indicator (CCI). Arch Gerontol Geriatr 2011;52(3):284-289. [doi: 10.1016/j.archger.2010.04.017] [Medline: 20452688]
3. Davies N. Reducing inequalities in healthcare provision for older adults. Nurs Stand 2011;25(41):49-55; quiz 58. [Medline: 21815517]
4. Oliver D. Age based discrimination in health and social care services. BMJ 2009;339:b3400. [Medline: 19706655]
5. Thorpe KE, Howard DH. The rise in spending among Medicare beneficiaries: the role of chronic disease prevalence and changes in treatment intensity. Health Aff (Millwood) 2006;25(5):w378-w388 [FREE Full text] [doi: 10.1377/hlthaff.25.w378] [Medline: 16926180]
6. Welch G, Garb J, Zagarins S, Lendel I, Gabbay RA. Nurse diabetes case management interventions and blood glucose control: results of a meta-analysis. Diabetes Res Clin Pract 2010 Apr;88(1):1-6. [doi: 10.1016/j.diabres.2009.12.026] [Medline: 20116879]
7. Garcia-Aymerich J, Hernandez C, Alonso A, Casas A, Rodriguez-Roisin R, Anto JM, et al. Effects of an integrated care intervention on risk factors of COPD readmission. Respir Med 2007 Jul;101(7):1462-1469. [doi: 10.1016/j.rmed.2007.01.012] [Medline: 17339106]
8. Maron DJ, Boden WE, O'Rourke RA, Hartigan PM, Calfas KJ, Mancini GB, COURAGE Trial Research Group. Intensive multifactorial intervention for stable coronary artery disease: optimal medical therapy in the COURAGE (Clinical Outcomes Utilizing Revascularization and Aggressive Drug Evaluation) trial. J Am Coll Cardiol 2010 Mar 30;55(13):1348-1358. [doi: 10.1016/j.jacc.2009.10.062] [Medline: 20338496]
9. Clark RA, Inglis SC, McAlister FA, Cleland JG, 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 [FREE Full text] [doi: 10.1136/bmj.39156.536968.55] [Medline: 17426062]
10. Rosamond W, Flegal K, Friday G, Furie K, Go A, Greenlund K, American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics--2007 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation 2007 Feb 6;115(5):e69-171 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.106.179918] [Medline: 17194875]
11. Klersy C, De Silvestri A, Gabutti G, Regoli F, Auricchio A. A meta-analysis of remote monitoring of heart failure patients. J Am Coll Cardiol 2009 Oct 27;54(18):1683-1694. [doi: 10.1016/j.jacc.2009.08.017] [Medline: 19850208]
12. Clarke M, Shah A, Sharma U. Systematic review of studies on telemonitoring of patients with congestive heart failure: a meta-analysis. J Telemed Telecare 2011;17(1):7-14. [doi: 10.1258/jtt.2010.100113] [Medline: 21097564]
13. Inglis SC, Clark RA, McAlister FA, Stewart S, Cleland JG. Which components of heart failure programmes are effective? A systematic review and meta-analysis of the outcomes of structured telephone support or telemonitoring as the primary component of chronic heart failure management in 8323 patients: Abridged Cochrane Review. Eur J Heart Fail 2011 Sep;13(9):1028-1040. [doi: 10.1093/eurjhf/hfr039] [Medline: 21733889]

14. Chen YH, Ho YL, Huang HC, Wu HW, Lee CY, Hsu TP, et al. Assessment of the clinical outcomes and cost-effectiveness of the management of systolic heart failure in Chinese patients using a home-based intervention. J Int Med Res 2010;38(1):242-252. [Medline: 20233536]
15. Ho YL, Hsu TP, Chen CP, Lee CY, Lin YH, Hsu RB, et al. Improved cost-effectiveness for management of chronic heart failure by combined home-based intervention with clinical nursing specialists. J Formos Med Assoc 2007 Apr;106(4):313-319. [doi: 10.1016/S0929-6646(09)60258-8] [Medline: 17475609]
16. Hunt SA, Abraham WT, Chin MH, Feldman AM, Francis GS, Ganiats TG, et al. 2009 focused update incorporated into the ACC/AHA 2005 Guidelines for the Diagnosis and Management of Heart Failure in Adults: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines: developed in collaboration with the International Society for Heart and Lung Transplantation. Circulation 2009 Apr 14;119(14):e391-e479 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.109.192065] [Medline: 19324966]
17. Koehler F, Winkler S, Schieber M, Sechtem U, Stangl K, Böhm M, Telemedical Interventional Monitoring in Heart Failure Investigators. Impact of remote telemedical management on mortality and hospitalizations in ambulatory patients with chronic heart failure: the telemedical interventional monitoring in heart failure study. Circulation 2011 May 3;123(17):1873-1880 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.111.018473] [Medline: 21444883]
18. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. N Engl J Med 2010 Dec 9;363(24):2301-2309 [FREE Full text] [doi: 10.1056/NEJMoa1010029] [Medline: 21080835]
19. Bui AL, Fonarow GC. Home monitoring for heart failure management. J Am Coll Cardiol 2012 Jan 10;59(2):97-104 [FREE Full text] [doi: 10.1016/j.jacc.2011.09.044] [Medline: 22222071]
20. Anker SD, Koehler F, Abraham WT. Telemedicine and remote management of patients with heart failure. Lancet 2011 Aug 20;378(9792):731-739. [doi: 10.1016/S0140-6736(11)61229-4] [Medline: 21856487]
21. Hunt SA, Baker DW, Chin MH, Cinquegrani MP, Feldman AM, Francis GS, American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Revise the 1995 Guidelines for the Evaluation and Management of Heart Failure), International Society for Heart and Lung Transplantation, Heart Failure Society of America. ACC/AHA Guidelines for the Evaluation and Management of Chronic Heart Failure in the Adult: Executive Summary A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Revise the 1995 Guidelines for the Evaluation and Management of Heart Failure): Developed in Collaboration With the International Society for Heart and Lung Transplantation; Endorsed by the Heart Failure Society of America. Circulation 2001 Dec 11;104(24):2996-3007 [FREE Full text] [Medline: 11739319]
22. Hunt SA, American College of Cardiology, American Heart Association Task Force on Practice Guidelines (Writing Committee to Update the 2001 Guidelines for the Evaluation and Management of Heart Failure). ACC/AHA 2005 guideline update for the diagnosis and management of chronic heart failure in the adult: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Update the 2001 Guidelines for the Evaluation and Management of Heart Failure). J Am Coll Cardiol 2005 Sep 20;46(6):e1-82. [doi: 10.1016/j.jacc.2005.08.022] [Medline: 16168273]
23. Bartoli L, Zanaboni P, Masella C, Ursini N. Systematic review of telemedicine services for patients affected by chronic obstructive pulmonary disease (COPD). Telemed J E Health 2009 Nov;15(9):877-883. [doi: 10.1089/tmj.2009.0044] [Medline: 19919194]
24. Davis RM, Hitch AD, Salaam MM, Herman WH, Zimmer-Galler IE, Mayer-Davis EJ. TeleHealth improves diabetes self-management in an underserved community: diabetes TeleCare. Diabetes Care 2010 Aug;33(8):1712-1717 [FREE Full text] [doi: 10.2337/dc09-1919] [Medline: 20484125]
25. Berman SJ, Wada C, Minatodani D, Halliday T, Miyamoto R, Lindo J, et al. Home-based preventative care in high-risk dialysis patients: a pilot study. Telemed J E Health 2011 May;17(4):283-287. [doi: 10.1089/tmj.2010.0169] [Medline: 21480788]
26. Maric B, Kaan A, Ignaszewski A, Lear SA. A systematic review of telemonitoring technologies in heart failure. Eur J Heart Fail 2009 May;11(5):506-517. [doi: 10.1093/eurjhf/hfp036] [Medline: 19332417]
27. Takahashi PY, Pecina JL, Upatising B, Chaudhry R, Shah ND, Van Houten H, et al. A randomized controlled trial of telemonitoring in older adults with multiple health issues to prevent hospitalizations and emergency department visits. Arch Intern Med 2012 May 28;172(10):773-779. [doi: 10.1001/archinternmed.2012.256] [Medline: 22507696]
28. Allely EB. Synchronous and asynchronous telemedicine. J Med Syst 1995 Jun;19(3):207-212. [Medline: 7643019]
29. Dendale P, De Keulenaer G, Troisfontaines P, Weytjens C, Mullens W, Elegeert I, et al. Effect of a telemonitoring-facilitated collaboration between general practitioner and heart failure clinic on mortality and rehospitalization rates in severe heart failure: the TEMA-HF 1 (TElemonitoring in the MAnagement of Heart Failure) study. Eur J Heart Fail 2012 Mar;14(3):333-340. [doi: 10.1093/eurjhf/hfr144] [Medline: 22045925]
30. Koehler F, Winkler S, Schieber M, Sechtem U, Stangl K, Böhm M, TIM-HF Investigators. Telemedical Interventional Monitoring in Heart Failure (TIM-HF), a randomized, controlled intervention trial investigating the impact of telemedicine on mortality in ambulatory patients with heart failure: study design. Eur J Heart Fail 2010 Dec;12(12):1354-1362. [doi: 10.1093/eurjhf/hfq199] [Medline: 21098580]

31. Spaeder J, Najjar SS, Gerstenblith G, Hefter G, Kern L, Palmer JG, et al. Rapid titration of carvedilol in patients with congestive heart failure: a randomized trial of automated telemedicine versus frequent outpatient clinic visits. Am Heart J 2006 Apr;151(4):844.e1-844.10. [doi: 10.1016/j.ahj.2005.06.044] [Medline: 16569544]
32. 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]
33. Anderson ER, Smith B, Ido M, Frankel M. Remote assessment of stroke using the iPhone 4. J Stroke Cerebrovasc Dis 2011 Oct 21:1-5. [doi: 10.1016/j.jstrokecerebrovasdis.2011.09.013] [Medline: 22018507]
34. Cusack CM, Pan E, Hook JM, Vincent A, Kaelber DC, Middleton B. The value proposition in the widespread use of telehealth. J Telemed Telecare 2008;14(4):167-168. [doi: 10.1258/jtt.2007.007043] [Medline: 18534947]
35. Bugeja G, Kumar A, Banerjee AK. Exclusion of elderly people from clinical research: a descriptive study of published reports. BMJ 1997 Oct 25;315(7115):1059 [FREE Full text] [Medline: 9366735]


## Abbreviations

**ACC:** American College of Cardiology
**AHA:** American Heart Association
**CAD:** coronary artery disease
**eGFR:** estimated glomerular filtration rate
**MDRD:** Modification of Diet in Renal Disease
**NTUH:** National Taiwan University Hospital


*Edited by J Keelan; submitted 22.02.12; peer-reviewed by C Lynch, M Stuart, G Chen; comments to author 24.05.12; revised version received 25.07.12; accepted 19.03.13; published 24.04.13.*

*Please cite as:*
*Chen YH, Lin YH, Hung CS, Huang CC, Yeih DF, Chuang PY, Ho YL, Chen MF*
*Clinical Outcome and Cost-Effectiveness of a Synchronous Telehealth Service for Seniors and Nonseniors with Cardiovascular Diseases: Quasi-Experimental Study*
*J Med Internet Res 2013;15(4):e87*
*URL: http://www.jmir.org/2013/4/e87/*
*doi:10.2196/jmir.2091*
*PMID:23615318*

Original Paper

# How Valid are Web-Based Self-Reports of Weight?

Stephanie Erika Bonn[1], MSc; Ylva Trolle Lagerros[2], MD, PhD; Katarina Bälter[1], PhD

[1]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
[2]Clinical Epidemiology Unit, Department of Medicine, Karolinska Institutet, Stockholm, Sweden

**Corresponding Author:**
Stephanie Erika Bonn, MSc
Department of Medical Epidemiology and Biostatistics
Karolinska Institutet
Nobels väg 12a
Stockholm, SE-17177
Sweden
Phone: 46 852482298
Fax: 46 8314975
Email: stephanie.bonn@ki.se

## Abstract

**Background:** Many studies rely on self-reported anthropometric data. While paper-based self-reports have been the standard collection mode, the number of studies collecting self-reported data via the Web is increasing rapidly. Although numerous studies have shown good agreement between self-reported and measured weight using paper-based questionnaires, the validity of using the Web to inquire about weight is unknown.

**Objective:** The objective of this study was to validate Web-based self-reports of bodyweight compared to weight measured at the study center.

**Methods:** The validity of weight self-reported via the Web was assessed by comparing self-reports against measurements of weight in a convenience sample of 149 individuals (77.2% women, 115/149), aged 20-65 years. Study participants self-reported their weight via a Web-based questionnaire and thereafter had their weight measured in the research center.

**Results:** The Spearman correlation coefficient between self-reported and measured weight was 0.98 ($P$<.001). The mean difference between self-reported and measured weight was -1.2 (SD 2.6) kg. There was a statistically significant difference between self-reported and measured weight with the self-reported being lower ($P$<.001). Subjects with a body mass index (BMI) ≥25 kg/m$^2$, and subjects ≥30 years of age, under-reported their weight statistically significantly more than subjects with a BMI <25 kg/m$^2$, and subjects <30 years of age, respectively.

**Conclusions:** Our results show that self-reported weight via the Web can be a valid method of data collection.

***(J Med Internet Res 2013;15(4):e52)*** doi:10.2196/jmir.2393

**KEYWORDS**

body weight; Internet; validity

## Introduction

Many studies rely on self-reported data on anthropometric variables. This is more practical and less expensive to collect, than measured data, especially when handling a large sample size. Nonetheless, self-reported data is limited by information bias [1]. Considering the increasing number of studies collecting Web-based data, there is a need to determine the validity of Web-based self-reports of body weight.

It has been suggested that using the Internet creates a distance between the investigator and the subject, as a self-reported questionnaire increases the anonymity of the subject responding. This makes it easier for the subject to answer personal questions truthfully and encourages reporting of uncensored personal information [2]. Sensitive or stigmatized behaviors have been reported more truthfully in self-administered surveys compared to interview-conducted surveys [3,4]. A previous study investigating the validity of self-reported weight by telephone interview and a paper questionnaire found that overweight and obese women reported weight more accurately in the questionnaire [5]. Heavy drinking of alcohol is another example of a stigmatized behavior where a Web-based questionnaire

captured more heavy drinkers compared to telephone interviews [6].

While many studies have evaluated paper-based self-reports of weight, it is unclear how valid Web-based self-reports are. The aim of this study was to validate Web-based self-reports of weight against weight measurements made by the study personnel at a research center.

## Methods

### Study Design

Individuals between the ages of 20 and 65 years were recruited for participation during the spring of 2009 by public advertisements in Stockholm County, Sweden. The primary aim of the study was to validate methods assessing diet and physical activity. Access to the Internet and an email address were the requirements for participation. Subjects who were on any form of weight alteration diet, pregnant, or had given birth during the 10 months prior to the start of the study were excluded from participation. Participants were provided with information about the study and gave their written informed consent prior to the start of the study. The study design was previously described in detail by our research group [7]. The study was approved by the research ethics committee at Karolinska Institutet, Stockholm, Sweden.

In total, 179 subjects were included for participation in the study. Out of them, 150 subjects applied via the Web by filling in a questionnaire with personal information, including height and current weight. The questionnaire was downloaded from the Web, filled out by applicants on the computer, and thereafter sent via email to the researchers. There were no instructions regarding clothing or of subjects having to weigh themselves specifically for the questionnaire. Subjects were not aware that their weight was going to be measured at the start of the study when applying for participation. The remaining 29 subjects applied via personal contact or regular paper mail and were not included in the analysis of the present study. Approximately 2 months after applying for participation, subjects attended an introductory meeting and had their weight measured by study personnel. Subjects wore indoor clothing and no shoes during weight measurements. Measurements were made using a digital scale displaying weight to a tenth of a kg. All measurements were made by the same study personnel using the same scale. One subject out of the 150 who applied via the Web had no data on measured weight at the study center and was excluded from further analyses. Thus, 149 subjects were included in statistical analyses.

### Statistical Methods

Characteristics of the study participants are presented as descriptive statistics. Results of self-reported weight and height, measured weight, and differences between assessments, are reported as mean values and standard deviations. Potential differences between men and women, subjects <30 and ≥30 years of age, and subjects with a body mass index (BMI) <25 and ≥25 kg/m$^2$, were assessed using paired *t* tests for continuous variables and chi-square tests for categorical variables. We also performed multivariate linear regression controlling for sex, age (<30 or ≥30 years), BMI (<25 or ≥25 kg/m$^2$), education, and smoking. The degree of association between self-reported and measured weight was assessed using Spearman and intraclass correlation coefficients [8]. Because systematic differences between assessments cannot be detected using Spearman correlation coefficients, absolute agreement between the assessments was determined by using the Bland-Altman technique [9]. The difference between self-reported and measured weight (y-axis) was plotted against the mean of the two assessments (x-axis). The significance level was set to $P<.05$. All analyses were performed using STATA 11.1

## Results

Characteristics of subjects included in analyses are presented in Table 1. The majority of subjects were female (77.2%, 115/149), <30 years of age (51.7%, 77/149), and had a BMI <25 kg/m$^2$ (76.5%, 114/149). Men reported using Swedish snuff, a moist form of snuff, to a higher extent than women ($P=.000$). No other statistically significant differences were found between men and women, or between subjects with a BMI <25 or ≥25 kg/m$^2$.

The Spearman correlation coefficient between self-reported weight via the Web and weight measured by study personnel was 0.98 ($P<.001$) for all subjects (Figure 1). The Spearman correlation coefficients for men and women were 0.89 ($P<.001$) and 0.97 ($P<.001$) respectively. For subjects with a BMI <25 kg/m$^2$ and ≥25 kg/m$^2$, the Spearman correlation coefficients between self-reported and measured weight were 0.97 ($P<.001$) and 0.98 ($P<.001$) respectively. Intraclass correlation coefficients were almost identical to the Spearman correlation coefficients (results not shown). There was a statistically significant difference between self-reported and measured weight in the whole group ($P<.001$), and in women ($P<.001$). There was a borderline statistical difference in men ($P=.066$). Further, there were statistically significant differences between self-reported and measured weight independent of BMI (<25 kg/m$^2$, $P<.001$ and ≥25 kg/m$^2$, $P=.002$).

Figure 2 shows a Bland-Altman plot graphically illustrating differences between self-reported and measured weight among all subjects, the mean difference between self-reported and measured weight was -1.2 kg (SD 2.64). The plot shows good agreement between the two methods. However, the trend in the plot indicates increased under-reporting of self-reported weight with increasing body weight.

Table 2 shows self-reported weight, height and BMI, measured weight, and differences between self-reported and measured weight. Self-reported weight was under-reported compared to measured weight for all subjects ($P<.001$). The difference between assessments remained highly significant among women, -1.3 kg ($P<.001$), but not among men, -0.9 kg ($P=.07$). The difference between self-reported and measured weight was not statistically significantly different between men and women. For subjects with a BMI <25 kg/m$^2$ and ≥25 kg/m$^2$, there was a statistically significant difference between the groups ($P=.02$), with heavier subjects under-reporting on average -1.2 kg more

compared to leaner subjects. Years of education did not appear to affect self-reports of weight while subjects ≥30 years of age under-reported their weight more than subjects <30 years of age ($P$=.02).

Results from the multivariate regression controlling for sex, age (<30 or ≥30 years), BMI (<25 or ≥25 kg/m$^2$), education, and smoking were similar to the results described above. There were statistically significantly higher under-reporting among subjects ≥30 years and subjects with a BMI ≥25 kg/m$^2$ compared to subjects <30 years of age and a BMI of <25 kg/m$^2$, respectively. Sex, education, and smoking did not affect the difference between self-reported and measured weight (data not shown).

**Table 1.** Characteristics of study participants.

| | All | | Men | | Women | | $P$ value[a] |
|---|---|---|---|---|---|---|---|
| | n=149 | | n=34 | | n=115 | | |
| | n | % | n | % | n | % | |
| **Age (years)** | | | | | | | .45 |
| <30 | 77 | 51.7 | 18 | 52.9 | 59 | 51.3 | |
| 30-39 | 28 | 18.8 | 7 | 20.6 | 21 | 18.3 | |
| 40-49 | 22 | 14.8 | 7 | 20.6 | 15 | 13.0 | |
| 50-59 | 18 | 12.1 | 2 | 5.9 | 16 | 13.9 | |
| >60 | 4 | 2.7 | 0 | 0.0 | 4 | 3.5 | |
| **Education (years)** | | | | | | | .46 |
| 9-12[b] | 32 | 21.5 | 9 | 26.5 | 23 | 20.0 | |
| >12 | 114 | 76.5 | 25 | 73.5 | 89 | 77.4 | |
| Missing data | 3 | 2.0 | 0 | 0.0 | 3 | 2.6 | |
| **Smoker** | | | | | | | .12 |
| Current | 11 | 7.4 | 5 | 14.7 | 6 | 5.2 | |
| Previous | 38 | 25.5 | 6 | 17.6 | 32 | 27.8 | |
| Never | 96 | 64.4 | 23 | 67.6 | 73 | 63.5 | |
| Missing data | 4 | 2.7 | 0 | 0.0 | 4 | 3.5 | |
| **Swedish snuff[c] user** | | | | | | | .000 |
| Current | 7 | 4.7 | 6 | 17.6 | 1 | 0.9 | |
| Previous | 16 | 10.7 | 7 | 20.6 | 9 | 7.8 | |
| Never | 122 | 81.9 | 21 | 61.8 | 101 | 87.8 | |
| Missing data | 4 | 2.7 | 0 | 0.0 | 4 | 3.5 | |
| **BMI (kg/m$^2$)[d]** | | | | | | | .11 |
| <25 | 114 | 76.5 | 27 | 79.4 | 87 | 75.7 | |
| ≥25 | 35 | 23.5 | 7 | 20.6 | 28 | 24.3 | |

[a]comparing men and women using chi-square tests

[b]primary school

[c]a moist form of snuff

[d]BMI based on self-reported weight and height

**Table 2.** Weight, height, and BMI based on self-reported and measured data.

| | | Self-report Mean (SD) | | | Measured Mean (SD) | Difference[a] Mean (SD) |
|---|---|---|---|---|---|---|
| | n | Weight (kg) | Height (cm) | BMI (kg/m$^2$) | Weight (kg) | Weight (kg) |
| **All** | | | | | | |
| | 149 | 68.7 (12.6) | 170.6 (9.5) | 23.6 (3.8) | 69.9 (13.2) | -1.2 (2.6) |
| **Gender** | | | | | | |
| Men | 34 | 80.0 (7.8) | 183.2 (7.3) | 23.9 (2.2) | 80.9 (8.5) | -0.9 (2.7) |
| Women | 115 | 65.3 (11.7) | 166.9 (6.4) | 23.5 (4.1) | 66.6 (12.5) | -1.3 (2.6) |
| *P* value | | <.001 | <.001 | .55 | <.001 | .44 |
| **BMI (kg/m$^2$)** | | | | | | |
| <25 | 114 | 64.9 (9.7) | 171.4 (9.6) | 22.0 (1.7) | 65.8 (10.0) | -0.9 (2.1) |
| ≥25 | 35 | 80.8 (13.5) | 168.3 (8.8) | 28.5 (4.3) | 83.0 (13.9) | -2.1 (3.8) |
| *P* value | | <.001 | .09 | <.001 | <.001 | .02 |
| **Age (years)** | | | | | | |
| <30 | 77 | 66.6 (10.8) | 171.7 (9.7) | 22.5 (2.6) | 67.3 (11.5) | -0.7 (2.5) |
| ≥30 | 72 | 70.9 (14.0) | 169.5 (9.2) | 24.7 (4.4) | 72.6 (14.3) | -1.7 (2.7) |
| *P* value | | .04 | .14 | <.001 | 0.01 | .03 |

[a]difference between measured and self-reported weight

**Figure 1.** Scatter plot showing the correlation between self-reported (y-axis) and measured (x-axis) weight (kg). The Spearman correlation coefficient was 0.98 (*P*<.001). Each data point represents one subject. n=149.

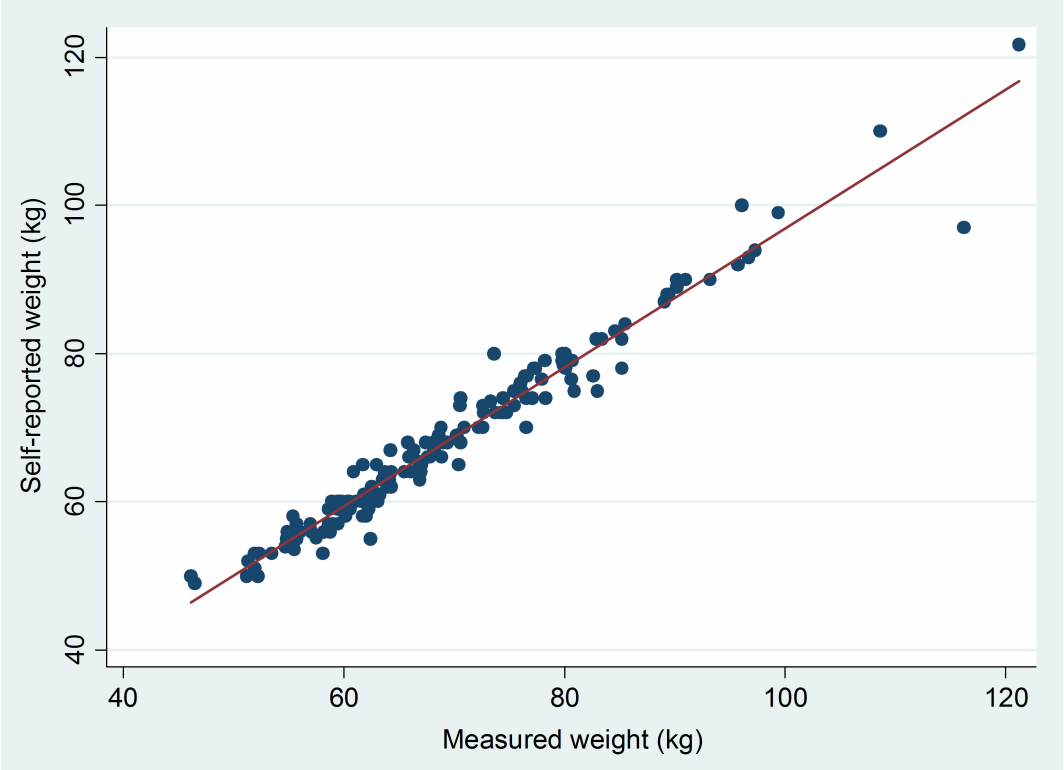
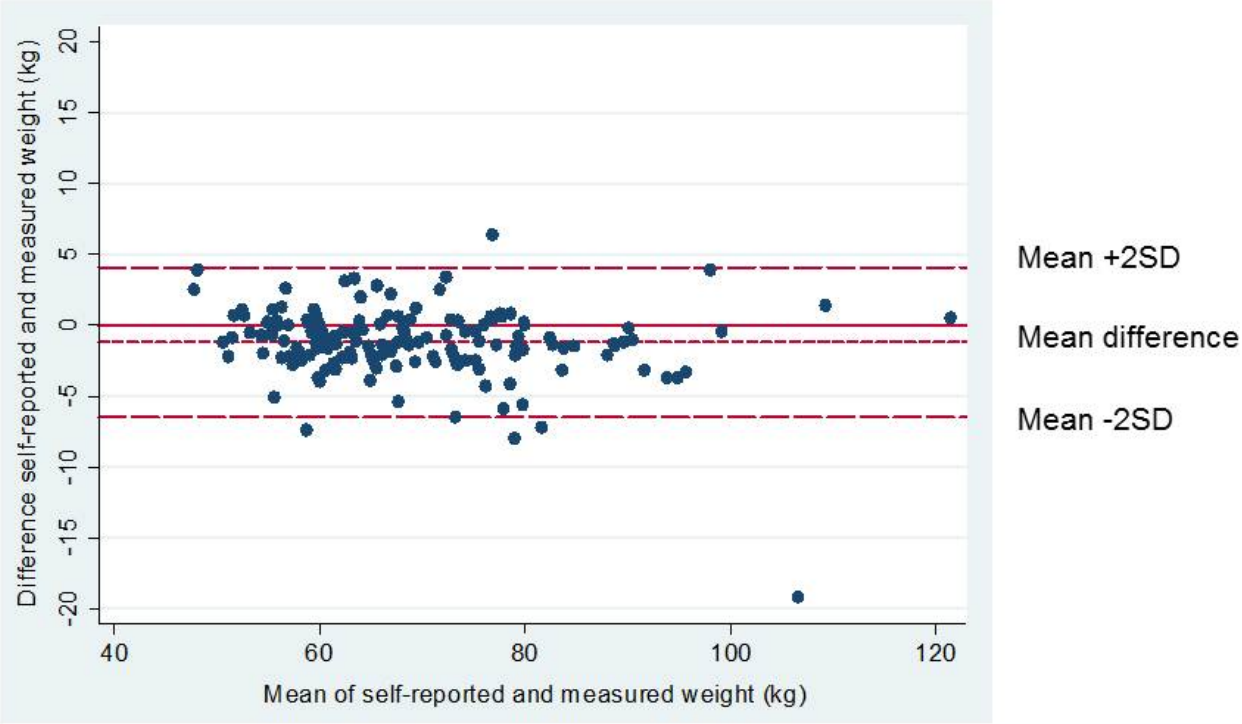


**Figure 2.** Bland-Altman plot showing the difference between self-reported and measured weight (y-axis) in relation to measured weights (kg). The mean difference was -1.2 kg. The 95% limits of agreement for the observations are represented as ±2 SD of the mean difference. Each data point represents one subject. n=149.




## Discussion

### General Discussion

Results from the present study show that Web-based self-reports of weight are highly correlated, although somewhat under-reported, with measured weight in a Swedish population. Although there were small differences between subgroups, overweight and obese subjects with a BMI ≥25 kg/m$^2$ under-reported their weight on average by 1.2 kg more than normal weight subjects. Further, we found that subjects ≥30 years of age, under-reported their weight on average by 1.0 kg more than younger subjects. Under-reporting was associated with both increasing BMI and age independently. No differences were seen between men and women. To the best of our knowledge, the present study is among the first to examine the validity of Web-based self-reports of weight.

Our results are in line with previous studies validating paper-based self-reports of weight against measurements of weight, which have shown high correlations and high validity [5,10-14]. Nevertheless, self-reported weight is under-estimated in most populations [15], and by most individuals [5,12,16], although under-weight subjects have been seen to over-report weight [5]. Consistent with the results of previous studies where increasing body mass was associated with increased under-reporting [15,17,18], our results demonstrated increased under-reporting among subjects with a high BMI.

As was also shown in a previous study [19], we saw an indication of gender difference, with women under-reporting their weight to a greater degree than men. The lack of statistical significance between the sexes' degree of under-reporting might be due to our small sample size and fewer participating men. The greater under-reporting detected among older subjects may partly be explained by the positive association between BMI and age and the increasing under-reporting seen with increasing BMI. Increased under-reporting in older subjects could also partly be explained by memory bias due to older subjects not keeping track of their current weight to the same extent as younger subjects. Previous studies have also shown age to have an effect on under-reporting [15].

The observed under-reporting of self-reported weight compared to measured weight, may partly be explained by true differences between the time of self-reported and measured weight. However, the time period between the self-report and the measurement was fairly short and participants were healthy and not on any weight changing regimen. Thus, we believe that large changes in weight during this period were unlikely to have occurred, and does not fully explain the difference. Nonetheless, a narrower time period between the self-reported and measured weights would be ideal to avoid fluctuations in weight over time. Some of the under-estimation might be explained by subjects self-reporting their weight without clothes, while the measurements of weight were made while wearing clothing. Although this may explain part of the under-estimation, it is not likely to have caused the entire difference. Another explanation for differences might be the fact that subjects self-reported their weight in different manners as no specific instructions for self-reports were given, causing random errors. Errors in the measurement of weight were unlikely, given that the same scale was used by the study personnel for all measurements.

Furthermore, the lack of measured height was also a limitation to the current study, and may have had an effect on estimates of BMI. However, previous studies have shown good validity of categorization according to BMI assessed from self-reported

data compared to measured data [5]. Therefore, we believe that our estimates of BMI assessed from self-reported weight and height were fairly accurate.

The use of Web-based self-reports of data has numerous advantages compared to collecting data using paper-based questionnaires. Data quality can be improved by implementing automatic controls for missing data and answers out of a reasonable range. The quality of anthropometric data, with regard to missing and plausible answers, collected using a Web-based questionnaire, has been shown to be equal to, or better than, that of data collected using a paper version of the questionnaire [20]. While Web-based questionnaires may be superior to their paper-based predecessors when it comes to completeness of data, concerns have been raised regarding response rates and selection bias introduced by using online surveys [21]. We do not believe, however, that these concerns will be a problem when collecting data in future studies. Access to, and use of, the Internet in Sweden has increased considerably during the last decade with more than 90% of the adult population having access today [22].

Like many other convenience samples, the present study population comprised of men and women of different ages, but the majority of subjects were nonetheless young females residing in Stockholm County, limiting the generalizability of the results. A truly population-based study would improve generalizability, but was clearly unfeasible for practical reasons. In addition, the study subjects who applied for participation in the study knew that the aim of the study was to evaluate methods assessing diet and physical activity. They may therefore be more health conscious and motivated to participate than the average population, and thus prone to self-reporting their weight in a more truthful manner. Nonetheless, the study participants were not initially aware of the comparisons between self-reported and measured weight.

The self-selection of participants might have created a sampling bias yielding a stronger correlation and a decreased difference between assessment methods than in the general population. Of note is that the study subjects had a mean BMI comparable to the general Swedish population [23]. Nonetheless, future studies should focus on validity of self-reports by other sub-groups, such as elderly or obese. Despite these potential limitations, our results may be helpful when using self-reported weights in studies in young and normal weight populations.

## Conclusions

We have found Web-based self-reports of weight to be as good as paper-based self-reports of weight found in previous studies. Although more validation studies may be needed, our results showed that self-reported weight via the Internet is a suitable method of data collection for use in research and clinical work.

## Acknowledgments

This work was supported by grants from the Swedish Cancer Foundation as well as grants to LifeGene from AFA Insurance, Torsten and Ragnar Söderbergs Foundation, Karolinska Institutet, Stockholm's County Council, and the Swedish Research Council. The authors thank Sara Christensen and Elisabeth Möller for their invaluable contributions during the collection of data.

## Conflicts of Interest

None declared.

## References

1. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Selection bias and information bias in clinical research. Nephron Clin Pract 2010;115(2):c94-c99 [FREE Full text] [doi: 10.1159/000312871] [Medline: 20407272]
2. Joinson AN. Knowing me, knowing you: reciprocal self-disclosure in Internet-based surveys. Cyberpsychol Behav 2001 Oct;4(5):587-591. [Medline: 11725651]
3. Couper M. Web surveys: a review of issues and approaches. Public Opin Q 2000;64(4):464-494. [Medline: 11171027]
4. Okamoto K, Ohsuka K, Shiraishi T, Hukazawa E, Wakasugi S, Furuta K. Comparability of epidemiological information between self- and interviewer-administered questionnaires. J Clin Epidemiol 2002 May;55(5):505-511. [Medline: 12007554]
5. Lin CJ, DeRoo LA, Jacobs SR, Sandler DP. Accuracy and reliability of self-reported weight and height in the Sister Study. Public Health Nutr 2012 Jun;15(6):989-999 [FREE Full text] [doi: 10.1017/S1368980011003193] [Medline: 22152926]
6. Link MW, Mokdad AH. Effects of survey mode on self-reports of adult alcohol consumption: a comparison of mail, web and telephone approaches. J Stud Alcohol 2005 Mar;66(2):239-245. [Medline: 15957675]
7. Bonn SE, Trolle Lagerros Y, Christensen SE, Möller E, Wright A, Sjölander A, et al. Active-Q: validation of the web-based physical activity questionnaire using doubly labeled water. J Med Internet Res 2012;14(1):e29 [FREE Full text] [doi: 10.2196/jmir.1974] [Medline: 22356755]
8. Lee J, Koh D, Ong CN. Statistical evaluation of agreement between two methods for measuring a quantitative variable. Comput Biol Med 1989;19(1):61-70. [Medline: 2917462]
9. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307-310. [Medline: 2868172]
10. Dahl AK, Hassing LB, Fransson EI, Pedersen NL. Agreement between self-reported and measured height, weight and body mass index in old age--a longitudinal study with 20 years of follow-up. Age Ageing 2010 Jul;39(4):445-451 [FREE Full text] [doi: 10.1093/ageing/afq038] [Medline: 20453247]

11. Lawlor DA, Bedford C, Taylor M, Ebrahim S. Agreement between measured and self-reported weight in older women. Results from the British Women's Heart and Health Study. Age Ageing 2002 May;31(3):169-174 [FREE Full text] [Medline: 12006304]
12. Bes-Rastrollo M, Sabaté J, Jaceldo-Siegl K, Fraser GE. Validation of self-reported anthropometrics in the Adventist Health Study 2. BMC Public Health 2011;11:213 [FREE Full text] [doi: 10.1186/1471-2458-11-213] [Medline: 21466678]
13. Burton NW, Brown W, Dobson A. Accuracy of body mass index estimated from self-reported height and weight in mid-aged Australian women. Aust N Z J Public Health 2010 Dec;34(6):620-623. [doi: 10.1111/j.1753-6405.2010.00618.x] [Medline: 21134066]
14. Spencer EA, Appleby PN, Davey GK, Key TJ. Validity of self-reported height and weight in 4808 EPIC-Oxford participants. Public Health Nutr 2002 Aug;5(4):561-565. [doi: 10.1079/PHN2001322] [Medline: 12186665]
15. Connor Gorber S, Tremblay M, Moher D, Gorber B. A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. Obes Rev 2007 Jul;8(4):307-326. [doi: 10.1111/j.1467-789X.2007.00347.x] [Medline: 17578381]
16. Engstrom JL, Paterson SA, Doherty A, Trabulsi M, Speer KL. Accuracy of self-reported height and weight in women: an integrative review of the literature. J Midwifery Womens Health 2003;48(5):338-345. [Medline: 14526347]
17. Larsen JK, Ouwens M, Engels RC, Eisinga R, van Strien T. Validity of self-reported weight and height and predictors of weight bias in female college students. Appetite 2008;50(2-3):386-389. [doi: 10.1016/j.appet.2007.09.002] [Medline: 18006115]
18. Nyholm M, Gullberg B, Merlo J, Lundqvist-Persson C, Råstam L, Lindblad U. The validity of obesity based on self-reported weight and height: Implications for population studies. Obesity (Silver Spring) 2007 Jan;15(1):197-208. [doi: 10.1038/oby.2007.536] [Medline: 17228048]
19. Merrill RM, Richardson JS. Validity of self-reported height, weight, and body mass index: findings from the National Health and Nutrition Examination Survey, 2001-2006. Prev Chronic Dis 2009 Oct;6(4):A121 [FREE Full text] [Medline: 19754997]
20. Touvier M, Méjean C, Kesse-Guyot E, Pollet C, Malon A, Castetbon K, et al. Comparison between web-based and paper versions of a self-administered anthropometric questionnaire. Eur J Epidemiol 2010 May;25(5):287-296. [doi: 10.1007/s10654-010-9433-9] [Medline: 20191377]
21. Kongsved SM, Basnov M, Holm-Christensen K, Hjollund NH. Response rate and completeness of questionnaires: a randomized study of Internet versus paper-and-pencil versions. J Med Internet Res 2007;9(3):e25 [FREE Full text] [doi: 10.2196/jmir.9.3.e25] [Medline: 17942387]
22. Arrhenius C, Ewerdahl D. Statistics Sweden. 2011. Use of computers and the Internet by private persons in 2010 URL: http://www.scb.se/statistik/_publikationer/LE0108_2010A01_BR_IT01BR1101.pdf [accessed 2011-10-18] [WebCite Cache ID 62Wgb5Nr0]
23. Sundquist J, Johansson SE, Sundquist K. Levelling off of prevalence of obesity in the adult population of Sweden between 2000/01 and 2004/05. BMC Public Health 2010;10:119 [FREE Full text] [doi: 10.1186/1471-2458-10-119] [Medline: 20214805]


## Abbreviations

**BMI:** body mass index


*Edited by G Eysenbach; submitted 12.10.12; peer-reviewed by J Kraschnewski, W You, J Harvey-Berino, M Bes-Rastrollo, S Lim Choi Keung, B Brumen; comments to author 09.11.12; revised version received 20.12.12; accepted 08.02.13; published 09.04.13.*

*Please cite as:*
*Bonn SE, Trolle Lagerros Y, Bälter K*
*How Valid are Web-Based Self-Reports of Weight?*
*J Med Internet Res 2013;15(4):e52*
*URL: http://www.jmir.org/2013/4/e52/*
*doi:10.2196/jmir.2393*
*PMID:23570956*

Original Paper

# Self-Test Web-Based Pure-Tone Audiometry: Validity Evaluation and Measurement Error Analysis

Marcin Masalski[1,2], MD, PhD; Tomasz Kręcicki[1], MD, PhD

[1]Department and Clinic of Otolaryngology, Head and Neck Surgery, Wroclaw Medical University, Wrocław, Poland

[2]Institute of Biomedical Engineering and Instrumentation, Wroclaw University of Technology, Wrocław, Poland

**Corresponding Author:**
Marcin Masalski, MD, PhD
Department and Clinic of Otolaryngology, Head and Neck Surgery
Wroclaw Medical University
Borowska 213
Wrocław, 50-556
Poland
Phone: 48 71 734 3700
Fax: 48 71 733 12 09
Email: marcin.masalski@pwr.wroc.pl

## Abstract

**Background:** Potential methods of application of self-administered Web-based pure-tone audiometry conducted at home on a PC with a sound card and ordinary headphones depend on the value of measurement error in such tests.

**Objective:** The aim of this research was to determine the measurement error of the hearing threshold determined in the way described above and to identify and analyze factors influencing its value.

**Methods:** The evaluation of the hearing threshold was made in three series: (1) tests on a clinical audiometer, (2) self-tests done on a specially calibrated computer under the supervision of an audiologist, and (3) self-tests conducted at home. The research was carried out on the group of 51 participants selected from patients of an audiology outpatient clinic. From the group of 51 patients examined in the first two series, the third series was self-administered at home by 37 subjects (73%).

**Results:** The average difference between the value of the hearing threshold determined in series 1 and in series 2 was -1.54dB with standard deviation of 7.88dB and a Pearson correlation coefficient of .90. Between the first and third series, these values were -1.35dB±10.66dB and .84, respectively. In series 3, the standard deviation was most influenced by the error connected with the procedure of hearing threshold identification (6.64dB), calibration error (6.19dB), and additionally at the frequency of 250Hz by frequency nonlinearity error (7.28dB).

**Conclusions:** The obtained results confirm the possibility of applying Web-based pure-tone audiometry in screening tests. In the future, modifications of the method leading to the decrease in measurement error can broaden the scope of Web-based pure-tone audiometry application.

***(J Med Internet Res 2013;15(4):e71)*** doi:10.2196/jmir.2222

**KEYWORDS**

pure tone audiometry; computer-assisted instruction; self-examination

## Introduction

The development of Internet technologies combined with technological progress in the construction of personal computers, especially in terms of improving performance of sound cards, offer the possibility of conducting self-administered hearing tests at home. Moreover, in a group of elderly people who often suffer from hearing loss, there is an increasing number of Internet users. The validity of such hearing tests needs to be evaluated before applying the test results in the diagnostic and therapeutic process.

The remote examinations of hearing conducted with the use of PCs have been carried out for about 10 years [1]. These examinations can be divided into surveys [2-4] and examinations done with the use of sound signals. The sound signals can be generated through a dedicated external device connected to a PC, usually an audiometer [5,6] or a PC sound card [3,4,7-12]. Among hearing examinations, we can distinguish screening

tests [3,7,11,12] and examinations such as pure-tone audiometry [4-6,8-10], which provide additional information.

The hearing screening tests done remotely with the use of a PC sound card are most often in the form of a speech-in-noise test [3,7,11,12]. The speech-in-noise test is preferred over surveys [3] as it improves the detection of hearing impairment among the population [7], especially after introducing a low-pass noise [12]. The evaluation of validity of pure-tone audiometry conducted in similar conditions, ie, with the use of a PC sound card and ordinary headphones [8-10] is ambiguous and depends on the adopted solutions, which include, eg, calibration, hearing threshold evaluation method, and presence or lack of a person supervising the test. In the supervised tests, the mean error concerning the determination of a hearing threshold on a specially constructed and calibrated PC-based device was below 2.3dB [8]. In unsupervised tests conducted after computer calibration performed by a person with good hearing, the greatest error occurred at the frequency of 2kHz and 4kHz and was −5.6dB and −5.1dB respectively, with standard deviation of 8.29dB and 6.9dB [10]. In unsupervised tests conducted without calibration, the maximum mean difference occurred at the frequency of 500Hz and was 11.3dB [9].

The potential application of pure-tone audiometry conducted on a PC depends on the value of the measurement error. The PC-based test will not substitute the clinical pure-tone audiometry. However, it can be applied for conducting self-administered check-ups in cases of limited access to clinical devices, eg, at the general practitioner. Alternatively, it can be used as an initial telemedical examination combined with a survey, which will help determine the direction of future treatment. The aim of the research was to identify the measurement error connected with determining the hearing threshold conducted by means of self-administered Web-based pure tone audiometry, as well as identify and analyze factors affecting its value.

## Methods

The hearing threshold evaluation was done in three series: (1) audiologist-performed tests on clinical audiometer at an audiology outpatient clinic, (2) self-administered Web-based tests on a specially calibrated computer at an audiology outpatient clinic under the supervision of an audiologist, and (3) self-administered Web-based tests conducted at home.

In series 1 and 2, 51 participants (32 men, 19 women), aged 11-60 years (the median age was 34) underwent examination. The research participants were selected from among the patients of an audiology outpatient clinic. The qualification criterion was the willingness to participate in the research, owning a PC, basic skills to operate it, and having an email account. 102 ears were examined from which 45 (44%) were ears without a hearing loss, ie, with the hearing threshold of 25dBHL and less, 17 (16.7%) were ears with hearing impairment below 40dBHL, 31 (30.4%) with hearing impairment above 40dBHL and below 70dBHL, and 9 (8.8%) were above 70dBHL. From the group of 51 patients examined in the first two series, series 3 was self-administered at home by 37 subjects (73%). The examinations in series 1 and 2 were conducted on the same day, and examinations from series 3 were conducted up to 196 days later (median of 9 days). In the case of failure to conduct test in series 3 in a time of 2-3 weeks, the patients were reminded by a phone call and then after about 1 month, they were reminded by an email.

### Tests in Clinical Settings

All tests from series 1 were made in an acoustic cabin with the use of clinical audiometer Interacoustic AD229e and headphones TDH-39. The calibration of the audiometer was conducted according to ISO 389-1:1998. The hearing threshold in pure-tone audiometry was determined with the use of the ascending method, according to ISO 8253-1:2010. The level of the tone was reduced in 10dB steps until no further response occurred, and then it was increased in 5dB steps until the subject responded. The threshold was defined as the lowest level at which responses occurred in at least half of the series of ascending trials with a minimum of two responses required at that level [13]. The examinations were conducted at the frequencies of 250Hz, 500Hz, 1kH, 2kHz, 4kHz, and 8kHz [13].

### Self-administered Web-Based Test

The examinations in series 2 and 3 were conducted on personal computers following their calibration. Both the calibration and the examination were performed at system volume set at the maximum level. The calibration consisted of determining the reference sound level by a person with good hearing. During calibration, two sounds differing in intensity were presented bilaterally in turns for 1 second at a frequency of 1kHz. The difference in sound intensity was stable and equalled 5dB. The task of the reference person was to set the volume in a way that ensured that only the louder of the two sounds presented was audible. The volume was controlled by means of a slider with the step of 1dB. The reference person could listen to the sound for unlimited time, adjusting the volume using the slider with 1dB step any number of times, in order to finally confirm the selected level with a button. The mean intensity of the two presented sounds determined the hearing level of the reference person and was marked 0dBRP-HL. The reference values at frequencies differing from the calibration frequency, ie, 1kHz, were calculated using the model based on the A-weighting filter [14].

The procedure of determining the hearing threshold during the test, both in series 2 and 3, differed from that applied during calibration. The task of the subject was to set the intensity of the presented pure tone modulated by sinusoidal envelope with a 2-second period at such a level that the sound was on the verge of audibility. Patients could listen to the sound for unlimited time and change the intensity of the presented sound signal by themselves using a slider with 5dB step any number of times, and then confirm the chosen level with a button. The hearing threshold was determined at frequencies as in series 1, ie, 250Hz, 500Hz, 1kHz, 2kHz, 4kHz, and 8kHz. For hearing thresholds below 0dBHL-RP, the value was assumed to be 0dBHL-RP. The time spent on calibrating and examinations was not recorded.

All the examinations in series 2 were conducted on a notebook Dell Vostro 1320 with operational system Microsoft Windows

7 and headphones Technics RP-F290 placed in an acoustic cabin of an audiology outpatient clinic and connected to the Internet. The calibration was repeated six times by 3 people whose hearing threshold in pure-tone audiometry at the calibration frequency was 0dBHL. Each person performed calibration twice. The final value of the calibration coefficient was determined as the mean of all values obtained from single calibrations and was the same for all examinations in series 2. Examinations were supervised by an audiologist, whose task was to train a patient and detect hisher mistakes, eg, changing the sides of headphones, omission of the frequency, or accidental marking of a threshold that was significantly different from the actual one.

Examinations in series 3 were conducted by the patients themselves on their own personal computers at home. Each examination was conducted on a different computer using different headphones. The test station was calibrated by other household members with no previous hearing problems. In that series, the hearing threshold of the reference person was not controlled. The trial participants were instructed to conduct the test at home in a quiet place, preferably in the evening or at night. Moreover, they were informed that the calibration should be performed by a person up to 35 years old, with no previous hearing problems.

## Results

### Comparison of the Results Between Series

The hearing threshold determined in series 1 was compared to hearing thresholds from series 2 and 3. If there was no response at a given frequency, measurement was not taken into account for further calculations. Figure 1 presents relationships between the series: the relationship between hearing thresholds with division into frequencies, the relationship between total hearing thresholds, and that between the mean hearing thresholds calculated on the basis of the value obtained at all the examined frequencies.

The mean difference between thresholds in series, its standard deviation, Pearson's correlation coefficient, and linear estimators (Table 1) were calculated for the relationships described above. Linear estimators were determined for Deming's regression because the explanatory variable, which constitutes the hearing threshold in series 1, is also burdened with measurement error. The mean difference in the hearing threshold between series 1 and 2 and between series 1 and 3 was −1.54dB and −1.34dB respectively, with standard deviation 7.88dB and 10.66dB respectively, and Pearson's correlation coefficients .90 and .84 respectively. In both comparisons, the lowest values of Pearson's correlation coefficient were obtained for the frequency of 250Hz at the level of .88 and .69 respectively. The highest value of the standard deviation occurred at the frequency of 8kHz (8.88dB) and 500Hz (12.03dB) respectively. Pearson's correlation coefficients calculated for the mean threshold were at the level of .94 and .89 respectively (Table 1).

The mean difference between thresholds in series, its standard deviation, Pearson's correlation coefficient, and linear estimators of Deming's regression were determined for the hearing loss below 40dBHL as well as greater than or equal to 40dBHL (Table 2). The division was made on the basis of the mean value of the hearing threshold in two comparable series. The mean difference in the hearing threshold did not exceed 2dB in any of the groups, and its standard deviation increases together with the increase in the hearing loss. Pearson's correlation coefficients reflect changes in the standard deviation and reach low values due to narrow ranges of random variables compared to their standard deviations. For the same reason, the values of linear estimators deviate from the values established for the whole group.

On the basis of the obtained results, the sensitivity and specificity of noise-induced hearing loss detection was calculated, according to the criteria proposed in [15] adopted for the purposes of this paper. The noise-induced hearing loss was detected when the hearing threshold exceeded 30dB at one of the following frequencies: 500Hz, 1kHz, 2kHz, or 25dB at more than one, or when the hearing threshold exceeded 50dB at 4Hz. For series 2, the sensitivity was 0.92 with confidence interval of (0.81, 1.0) at $P$=.05, and the specificity was 0.96 with confidence intervals of (0.88, 1.0) at $P$=.05. For series 3, the respective values were sensitivity 0.89 (0.74, 1.0) and specificity 0.89 (0.76, 1.0).

### Analysis of the Measurement Error

This paper attempts to identify and assess the values of the sources of error in determining the hearing threshold of Web-based examinations. The literature data on standard deviation of the hearing threshold determined in test-retest examinations carried out in clinical settings are presented in Table 3. On the basis of these data, the value of the standard deviation of the hearing threshold determined in test-retest examinations was adopted for the further calculations at the level of 6dB.

**Table 1.** The mean difference *m* in the hearing threshold value between the series calculated on the basis of n data points collected on a group of N subjects, its standard deviation σ, Pearson's correlation coefficient *r* and linear estimators *a*, *b* of Deming's regression *y=ax+b* and corresponding confidence intervals CI at *P*=.05.

| | f | n | *m* | σ | *r* (CI) | *a* (CI) | *b* (CI) |
|---|---|---|---|---|---|---|---|
| **Series 1 vs 2, N=51** | | | | | | | |
| | 250Hz | 100 | −3.70 | 7.30 | .88 (0.82, 0.92) | 1.08 (1.06, 1.09) | −5.49 (−5.84, −5.13) |
| | 500Hz | 101 | 0.52 | 7.20 | .88 (0.83, 0.92) | 0.93 (0.92, 0.94) | 2.09 (1.83, 2.34) |
| | 1kHz | 101 | −4.59 | 6.91 | .92 (0.88, 0.94) | 0.85 (0.84, 0.86) | −1.16 (−1.39, −0.92) |
| | 2kHz | 101 | −2.42 | 7.22 | .93 (0.90, 0.95) | 0.95 (0.94, 0.96) | −1.18 (−1.44, −0.91) |
| | 4kHz | 100 | 2.15 | 7.68 | .93 (0.90, 0.95) | 0.97 (0.96, 0.98) | 2.88 (2.64, 3.12) |
| | 8kHz | 100 | −1.20 | 8.88 | .91 (0.86, 0.94) | 0.93 (0.92, 0.93) | 0.71 (0.47, 0.95) |
| | Total | 603 | −1.54 | 7.88 | .90 (0.89, 0.92) | 0.96 (0.96, 0.96) | −0.52 (−0.57, −0.48) |
| | Mean | 99 | −1.52 | 5.42 | .94 (0.91, 0.96) | 0.94 (0.93, 0.96) | −0.23 (−0.46, −0.01) |
| **Series 1 vs 3, N=37** | | | | | | | |
| | 250Hz | 71 | −5.07 | 12.03 | .69 (0.55, 0.80) | 1.23 (1.19, 1.26) | −10.42 (−11.14, −9.70) |
| | 500Hz | 71 | −0.35 | 10.43 | .79 (0.68, 0.86) | 1.05 (1.03, 1.08) | −1.57 (−2.08, −1.05) |
| | 1kHz | 72 | −1.24 | 9.28 | .85 (0.76, 0.90) | 1.01 (0.99, 1.03) | −1.49 (−1.96, −1.02) |
| | 2kHz | 72 | −1.58 | 10.46 | .88 (0.81, 0.92) | 0.90 (0.87, 0.92) | 1.04 (0.48, 1.59) |
| | 4kHz | 72 | −1.42 | 10.77 | .87 (0.80, 0.92) | 0.98 (0.96, 1.00) | −0.83 (−1.33, −0.33) |
| | 8kHz | 71 | 1.63 | 10.05 | .88 (0.82, 0.93) | 1.05 (1.04, 1.07) | 0.31 (−0.11, 0.73) |
| | Total | 429 | −1.34 | 10.66 | .84 (0.81, 0.86) | 1.02 (1.01, 1.02) | −1.71 (−1.80, −1.62) |
| | Mean | 68 | −1.44 | 7.59 | .89 (0.83, 0.93) | 1.01 (1.00, 1.03) | −1.73 (−2.14, −1.33) |

**Table 2.** The mean difference *m* in the hearing threshold *t* between the series calculated on the basis of n data points collected on a group of N subjects, its standard deviation σ, Pearson's correlation coefficient *r* and linear estimators *a*, *b* of Deming's regression *y=ax+b* and corresponding confidence intervals CI at *P*=.05 for the hearing threshold below 40dBHL as well as greater than or equal to 40dBHL.

| Group | | n | *m* | σ | *r* (CI) | *a* (CI) | *b* (CI) |
|---|---|---|---|---|---|---|---|
| **Series 1 vs 2, N=51** | | | | | | | |
| | *t*<40dBHL | 498 | −1.45 | 7.37 | .74 (0.70, 0.78) | 0.84 (0.84, 0.84) | 1.24 (1.19, 1.30) |
| | *t*≥40dBHL | 105 | −1.99 | 9.88 | .64 (0.51, 0.74) | 1.07 (1.04, 1.09) | −5.67 (−7.04, −4.30) |
| **Series 1 vs 3, N=37** | | | | | | | |
| | *t*<40dBHL | 339 | −1.52 | 9.72 | .58 (0.51, 0.65) | 1.11 (1.10, 1.12) | −3.38 (−3.53, −3.22) |
| | *t*≥40dBHL | 90 | −0.64 | 13.67 | .44 (0.25, 0.59) | 0.66 (0.62, 0.69) | 17.59 (15.56, 19.62) |

**Figure 1.** The relationship between hearing thresholds in series 1 and 2 (first two columns on left) and between series 1 and 3 (last two columns on right) calculated in a group of 51 and 37 subjects, respectively.

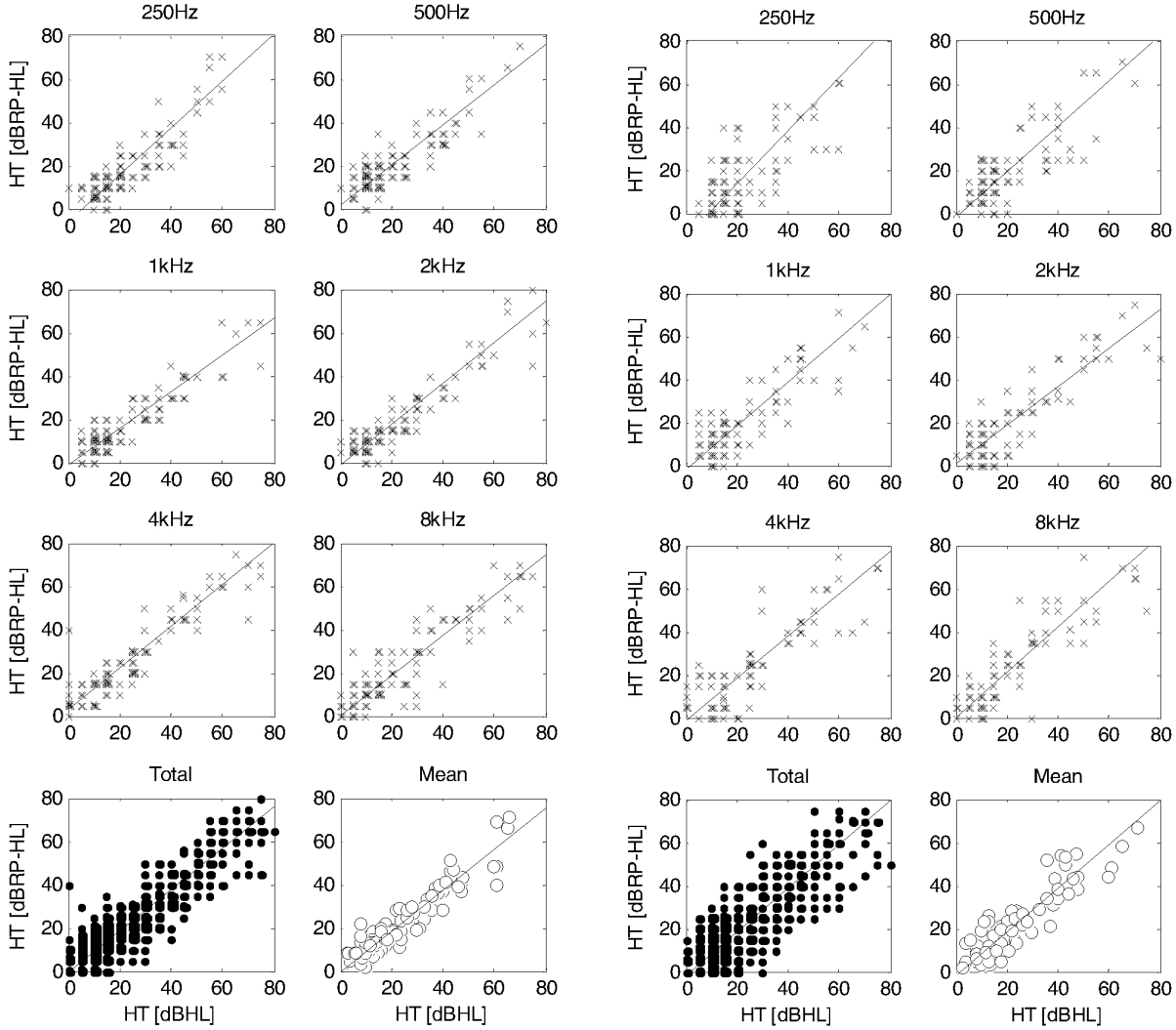



**Table 3.** Standard deviations of the hearing threshold obtained in test-retest examination (literature data) (separate measurements were carried out for supra-aural headphones (series I, II, III) and in-ear headphones (IV, V, VI) and for the group of young (series I and IV), older (II, V) and the oldest (III, VI) subjects; N=number of subjects) [16-18].

| Investigator | Series | N | 250 | 500 | 1kHz | 2kHz | 4kHz | 8kHz | Mean |
|---|---|---|---|---|---|---|---|---|---|
| Brown 1948 [16] | | 30 | 5.3 | 4.4 | 5.8 | 4.5 | 5.4 | 7.1 | 5.4 |
| Erlandsson et al 1979 [17][a] | | 10 | | 6.3 | 6.2 | 4.7 | 5.3 | 10.4 | 6.6 |
| Landry et al 1999 [18] | I | 20 | 5.0 | 4.6 | 5.2 | 4.1 | 4.3 | 6.0 | 5.9 |
| | II | 10 | 8.5 | 7.9 | 5.2 | 6.4 | 9.5 | 6.9 | |
| | III | 10 | 5.9 | 5.4 | 6.4 | 2.1 | 6.7 | 7.1 | |
| | IV | 20 | 8.8 | 7.7 | 4.6 | 5.5 | 2.9 | 3.6 | |
| | V | 10 | 7.1 | 4.4 | 7.4 | 5.5 | 2.6 | 5.4 | |
| | VI | 10 | 9.3 | 4.6 | 3.3 | 5.3 | 9.0 | 9.4 | |
| Mean | | | 7.1 | 5.7 | 5.5 | 4.8 | 5.7 | 7.0 | 6.0 |

[a]Estimated by comparison with Békésy audiometry performed 5 times on a group of 10 subjects.

Knowing that the variance of the sum of the two independent random variables $X$ and $Y$ is the sum of their variance (see Multimedia Appendix 1), the standard deviation for a single measurement of the hearing threshold in clinical settings $\sigma_{clin}$ was calculated at the level of 4.25dB by dividing the $\sigma_{2clin}$ by the square root of two (see Multimedia Appendix 2).

Therefore, assuming the variability of the hearing threshold measurement in series 1 $\sigma_{(i)}$ at the level of literature data $\sigma_{clin}$, the standard deviation of the measurement error in Web-based hearing tests for series 2 $\sigma_{(ii)}$ was calculated at 6.64dB based on the value $\sigma_{(i)}$ and the value $\sigma_{(i)(ii)}$, which is the standard deviation of the hearing threshold difference between the values determined in series 1 and 2. Similarly, for series 3, we obtain the standard deviation $\sigma_{(iii)}$ equal to 9.78dB (see Multimedia Appendix 3).

In series 2, the standard deviation of the measurement error depends mainly on the standard deviation of the error connected with the procedure of determining threshold value $\sigma_{proc}$. In series 3, apart from the error connected with the procedure of determining the threshold value, we can distinguish other sources of error influencing the standard deviation: the standard deviation of the calibration error $\sigma_{cal}$, which is the reference sound level evaluation error at the calibration frequency, the frequency nonlinearity error $\sigma_{nonlin}$, which is the difference between the actual reference sound level and that set by the model, the gain error $\sigma_{gain}$, and the error connected with background noise $\sigma_{noise}$ (see Multimedia Appendix 4).

Of course in series 2, the calibration error, frequency nonlinearity error, and gain error were the same in all tests and did not influence the value of standard deviation, while the error connected with background noise can be omitted since the tests were conducted in an acoustic cabin.

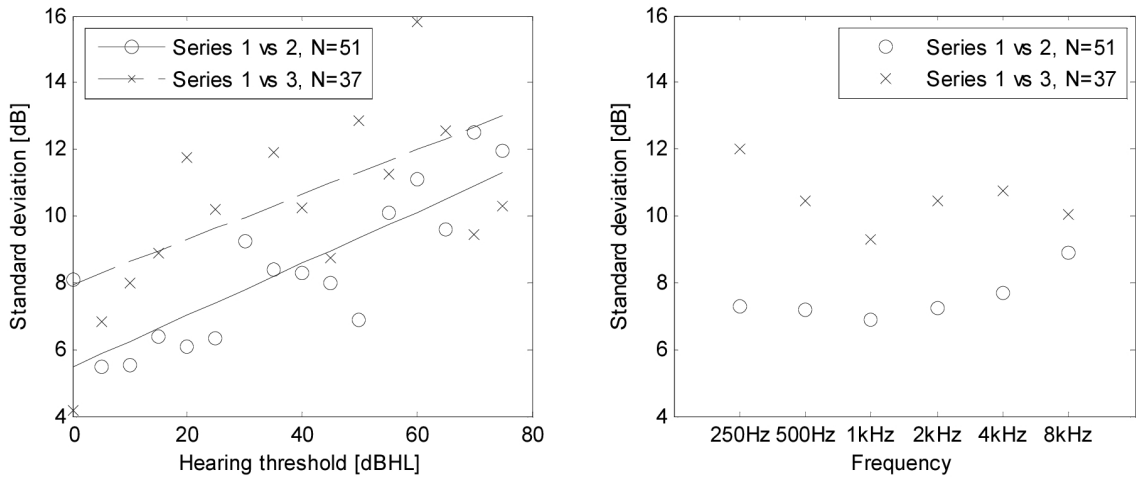
Figure 2 shows the standard deviation of threshold differences between series in relation to the hearing threshold for series 1 and the frequency. The lowest values of standard deviations are observed in small hearing losses, ie, for those measurements that should be most affected by the background noise. Therefore, we can assume that the error connected with background noise is significantly smaller than other components of the measurement error. In the following considerations, the standard deviation of error connected with background noise was omitted ($\sigma_{noise}$≈0dB), both for series 2 and for series 3.

With the increase in intensity, increment of the standard deviation of measurement error in series 3 is slightly smaller than in series 2. In series 2, the standard deviation of the gain error equals zero since all tests are done on the same set. The lack of increase in standard deviation in series 3 above the values observed in series 2 indicates a negligible effect of the gain error on the measurement values ($\sigma_{gain}$≈0dB).

The standard deviation of the calibration error $\sigma_{cal}$ can be estimated by comparing the standard deviations for series 2 and 3 at the calibration frequency, ie, 1kHz. Since the calibration is performed on the same set as the test, the error connected with the nonlinearity of the frequency response at the calibration frequency is zero ($_{1kHz}\sigma_{nonlin}$=0dB). Taking into account the above, we obtain the calibration error $\sigma_{cal}$ equal to 6.19dB (see Multimedia Appendix 5). Finally, knowing the calibration error $\sigma_{cal}$ and assuming the insignificance of the error connected with background noise as well as the gain error, the standard deviations of the frequency nonlinearity error were calculated. The highest value $_{250Hz}\sigma_{nonlin}$ was observed at the frequency of 250Hz at the level of 7.28dB (see Multimedia Appendix 5).

No asymmetry of measurement error between the left and right sides was detected in any of the tests, which could raise suspicion concerning incorrect audio balance.

**Figure 2.** Standard deviation of the hearing threshold differences between the series calculated for the group of N subjects in relation to the hearing threshold evaluated in series 1 (left) and the frequency (right).




## Discussion

The aim of the research was to identify the measurement error connected with determining the hearing threshold conducted by means of self-administered Web-based pure tone audiometry, as well as identify and analyze factors affecting its value.

The average difference between the value of the hearing threshold determined in series 1 and in series 2 was −1.54dB with standard deviation of 7.88dB and Pearson's correlation

coefficient of .90. Between series 1 and series 3, these values were −1.35dB±10.66dB and .84 respectively. For the average hearing threshold, these values were, for series 2 −1.52dB±5.42dB and .94 respectively, and for series 3, −1.44dB±7.59dB and .89. The results from series 2 are consistent with the results presented by Honeth et al, who determined the correlation coefficient of the average hearing threshold for the group of 72 people at the level of .94 and .93 respectively for the right and left ears [10]. 64 out of 72 people (89%) were examined on the same test set calibrated by the same person.

In series 3, the standard deviation values were most influenced by the error connected with the procedure of determining the hearing threshold, the same as the error in series 2 ($\sigma_{proc}$=6.64dB), the calibration error ($\sigma_{cal}$=6.19dB), and additionally at the frequency of 250Hz, the frequency response nonlinearity error ($_{250Hz}\sigma_{nonlin}$=7.28dB).

The error of determining the hearing threshold during examination, which consisted of self-adjustment of the position of the volume slider in 5dB steps, was estimated at the level of $\sigma_{proc}$=6.64dB. The reason for such high error value compared to the error value of the ascending method ($\sigma_{clin}$=4.25dB) may be attributed to the difficulty connected with self-assessment of the hearing threshold. Standard deviations between series 1 and 2 as well as between series 1 and 3 increase together with the increase in the hearing loss (Figure 2). This suggests that the task of self-evaluation of the hearing threshold proves more difficult to perform for subjects with moderate and profound hearing impairment compared to normal hearing subjects. Replacement of this method with the ascending method will contribute to improving the accuracy of the examination. The method of self-evaluation of the hearing threshold was originally chosen as simpler, faster and more flexible. Consequently, it was less monotonous and more attractive for the patient.

The purpose in the calibration was to approximate the 0dBHL level at the examined frequencies. The approximation can be conducted in a number of ways. The reference sound level can be determined independently for each frequency in relation to the hearing threshold of the reference person [10]. In this case, the calibration error at each frequency is burdened with measurement error of the hearing threshold and the difference between the hearing threshold of the reference person and 0dBHL. Another solution, which was adopted in this paper, is to measure the hearing threshold at one frequency, at which differences between the values of the hearing threshold among the population are the smallest and then determine the reference sound level at other frequencies on the basis of the model. Alternatively, calibration can be performed at a number of frequencies and the correctness of the calibration may be assessed based on its accordance with the model. At the same time, in the case of a large discrepancy between the measurement results, the calibration can be repeated by another person or on different headphones. The choice of the optimal method requires further research.

The calibration also involves the problem of selecting a procedure for determining the hearing threshold of the reference person. Calibration is usually conducted by young persons with normal hearing. Therefore, in order to achieve more accurate calibration results, one can use the procedure requiring greater hearing efficiency compared to the procedure used in examination. This study applies calibration consisting in changing the intensity of two sounds with the stable difference of 5dB. This method requires considerable concentration and optimal hearing ability. However, in contrast to the ascending method, it is not burdened with discretization errors.

Calibration error ($\sigma_{cal}$=6.19dB) is connected with the error of hearing threshold assessment $\sigma_{proc\_cal}$ and the standard deviation in the hearing threshold of the reference persons $\sigma_{pop}$ at the calibration frequency, ie, 1kHz. The standard deviation of the hearing threshold determined by means of the ascending method among the population of young subjects without prior hearing problems, $\sigma_{pop\_asc}$, can be estimated on the basis of literature data [19]. The distribution of the hearing threshold at 1kHz determined for the population of 2490 subjects was approximated with normal distribution. The standard deviation was calculated based on the values of 1st and 3rd quartile at the level of $\sigma_{pop\_asc}$=5.6dB. Assuming the standard deviation of the ascending method $\sigma_{clin}$ is at the level of 4.25dB, we obtain $\sigma_{pop}$ at 3.65dB and then $\sigma_{proc\_cal}$ at 5.00dB (see Multimedia Appendix 6).

The method of the hearing threshold evaluation used during calibration ($\sigma_{proc\_cal}$=5.00dB), despite the possibility of setting the threshold with the accuracy of 1dB step, is characterized by a slightly higher error than in the ascending method $\sigma_{clin}$=4.25dB, whose step equals 5dB. Therefore, accuracy improvement of the Web-based examination can be achieved by modifying the method. An interesting solution is offered by, eg, the Békésy's method, as it is characterized by lower standard deviation in the test-retest examination compared to the ascending method [17], and additionally it is not burdened with discretization errors.

The frequency nonlinearity error is the difference between the actual values of the reference sound level and the values determined by the model. The greatest value of $_{250Hz}\sigma_{nonlin}$=7.28dB was observed at 250Hz with the mean value calculated for all frequencies, excluding the calibration frequency $_{mean}\sigma_{nonlin}$=4.05dB. The frequency nonlinearity error is connected with different values of Reference Equivalent Threshold Sound Pressure Level (RETSPL) of the headphones used in the examination and the frequency nonlinearity of the sound card. The biggest differences between the RETSPL values occur at low frequencies [20], also in the low frequency range, more often than at mid and high frequencies, frequency distortion of the sound card can be observed, eg, when bass boost option or equalizer settings are enabled. The improvement in accuracy, especially at low frequencies, can be acquired through selection of headphones for which RETSPL values are known. However, in practice this will be hard to achieve for tests carried out at home. An alternative solution would be control determination of the calibration coefficient at low frequencies and verification against the value determined by the model.

Taking into account the significant value of the calibration error and frequency nonlinearity error, it seems interesting to conduct pure-tone audiometry examination on generally available appliances with known electro-acoustic parameters, eg, smartphones. On determining frequency characteristics of the selected smartphone model with bundled headphones, it seems possible to omit the calibration stage and thus obtain more precise examination results.

Web-based pure-tone audiometry can be used without previous training in its conducting. However, it requires the knowledge of basic terms such as the hearing threshold or frequency. On the other hand, an attractive, intuitive and user-friendly interface can largely replace training. Prior to the tests in series 3, patients were instructed on how to perform the test and had performed similar tests before in series 2 under the supervision of an audiologist. The knowledge of the application could lead to reduction in the value of the measurement error.

The obtained results of sensitivity and specificity confirm the possibility of applying Web-based pure-tone audiometry in screening tests. Moreover, Web-based pure-tone audiometry may be used for self-monitoring of hearing, especially if tests are to be conducted under the same conditions. If the same equipment is used, the relative error between subsequent examinations will be reduced by frequency nonlinearity error, and in the case of the same calibration coefficients, relative error will be reduced by the calibration error. Self-monitoring of hearing may become applicable in hearing disorders, such as fluctuating hearing loss, tinnitus, sudden hearing loss, otosclerosis, Ménière's disease, as well as during treatment with ototoxic drugs. In the future, modifications of the method leading to the decrease in measurement error can broaden the scope of Web-based pure-tone audiometry application.

## Acknowledgments

The trial was financed by Wroclaw Medical University as part of Research Grant for Young Scientists. Grant No. 38/Pbmn.

## Conflicts of Interest

The author of this article is the owner of the Internet portal, e-audiologia.pl, on the basis of which the tests above were conducted.

## Multimedia Appendix 1

Variance of the sum of the two independent random variables.

[PNG File, 4KB - jmir_v15i4e71_app1.png ]

## Multimedia Appendix 2

The standard deviation for a single measurement of the hearing threshold determined by means of the ascending method.

[PNG File, 69KB - jmir_v15i4e71_app2.png ]

## Multimedia Appendix 3

The standard deviation for a single measurement of the hearing threshold determined in series 2 and series 3.

[PNG File, 101KB - jmir_v15i4e71_app3.png ]

## Multimedia Appendix 4

Factors influencing the standard deviation of the measurement error of the hearing threshold in the series 2 and 3.

[PNG File, 145KB - jmir_v15i4e71_app4.png ]

## Multimedia Appendix 5

The standard deviation of the calibration error and frequency nonlinearity error.

[PNG File, 217KB - jmir_v15i4e71_app5.png ]

## Multimedia Appendix 6

The standard deviation of the hearing evaluation method used during calibration.

[PNG File, 136KB - jmir_v15i4e71_app6.png ]

## Multimedia Appendix 7

CONSORT-EHEALTH checklist V1.6.2 [21].

[PDF File (Adobe PDF File), 994KB - jmir_v15i4e71_app7.pdf ]

## References


1. Swanepoel de W, Hall JW. A systematic review of telehealth applications in audiology. Telemed J E Health 2010 Mar;16(2):181-200. [doi: 10.1089/tmj.2009.0111] [Medline: 20187743]
2. Chung JH, Des Roches CM, Meunier J, Eavey RD. Evaluation of noise-induced hearing loss in young people using a Web-based survey technique. Pediatrics 2005 Apr;115(4):861-867 [FREE Full text] [doi: 10.1542/peds.2004-0173] [Medline: 15805356]
3. Smits C, Kramer SE, Houtgast T. Speech reception thresholds in noise and self-reported hearing disability in a general adult population. Ear Hear 2006 Oct;27(5):538-549. [doi: 10.1097/01.aud.0000233917.72551.cf] [Medline: 16957503]
4. Bexelius C, Honeth L, Ekman A, Eriksson M, Sandin S, Bagger-Sjöbäck D, et al. Evaluation of an internet-based hearing test--comparison with established methods for detection of hearing loss. J Med Internet Res 2008;10(4):e32 [FREE Full text] [doi: 10.2196/jmir.1065] [Medline: 18940783]
5. Givens GD, Blanarovich A, Murphy T, Simmons S, Blach D, Elangovan S. Internet-based tele-audiometry system for the assessment of hearing: a pilot study. Telemed J E Health 2003;9(4):375-378. [doi: 10.1089/153056203772744707] [Medline: 14980095]
6. Swanepoel de W, Mngemane S, Molemong S, Mkwanazi H, Tutshini S. Hearing assessment-reliability, accuracy, and efficiency of automated audiometry. Telemed J E Health 2010 Jun;16(5):557-563. [doi: 10.1089/tmj.2009.0143] [Medline: 20575723]
7. Smits C, Merkus P, Houtgast T. How we do it: The Dutch functional hearing-screening tests by telephone and internet. Clin Otolaryngol 2006 Oct;31(5):436-440. [doi: 10.1111/j.1749-4486.2006.01195.x] [Medline: 17014457]
8. Choi JM, Lee HB, Park CS, Oh SH, Park KS. PC-based tele-audiometry. Telemed J E Health 2007 Oct;13(5):501-508. [doi: 10.1089/tmj.2007.0085] [Medline: 17999612]
9. Kimball SH. Inquiry into online hearing test raises doubts about its validity. Hearing Journal 2008 Mar;61(3):38-46. [doi: 10.1097/01.HJ.0000314717.42321.4a]
10. Honeth L, Bexelius C, Eriksson M, Sandin S, Litton JE, Rosenhall U, et al. An internet-based hearing test for simple audiometry in nonclinical settings: preliminary validation and proof of principle. Otol Neurotol 2010 Jul;31(5):708-714. [doi: 10.1097/MAO.0b013e3181de467a] [Medline: 20458255]
11. Leensen MC, de Laat JA, Dreschler WA. Speech-in-noise screening tests by internet, part 1: test evaluation for noise-induced hearing loss identification. Int J Audiol 2011 Nov;50(11):823-834. [doi: 10.3109/14992027.2011.595016] [Medline: 21988504]
12. Leensen MC, de Laat JA, Snik AF, Dreschler WA. Speech-in-noise screening tests by internet, part 2: improving test sensitivity for noise-induced hearing loss. Int J Audiol 2011 Nov;50(11):835-848. [doi: 10.3109/14992027.2011.595017] [Medline: 21970351]
13. British Society of Audiology. Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking. 2011. Recommended Procedure URL: http://www.thebsa.org.uk/docs/Guidelines/BSA_RP_PTA_FINAL_24Sept11.pdf [accessed 2012-05-27] [WebCite Cache ID 67yJElAP1]
14. International Electrotechnical Commission. Electroacoustics - Sound level meters - Part 2: Pattern evaluation tests. In: International Standard IEC 61672-2:2003(E). Geneva: IEC; 2003.
15. Klockhoff I, Drettner B, Hagelin KW, Lindholm L. A method for computerized classification of pure tone screening audiometry results in noise-exposed groups. Acta Otolaryngol 1973 Apr;75(4):339-340. [Medline: 4702637]
16. Brown RE. Experimental studies on the reliability of audiometry. J Laryngol Otol 1948 Aug;62(8):487-524. [Medline: 18876519]
17. Erlandsson B, Håkanson H, Ivarsson A, Nilsson P. Comparison of the hearing threshold measured by manual pure-tone and by self-recording (Békésy) audiometry. Audiology 1979;18(5):414-429. [Medline: 496724]
18. Landry JA, Green WB. Pure-Tone Audiometric Threshold Test-Retest Variability in Young and Elderly Adults. J Speech Lang Pathol Audiol 1999;23(2):74-80 [FREE Full text]
19. Engdahl B, Tambs K, Borchgrevink HM, Hoffman HJ. Screened and unscreened hearing threshold levels for the adult population: results from the Nord-Trøndelag Hearing Loss Study. Int J Audiol 2005 Apr;44(4):213-230. [Medline: 16011050]
20. Yost YA, Killion MC. Hearing thresholds. In: Crocker MJ, editor. Encyclopedia of acoustics. New York: John Wiley; 1997:1545-1553.
21. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

## Abbreviations

**dBHL:** decibel hearing level
**RETSPL:** Reference Equivalent Threshold Sound Pressure Level

*Edited by G Eysenbach; submitted 29.06.12; peer-reviewed by M Eriksson; comments to author 26.09.12; revised version received 04.11.12; accepted 09.02.13; published 12.04.13.*

*Please cite as:*
*Masalski M, Kręcicki T*
*Self-Test Web-Based Pure-Tone Audiometry: Validity Evaluation and Measurement Error Analysis*
*J Med Internet Res 2013;15(4):e71*
*URL: http://www.jmir.org/2013/4/e71/*
*doi:10.2196/jmir.2222*
*PMID:23583917*

Original Paper

# Attitudes of Patients Toward Adoption of 3D Technology in Pain Assessment: Qualitative Perspective

Fotios Spyridonis[1*], PhD; Gheorghita Ghinea[1*], PhD; Andrew O Frank[2*], FRCP

[1]Brunel University, Department of Information Systems and Computing, Uxbridge, United Kingdom

[2]Brunel University, Centre for Research in Rehabilitation, School of Health Studies and Social Care, Uxbridge, United Kingdom

[*]all authors contributed equally

**Corresponding Author:**
Fotios Spyridonis, PhD
Brunel University
Department of Information Systems and Computing
Kingston Lane
Uxbridge, UB8 3PH
United Kingdom
Phone: 44 1895265503
Fax: 44 1895251686
Email: fotios.spyridonis@brunel.ac.uk

## Abstract

**Background:** Past research has revealed that insufficient pain assessment could, and often, has negative implications on the provision of quality health care. While current available clinical approaches have proven to be valid interventions, they are expensive and can often fail in providing efficient pain measurements. The increase in the prevalence of pain calls for more intuitive pain assessment solutions. Computerized alternatives have already been proposed both in the literature and in commerce, but may lack essential qualities such as accuracy of the collected clinical information and effective patient-clinician interaction. In response to this concern, 3-dimensional (3D) technology could become the innovative intervention needed to support and improve the pain assessment process.

**Objective:** The purpose of this analysis was to describe qualitative findings from a study which was designed to explore patients' perceptions of adopting 3D technology in the assessment of their pain experience related to important themes that might positively or negatively influence the quality of the pain assessment process.

**Methods:** The perceptions of 60 individuals with some form of pain in the area of Greater London were collected through semi-structured interviews. Of the 60 respondents, 24 (43%) produced usable responses and were analyzed for content using principles of the grounded theory approach and thematic analysis, in order to gain insight into the participants' beliefs and attitudes towards adopting 3D technology in pain assessment.

**Results:** The analysis identified 4 high-level core themes that were representative of the participants' responses. These themes indicated that most respondents valued "the potential of 3D technology to facilitate better assessment of pain" as the most useful outcome of adopting a 3D approach. Respondents also expressed their opinions on the usability of the 3D approach, with no important concerns reported about its perceived ease of use. Our findings finally, showed that respondents appreciated the perceived clinical utility of the proposed approach, which could further have an influence on their intention to use it.

**Conclusions:** These findings highlighted factors that are seen as essential for improving the assessment of pain, and demonstrated the need for a strong focus on patient-clinician communication. The participants of this analysis believed that the introduction of 3D technology in the process might be a useful mechanism for such a positive health care outcome. The study's findings could also be used to make recommendations concerning the potential for inclusion of 3D technology in current clinical pain tools for the purpose of improving the quality of health care.

***(J Med Internet Res 2013;15(4):e55)*** doi:10.2196/jmir.2427

**KEYWORDS**

pain assessment; 3-dimensional image; health care systems; health care delivery; patient acceptance of health care; qualitative research

## Introduction

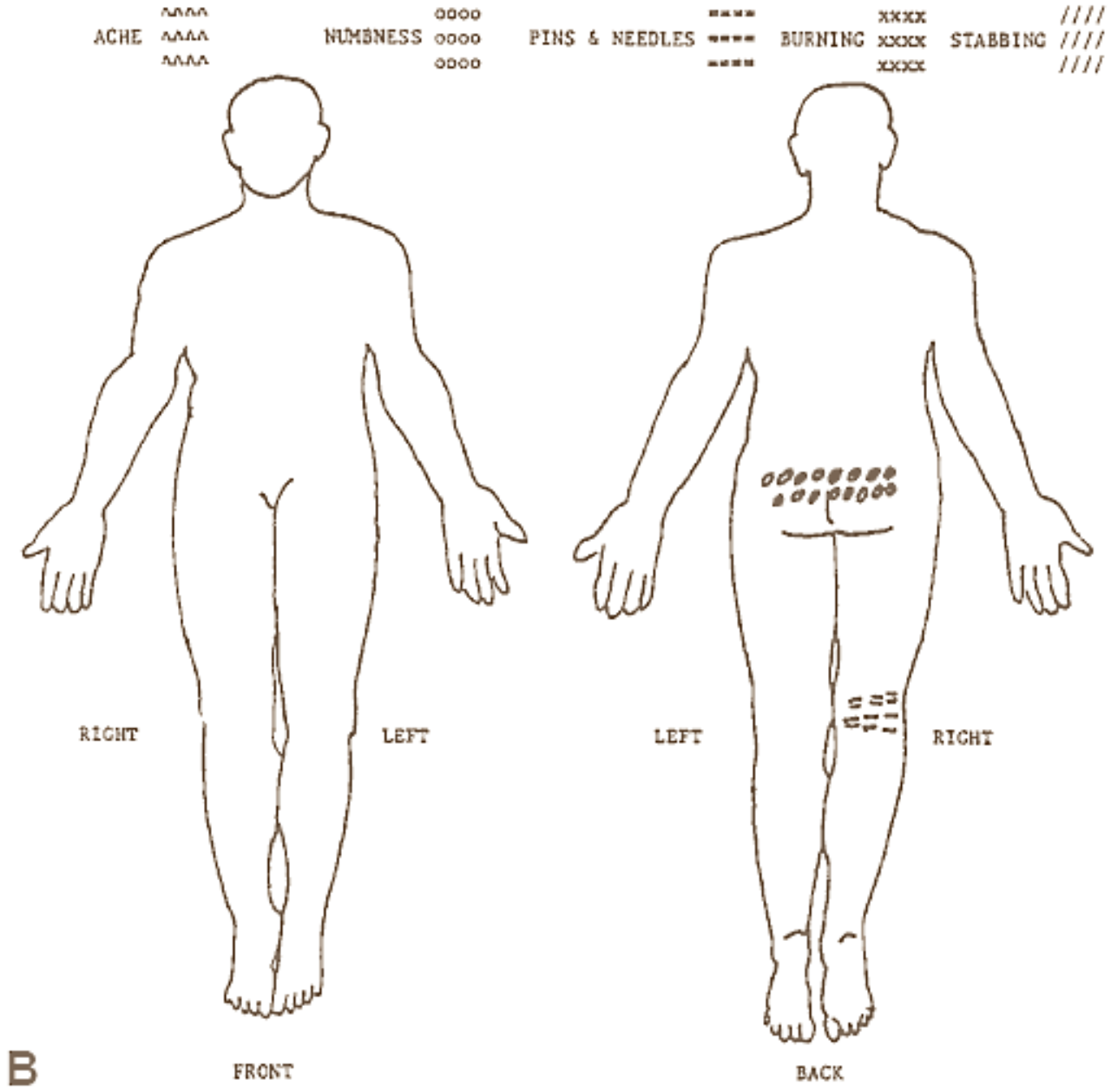
### Overview

In medical practice, high quality clinical information is essential in providing high quality health care. When assessing a patient for pain, the only valued information that can typically be used are suggestive descriptions or self-reports from the patient, which are considered as the best available source of information for pain measurement [1]. These, however, are also considered to be subjective in nature [2], as they are highly influenced by a variety of psychological and cultural factors [3]. In fact, it is not uncommon in some patients that psychological problems may have actually caused some of the pain by adding stress to the body, or the stress of the pain may have caused psychological problems [4]. As a result, the ability of a clinician to provide a wholly accurate diagnosis can be, and often is, undermined due to inaccurate pain descriptions from the patient. In response to this, clinicians in primary care have been using the "pain drawing" shown in Figure 1 as a means to collect and manage pain information of higher quality.

### The Pain Drawing in the Measurement of Pain

The pain drawing has long been applied to the field of pain documentation [5-9], offering patients the ability to self-record information about the *location* and *type* of pain they are experiencing on a paper-based diagram of a human figure. Such a topographical representation of pain is argued to be very useful in summarizing patients' description of their pain in an interpretable way for clinicians [5], and can be further used to better monitor change in patients' pain situation [10]. The patients typically mark the location of pain experienced on a blank human diagram, using either specific symbols [11] or colours [12], in order to indicate various pain types.

**Figure 1.** A completed pain drawing [24].



As a result of its potential, this tool has been employed over the years to assess a variety of pain-related conditions ranging from simple back pain to more serious medical problems. In some instances, the pain drawing was the most used tool to evaluate the location of back pain [13], and in other instances, the pain drawing was most commonly used to evaluate pain as a result of knee osteoarthritis [14] and to examine the association of depression with spinal stenosis surgery [15].

Nevertheless, the results from a pan-European consensus report [16] showed that one in five Europeans (19%) was estimated to suffer from pain, a figure that seems to indicate that there is still an overall lack of success in its assessment, despite the benefits of the pain drawing for the intended purpose. This might be an indication of underdeveloped clinical tools and limited on-going research with respect to improving them. This same report particularly highlighted that the percentage of 2019 people in 15 European countries whose pain was still not adequately managed was 38%, and suggested that this was indeed partially owed to the inappropriate and ineffective management and treatment of pain, which needs to be improved.

Accordingly, in the pain literature, recent studies similarly revealed that this situation is existent, evidencing that one of the most important pain measurement tools, the pain drawing, has indeed remained highly underdeveloped. In particular, it is noteworthy that in such studies (eg, [9,17-22] the same paper-based, 2-dimensional (2D) human diagram is still being used as the main means to report and show the location of pain that can typically occur in any part of our 3-dimensional (3D) human body.

This is certainly one of the biggest drawbacks of the pain drawing, as it implies serious concerns about its performance when it comes to accurately visualizing patients' pain descriptions. For instance, it is very common for pain to occur on the inside of a thigh, a location that is not easily captured in 2D pain drawings. It is therefore expected that a pain reporting mismatch is present in the current version of the pain drawing, resulting in (1) patients being unable to accurately report the pain that they are experiencing, and (2) making the assessment a time-consuming process with possible irrelevant medical data collected that can lead to the ineffective management of pain.

Results from a number of various studies in the area of information visualization seem to consent that 2D visualization is indeed not anymore useful for a complete understanding of the "object" under investigation, mainly because it lacks the natural depth cues (eg, perspective, shading, and occlusion) [23]. As such, notwithstanding its advantages, it is essential for new interventions for more accurate pain assessment to be developed.

### The Need for Adopting 3D Technology

With the emergence of 3D technology, the field of health care has already adopted 3D technology for a variety of uses, and it has become one of the most common methods for visualizing medical information. In the area of information technology, the 3D concept is used to describe a real or imagined environment that can be experienced visually in the 3 dimensions of width, height, and depth, and that may additionally provide an interactive experience. According to [24], 3D technology offers significant benefits over 2D, in particular: (1) displaying data in 3D can make it easier for users to manipulate the data, (2) improvement of the understanding of 3D structures when users have the ability to manipulate it, and (3) 3D makes it possible to make the layout of a designed object more consistent with its intended role and visualize it as perceived in its natural environment.

Along these lines, researchers have used 3D computer reconstructions to evaluate the pathology of a spinal cord injury [25] and to construct 3D virtual images from computerized medical scans [26]. In both examples, 3D technology was extremely beneficial because the models produced could be observed from many different viewpoints, while rotation and zooming features were combined to allow observer navigation within the tissue of interest. Such feature benefits were anticipated to provide and improve the depth cues that 2D pain drawings currently lack.

As opposed to a variety of other areas in health care, until recently, the traditional 2D pain drawings had never caught up with 3D technology, thus, it had always been lacking the benefits mentioned above. Previous work [2,27] attempted to address this issue by introducing a novel 3D pain drawing that employed the aforementioned features in the effort to provide better and more accurate measurements of pain. This new pain drawing was evaluated for its *usability* and *user satisfaction* in the self-reporting of pain by different groups of patients suffering from pain, producing very positive outcomes.

While clinicians seem to be embracing 3D technology to support a wide range of generic clinical activities, it is unclear in the literature how the category of patients suffering from pain would actually perceive the adoption of 3D technology in everyday practice for the intended purpose. To the best of the authors' knowledge, no previous studies exist that have attempted to investigate this aspect in the context of pain assessment. We were expecting that this study would produce the same positive feedback as in the usability testing studies mentioned above, and that pain sufferers would similarly embrace the benefits that the 3D pain drawing offers in supporting their everyday reporting of pain.

Accordingly, in order to address the above issues, the main aim of the analysis described in this paper was to identify and report patients' perceptions of adopting the developed 3D technological solution supporting improved assessment of pain, in the anticipation that technology adoption is best predicted by a patient's attitudes toward the technology, and perceptions about its usefulness.

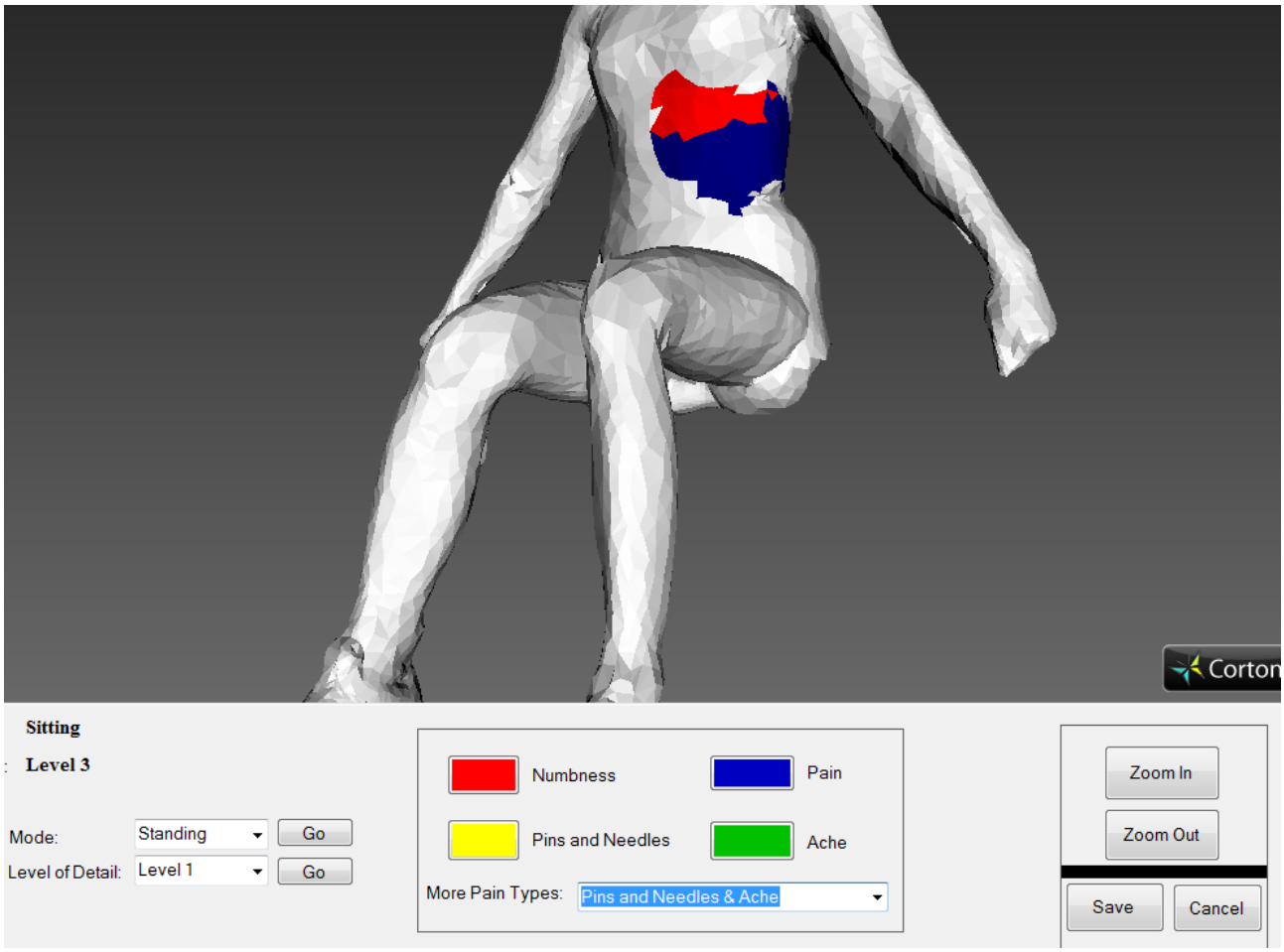
## Methods

### Design

The above aim was addressed as part of a within-subjects study that we conducted with a convenience sample of 3 different groups of patients [2,27]. Participants from each group of this study were randomly given the 3D pain drawing of Figure 2 and were asked to use it in order to report their pain experience at the moment of use. In order to record their perceptions and

understand the patients' attitudes with respect to the 3D pain drawing being adopted into everyday practice for the assessment of their pain experience, the authors employed a phenomenological approach. Ethical approval was obtained by both the Brunel University Research Ethics Committee and the North London 1 Research Ethics Committee.

**Figure 2.** Example screenshot of the 3D pain drawing used.




## Participants and Recruitment

For the purpose of this analysis, 3 different patient groups were targeted. The first group consisted of spinal cord injury patients (n=15), the second group were patients suffering from rheumatologic pain (n=13), and the third were individuals with some form of back pain (n=32). These participant groups totalled 60 individuals (26 female, 34 male; mean age 47.6 years, SD 12.9), who volunteered to take part in the study.

Recruitment was by convenience sampling—participants were patients in the Royal National Orthopaedic and Northwick Park hospitals, and members of the Hillingdon Independent Wheelchair User Group, all in the greater area of London. In order to approach the participants, we contacted the above organizations' clinical staff and/or administrators and asked them to recommend a number of people who could potentially satisfy our expected needs, and thus, take part in the study. The criteria for selection was that the participant had a medical condition that involved pain, was 18 years or older, and experienced some pain during the period of the evaluation.

Of the 60 participants who were asked to report their perceptions of using the 3D pain drawing to record their pain experience, the perceptions of only 24 (response rate 43%) were deemed usable and were consequently considered in the subsequent analysis.

## Instrumentation and Data Collection

Prior to collecting any data, informed consent was obtained from all participants. They were all given the 3D pain drawing to evaluate, the results of which are presented in previous work of ours [2,27]. For this study, participants were interviewed one week after using the 3D pain drawing to review their perceptions and attitudes towards adopting the technology to report their pain experience in everyday practice. Participants were not informed about the specific role of the interviewer as being part of the 3D pain drawing's development team in this study in order to avoid any potential risk of biased results, in the sense that participants could give overly favorable responses.

Semi-structured interviews were chosen for this purpose, as they are considered a valid and consistent method of data collection in phenomenological research. Specifically, participants were offered the opportunity to provide oral comments and/or suggestions about their experiences in assessing their pain using the 3D approach, through the interview questions presented in Textbox 1below.

**Textbox 1.** Interview questions agenda.

| |
|---|
| 1. What is your opinion about the pain drawing currently used to report your pain? |
| 1.1. What do you think is the best or worst aspect of this tool, and why? |
| 1.2. What do you think needs most improvement, and why? |
| 2. What is your opinion about the 3D pain drawing you recently used to report your pain? |
| 2.1. What do you think is the best or worst aspect of this tool, and why? |
| 2.2. What do you think needs most improvement, and why? |
| 3. Please state any comments/suggestions that you might have for both tools, if any. |
| 4. What would you want and need from a pain assessment tool? |

The above agenda was developed in order to ensure that participants would be able to provide their perceptions through an open-ended discussion that would allow them to focus on better exploring and communicating their thoughts and needs, as well as would help the authors identify issues related to their attitudes and particular needs regarding a 3D pain drawing. The interviews and data collection were conducted by the same person, and lasted approximately 20-30 minutes each, all producing rich information. All responses were recorded using a digital voice recorder, along with notes that would allow for further interpretation of the recorded information.

### Coding and Data Analysis

The responses of each participant were analyzed by employing a qualitative analysis approach, and particularly by combining thematic analysis with the principles of the grounded theory, which is a well-established approach in health care research (eg, [28-31]). Our selection of using the grounded theory lies in its capacity to help researchers formulate hypotheses or theories based on studied phenomena, or to discover participants' main concerns and how they continually try to resolve it [32]. Along these lines, identifying factors that could support the employment of such 3D technological advancements by patients in the assessment of pain can serve as important findings for further research studies. To discover what these factors are, and considering that there is a lack of prior research in this context, the grounded theory was employed, as it can provide the holistic view necessary to capture patients' perceptions from which the aforementioned factors could be derived.

Accordingly, thematic analysis and principles of the grounded theory were employed in the following systematic manner. Initially, the recorded interviews were transcribed and coding began by following the process of open-coding, which involves the systematic reading and consideration of every comment produced by each participant, using a line-by-line analysis [33]. The authors then read and reviewed the data, and subsequently developed a coding frame to facilitate the grouping of emerging issues from the data into core themes. This frame suggested the development of a core theme on the basis of grouping codes, which were abstracted from the various participant responses. As such, the data that were formed into core themes were the most relevant to the purpose of this study.

The collected data from each interview were further compared to each other, and to the data produced by the other interviews in order to find any similarities and repetitions, or differences in the emerging issues. Additional grouping codes were added as new issues emerged during the comparison process. Inter-rater reliability scores were not produced since the interviews were not structured [34].

Lastly, all findings were extensively re-reviewed by all authors, in order to validate the data, to gain a high-level understanding of the collected information that would help to identify potential participant attitudes, as well as reflect on any further implications and compare with previous work in the pain literature. All findings were then assessed with a thematic analysis.

## Results

### Overview

The thematic analysis produced 4 main high-level core themes related to the participants' responses to the interviews, namely: (1) better assessment, (2) perceived clinical utility, (3) intention to use, and (4) perceived ease of use.

The first core theme highlighted the potential that the 3D approach could offer towards the efforts to better facilitate the reporting, and thus, the assessment of the pain experience. A secondary core theme derived further suggested the participants' perceived clinical utility of the 3D pain drawing by focusing on its capacity to assist them in becoming better stakeholders in the pain management process.

The third and fourth core themes that were identified brought to the fore the participants' intentions to use the 3D pain drawing in their everyday pain assessment routine, and whether they believed it would be easy for them to systematically interact with, respectively. Table 1 presents a pool of selected participant responses that support these results, which are summarized and elaborated in the discussion that follows.

**Table 1.** Selected participant responses and core themes.

| Core theme | Code | Selected responses |
|---|---|---|
| **Better assessment** | | |
| | Accuracy | *It (3D approach) allows me to accurately pinpoint any location I choose.* [R1] |
| | Clarity | *(3D approach) is much better in showing where my pain is...the whole body seems to be closer to reality than the diagram (2D pain drawing) is.* [R2] |
| | Clarity | *You can actually focus better on that one (3D approach), as the body area is well represented and I can more easily show where my aches are.* [R3] |
| | Clarity | *The figure (2D pain drawing) was not adequate…I would definitely prefer something better.* [R4] |
| | Accuracy | *Being able to move it (ie, rotate it) and have a closer peek on my different body parts (ie, zoom in/out) makes me feel that I have a much better control of how to show where my pain is.* [R5] |
| **Perceived clinical utility** | | |
| | Allows correlation with activities | *It (3D approach) can allow me to better correlate the pain I am experiencing in certain parts of my body with the activities that I had been doing.* [R6] |
| | Allows correlation with medication | *It (3D approach) made me realize that I was taking my medication at the wrong time of the day.* [R7] |
| **Intention to use** | | |
| | User enthusiasm | *Amazing! The old one (2D pain drawing) should be retired now.* [R8] |
| | User preference | *It (3D approach) is more "friendly" to me… I actually prefer this since our body parts are now easier to see and show.* [R9] |
| | Increases user experience | *The diagram (2D pain drawing) is a bit "cold", and to be honest, I wouldn't mind using something better.* [R10] |
| | User preference | *I prefer something better than that (2D pain drawing) and I think your tool (3D approach) is better to show my pain.* [R11] |
| | User preference | *Now, I would never go back to the old one (2D pain drawing) again.* [R12] |
| | User preference | *I prefer your tool instead of the paper (2D drawing) figure.* [R13] |
| **Perceived ease of use** | | |
| | Ease of use | *Although it might be a bit hard to learn to use, I would like to see this tool (3D approach) again.* [R14] |

## Better Assessment

Responses revealed that participants were generally enthusiastic about the level of improvement that the 3D approach offered in assessing their pain experience, which was reported as being significantly better than its predecessor was. Particularly, participants highlighted two important feature advancements of the 3D pain drawing as compared to the 2D one—*accuracy* and *clarity*. The comments that were indicative of this view (Table 1, R1-R5) generally indicated the extent of the 3D approach to cover almost all assessment aspects of the existing conventional 2D pain drawing, while offering an enhanced level of detail. Building on this, participants overly reported that the 3D approach significantly contributed towards the improvement of current pain assessment practices.

## Perceived Clinical Utility

This second core theme revealed that participants were particularly enthusiastic with respect to the ability of the tool to offer a more accurate and structured means of correlating activities and medication intake with pain. It has to be pointed out that the above finding was not a direct result of 3D technology usage, but, was instead identified as an indirect implication of using a 3D tool to report pain.

Specifically, the study revealed that participants were provided the capability to better localize their pain location through 3D technology, whilst concurrently being able to monitor how various activities impacted their pain level at that localized body location. For instance, a participant found that she could manage her pain much better by reducing those activities that led to intense pain at that certain body location (R6).

Moreover, additional evidence came from another participant who claimed that he was also able to better monitor the progression and type of pain, vis-a-vis his prescribed medication/treatment (R7). This observed discrepancy between medication intake and experienced peaks of pain at certain body locations, ultimately resulted in that participant reducing his medication (strong analgesic) by 25%, with no deterioration in the pain levels encountered. Indeed, the reduction of medication intake as a result of self-monitoring of pain was not a singular observation, as this was also reported by 5 other members of our participant group.

### Intention to Use

The evidence produced from participant responses seemed to suggest that a 3D approach to pain assessment was seen as being of important clinical value to the participants. In fact, the majority of them were enthusiastic about the capabilities that the 3D approach provides with respect to better reporting pain to the clinicians involved in the assessment process, and they were generally in line with the response provided by one of the participants (R11) with respect to his preference for the 3D pain drawing to better report his pain experience as opposed to the 2D pain drawing (R9, R11-R13).

### Perceived Ease of Use

The overall responses seemed to indicate that the majority of the participants did not mention or was not concerned with the 3D pain drawing's potential ease of use. On the contrary, it must be remarked that from the transcribed responses, the general trend was that participants were enthusiastic about the functionality of the 3D pain drawing, with the majority of them expressing their expectations for the 3D drawing to better assist them in more efficiently reporting their pain.

## Discussion

### Principal Findings

As the need for quality health care provision continues to expand, 3D technology can be an efficient and intuitive approach for improving medical practices. The current paper set out to explore a qualitative perspective on the attitudes of 24 patients towards using a 3D pain drawing to report their pain experience, with a focus on understanding their perceptions and needs regarding the enhancement of current pain assessment practices with 3D technological advancements.

The findings from our reported study in the area of pain assessment suggested that participants generally accepted the introduction of a novel 3D approach for the purpose of reporting their pain characteristics as an alternative to the conventional 2D pain drawing, supporting the quantitative findings obtained in previous studies [2,27]. Specifically, the present findings revealed that participants generally appreciated the enhanced ability that the 3D approach offered with regard to reporting their pain. In particular, participants were enthusiastic about the capability it provided to better report their pain characteristics to the involved clinician(s), with results highlighting their preference and satisfaction in using the 3D approach to better describe and evaluate their pain in relation to everyday activities and/or medication received. This has led the authors to hypothesize that its employment in everyday practice could have a medium-term result, with the consequent improvement of the patient-clinician communication channel due to the more comprehensive communications for both parties using the 3D approach to assess pain.

### Comparison with Prior Work

In order to fit our results in the field of current pain research, we will compare our findings with those of previous work in the field. As such, our findings seemed to be consistent with the results of previous studies that looked at the added value of using computerized interfaces for pain assessment among various patients [35-38]. In addition, our study results support past findings regarding the success of 3D technology in the *management* and *treatment* of pain, where 3D technological solutions have been effectively used to treat conditions such as phantom limb pain [39-42], as a distraction technique for burn pain care [43-45], as an effective aid in reducing pain through hypnosis [46-48], and to decrease pain in cancer patients [49-51].

However, our review of the literature has shown that there is a paucity of research investigating the use of 3D technology for the *assessment* of pain, as opposed to past efforts that focused solely on the management and treatment aspects. Therefore, while 3D technology has been extensively used for health data reporting over the past 20 years, we are unaware of any work employing 3D technology for the collection of pain-related data, with the exception of the authors' previous work.

Consequently, the authors' view on this study is that 3D technology should not be applied to the above identified problem for the sake of it, but that any new 3D approach should be more intuitive than the existing solutions for pain assessment, and just as usable as the 3D technological tools for pain management and treatment. The hypothesis is that the 3D pain drawing tool presented in this study could offer a significant improvement to the current level of patient pain assessment.

Our study, therefore, provided additional qualitative evidence to further demonstrate the attitudes of patients about the adoption of 3D technology, and supported the above hypothesis by showing the important role that it could also play in *collecting*, *reporting*, and *assessing* their pain experience. The latter has been shown to be highly valued by those with pain-related conditions, as shown by this study's results.

### Potential Implications

Our findings seem to suggest that employing a 3D approach for pain assessment practices could, in the medium-term, have important implications. Specifically, in line with the authors' anticipation, our overall findings propose that the successful application of 3D technology in pain assessment could play an important starting role towards improving the provision of quality health care.

In particular, considering that individuals with pain often rely on several health care institutions (eg, hospitals, health, and social care institutes) for assistance, it is anticipated that the 3D pain drawing could be adopted by these institutions for the purpose of assisting the sufferers with their pain reporting by employing a more efficient and accurate pain assessment tool, supporting the use of 3D technology in the overall pain management process.

The capability therefore, offered by the 3D pain drawing is theorized to have significant implications in clinical practice. Even if usability issues are present, as one of the participants suggested (R14), we speculate that most individuals would overlook this, since the convenience factor associated with it would outweigh such considerations (eg, there could be less hospital visits).

The above assumption is also supported by the clinicians involved in the quantitative evaluations of the 3D pain drawings described in past work [27]. According to their comments, the possibility of individuals self-reporting their pain might have very positive implications towards improving the provision of quality care, since individuals could: (1) remotely monitor the progression and type of their pain, vis-a-vis their prescribed medication-a finding also confirmed by the present study's results, (2) become better stakeholders in managing their pain, and (3) also benefit from a psychological point of view.

From another perspective, it is generally accepted that while the cornerstone to efficient pain management is the successful assessment of the pain experience, this effort often relies on the health care professional's empathy, interest, and understanding of a patient's condition at the time of assessment [52]. However, the clinician's heavy workload or tiredness, for example, could often affect the aforementioned aspects. As such, although the 3D approach does not offer a diagnosis, it could potentially reduce the need for the above constant reliance on the clinician and ease his/her assessment, through its capacity to facilitate better pain self-reporting and remote assessment. This could also potentially ease the congestion and waiting times often experienced in health care settings, and further contribute to reducing hospital visits.

It is therefore speculated that the above could first, empower individuals with pain as a result of receiving improved quality of care, and second, could help moderate the role of health care institutions.

### Limitations and Future Work

Apart from the positive findings, this study has also raised certain limitations and avenues for future work that accordingly need to be acknowledged. First, we acknowledge that our participant sample size was relatively small. Unfortunately, it was rather impractical to recruit a large number of participants for this study, as only individuals with pain could be considered. It would, thus, be beneficial to involve a larger sample, in order to support a more informed study result. Second, it is recognized that the findings of the present study cannot be generalized with respect to every 3D technology used in health care. However, our findings could be used to draw generalized conclusions with regard to the usefulness of 3D models, and further the attitudes of patients towards the adoption of such 3D models in order to record health-related findings. Given the limited existing research efforts in this particular area of health care, our findings may be considerably useful as they could offer significant insight and could be used as an important point of reference for future efforts.

Third, although one of the beneficiaries of this study's findings is health care providers, the work presented in this study was limited to the perceptions of patients admitted to hospitals. As such, it has to be made clear that since this study was prototypical, it has not yet been tested in the remaining health care settings. In retrospect, an attractive future direction would be the investigation of the 3D approach with patients from the whole range of health care providers. Finally, future research may also pay attention to a wider range of pain-related medical conditions by examining the attitudes of, for instance, patients recovering from surgery or being treated for cancer.

### Conclusions

The findings of this study show that, despite past research efforts, individuals with pain are still not entirely satisfied with the adequacy of current pain assessment practices. As it has been revealed by our findings, this could be due to several factors including the individuals' limited understanding of their pain and the factors that may or may not affect it, as well as to their inability to accurately describe it to medical professionals using the conventional tools.

Consequently, it has been speculated that the use of 3D technology as an alternative could create the possibility for individuals with pain to become better stakeholders in the management of their pain experience. This in turn could have positive implications towards improving the provision of quality health care, particularly with respect to contributing towards time-effective and improved care for individuals with pain. These findings may be of considerable interest to health care providers, policy makers, researchers, and other parties that might be actively involved in the area of pain research and management.

## Acknowledgments

The authors are grateful to Dr. Jan Gawronski for his contribution and valued time in the realization of this work. The authors would like to acknowledge Cyberware, Inc. for providing a 3D model, which was adapted, manipulated, and extended to meet our needs.

## Conflicts of Interest

None declared.

## References

1. Reading AE. Testing pain mechanisms in persons in pain. In: Wall PD, Melzack R, editors. Textbook of pain. Edinburgh: Churchill Livingstone; 1989:269-283.
2. Ghinea G, Spyridonis F, Serif T, Frank AO. 3-D pain drawings-mobile data collection using a PDA. IEEE Trans Inf Technol Biomed 2008 Jan;12(1):27-33. [doi: 10.1109/TITB.2007.903266] [Medline: 18270034]
3. Coll AM, Ameen JR, Mead D. Postoperative pain assessment tools in day surgery: literature review. J Adv Nurs 2004 Apr;46(2):124-133. [doi: 10.1111/j.1365-2648.2003.02972.x] [Medline: 15056325]

4. Udén A, Aström M, Bergenudd H. Pain drawings in chronic back pain. Spine (Phila Pa 1976) 1988 Apr;13(4):389-392. [Medline: 2970123]
5. Takata K, Hirotani H. Pain drawing in the evaluation of low back pain. Int Orthop 1995;19(6):361-366. [Medline: 8567153]
6. Ohnmeiss DD, Vanharanta H, Guyer RD. The association between pain drawings and computed tomographic/discographic pain responses. Spine (Phila Pa 1976) 1995 Mar 15;20(6):729-733. [Medline: 7604350]
7. Samuelsson K, Larsson H, Thyberg M, Tropp H. Back pain and spinal deformity—common among wheelchair users with spinal cord injuries. Scand J Occup Ther 1996 Jan;3(1):28-32. [doi: 10.3109/11038129609106679]
8. Felix ER, Cruz-Almeida Y, Widerström-Noga EG. Chronic pain after spinal cord injury: what characteristics make some pains more disturbing than others? J Rehabil Res Dev 2007;44(5):703-715 [FREE Full text] [Medline: 17943682]
9. Felix ER, Galoian KA, Aarons C, Brown MD, Kearing SA, Heiss U. Utility of quantitative computerized pain drawings in a sample of spinal stenosis patients. Pain Med 2010 Mar;11(3):382-389. [doi: 10.1111/j.1526-4637.2009.00788.x] [Medline: 20088853]
10. Serif T, Ghinea G. Recording of time-varying back-pain data: a wireless solution. IEEE Trans Inf Technol Biomed 2005 Sep;9(3):447-458. [Medline: 16167699]
11. Sanders NW, Mann NH, Spengler DM. Pain drawing scoring is not improved by inclusion of patient-reported pain sensation. Spine (Phila Pa 1976) 2006 Nov 1;31(23):2735-41; discussion 2742. [doi: 10.1097/01.brs.0000244674.99258.f9] [Medline: 17077744]
12. Masferrer R, Prendergast V, Hagell P. Colored pain drawings: preliminary observations in a neurosurgical practice. Eur J Pain 2003;7(3):213-217. [doi: 10.1016/S1090-3801(02)00113-1] [Medline: 12725843]
13. Mannion AF, Balagué F, Pellisé F, Cedraschi C. Pain measurement in patients with low back pain. Nat Clin Pract Rheumatol 2007 Nov;3(11):610-618. [doi: 10.1038/ncprheum0646] [Medline: 17968331]
14. Arendt-Nielsen L, Nie H, Laursen MB, Laursen BS, Madeleine P, Simonsen OH, et al. Sensitization in patients with painful knee osteoarthritis. Pain 2010 Jun;149(3):573-581. [doi: 10.1016/j.pain.2010.04.003] [Medline: 20418016]
15. Sinikallio S, Aalto T, Airaksinen O, Herno A, Kröger H, Savolainen S, et al. Depression is associated with poorer outcome of lumbar spinal stenosis surgery. Eur Spine J 2007 Jul;16(7):905-912 [FREE Full text] [doi: 10.1007/s00586-007-0349-3] [Medline: 17394027]
16. Baker M. A European Consensus Report. 2010. Improving the current and future management of chronic pain URL: http://www.mijnpijn.nl/pdf/PainProposalEuropeanReport.pdf [WebCite Cache ID 6BziA5Xsb]
17. Jansen GB, Linder J, Ekholm KS, Ekholm J. Differences in symptoms, functioning, and quality of life between women on long-term sick-leave with musculoskeletal pain with and without concomitant depression. J Multidiscip Healthc 2011;4:281-292 [FREE Full text] [doi: 10.2147/JMDH.S21827] [Medline: 21847350]
18. Persson AL, Garametsos S, Pedersen J. Computer-aided surface estimation of pain drawings - intra- and inter-rater reliability. J Pain Res 2011;4:135-141 [FREE Full text] [doi: 10.2147/JPR.S18637] [Medline: 21647217]
19. von Baeyer CL, Lin V, Seidman LC, Tsao JC, Zeltzer LK. Pain charts (body maps or manikins) in assessment of the location of pediatric pain. Pain Manag 2011 Jan;1(1):61-68 [FREE Full text] [doi: 10.2217/pmt.10.2] [Medline: 21572558]
20. Andersen T, Christensen FB, Høy KW, Helmig P, Niedermann B, Hansen ES, et al. The predictive value of pain drawings in lumbar spinal fusion surgery. Spine J 2010 May;10(5):372-379. [doi: 10.1016/j.spinee.2010.02.002] [Medline: 20421073]
21. Wenngren A, Stålnacke BM. Computerized assessment of pain drawing area: A pilot study. Neuropsychiatr Dis Treat 2009;5:451-456 [FREE Full text] [Medline: 19721724]
22. Lauche R, Cramer H, Haller H, Musial F, Langhorst J, Dobos GJ, et al. My back has shrunk: the influence of traditional cupping on body image in patients with chronic non-specific neck pain. Forsch Komplementmed 2012;19(2):68-74. [doi: 10.1159/000337688] [Medline: 22585102]
23. St John M, Cowen MB, Smallman HS, Oonk HM. The use of 2D and 3D displays for shape-understanding versus relative-position tasks. Hum Factors 2001;43(1):79-98. [Medline: 11482314]
24. Marcus A, Feng L, Maletic JI. 3D representations for software visualization. 2003 Presented at: ACM Symposium on Software Visualization; 2003; San Diego, USA p. 11-12.
25. Duerstock BS, Bajaj CL, Pascucci V, Schikore D, Lin KN, Borgens RB. Advances in three-dimensional reconstruction of the experimental spinal cord injury. Comput Med Imaging Graph 2000 Dec;24(6):389-406. [Medline: 11008186]
26. Frank AO, De Souza LH. Conservative management of low back pain. Int J Clin Pract 2001 Feb;55(1):21-31. [Medline: 11219314]
27. Spyridonis F, Gawronski J, Ghinea G, Frank AO. An interactive 3-D application for pain management: Results from a pilot study in spinal cord injury rehabilitation. Comput Methods Programs Biomed 2012 Oct;108(1):356-366. [doi: 10.1016/j.cmpb.2012.02.005] [Medline: 22401773]
28. Liamputtong P, Ezzy D. Qualitative research methods. South Melbourne: Oxford University Press; 2005.
29. Brand CA, Abi HY, Couch DE, Vindigni A, Wark JD. Vitamin D deficiency: a study of community beliefs among dark skinned and veiled people. Int J Rheum Dis 2008 Apr;11(1):15-23. [doi: 10.1111/j.1756-185X.2008.00323.x]
30. Katsakou C, Rose D, Amos T, Bowers L, McCabe R, Oliver D, et al. Psychiatric patients' views on why their involuntary hospitalisation was right or wrong: a qualitative study. Soc Psychiatry Psychiatr Epidemiol 2012 Jul;47(7):1169-1179. [doi: 10.1007/s00127-011-0427-z] [Medline: 21863281]

31. Skordis-Worrall J, Hanson K, Mills A. Confusion, caring and tuberculosis diagnostic delay in Cape Town, South Africa. Int J Tuberc Lung Dis 2010 Feb;14(2):171-180. [Medline: 20074408]
32. Glaser B. Basics of grounded theory analysis: emergence vs forcing. Mill Valley, CA: Sociology Pr; 1992.
33. Strauss A, Corbin J. Basics of qualitative research: techniques and procedures for developing grounded theory. Thousand Oaks: Sage Publications; 1998.
34. Morse J. Perfectly healthy but dead: the myth of inter-rater reliability. Qual Health Res 1997 Nov;7(4):445-447. [doi: 10.1177/104973239700700401]
35. Lalloo C, Henry JL. Evaluation of the iconic pain assessment tool by a heterogeneous group of people in pain. Pain Res Manag 2011 Feb;16(1):13-18 [FREE Full text] [Medline: 21369536]
36. Jacob E, Stinson J, Duran J, Gupta A, Gerla M, Ann Lewis M, et al. Usability testing of a smartphone for accessing a web-based e-diary for self-monitoring of pain and symptoms in sickle cell disease. J Pediatr Hematol Oncol 2012 Jul;34(5):326-335. [doi: 10.1097/MPH.0b013e318257a13c] [Medline: 22627570]
37. Jamison RN, Fanciullo GJ, Baird JC. Computerized dynamic assessment of pain: comparison of chronic pain patients and healthy controls. Pain Med 2004 Jun;5(2):168-177. [doi: 10.1111/j.1526-4637.2004.04032.x] [Medline: 15209971]
38. Marceau LD, Link C, Jamison RN, Carolan S. Electronic diaries as a tool to improve pain management: is there any evidence? Pain Med 2007 Oct;8 Suppl 3:S101-S109. [doi: 10.1111/j.1526-4637.2007.00374.x] [Medline: 17877520]
39. Gaggioli A, Amoresano A, Gruppioni E, Verni G, Riva G. A myoelectric-controlled virtual hand for the assessment and treatment of phantom limb pain in trans-radial upper extremity amputees: a research protocol. Stud Health Technol Inform 2010;154:220-222. [Medline: 20543301]
40. Murray CD, Pettifer S, Howard T, Patchick EL, Caillette F, Kulkarni J, et al. The treatment of phantom limb pain using immersive virtual reality: three case studies. Disabil Rehabil 2007 Sep 30;29(18):1465-1469. [doi: 10.1080/09638280601107385] [Medline: 17729094]
41. Cole J, Crowle S, Austwick G, Slater DH. Exploratory findings with virtual reality for phantom limb pain; from stump motion to agency and analgesia. Disabil Rehabil 2009;31(10):846-854. [doi: 10.1080/09638280802355197] [Medline: 19191061]
42. Lamont K, Chin M, Kogan M. Mirror box therapy: seeing is believing. Explore (NY) 2011 Dec;7(6):369-372. [doi: 10.1016/j.explore.2011.08.002] [Medline: 22051561]
43. Maani CV, Hoffman HG, Fowler M, Maiers AJ, Gaylord KM, Desocio PA. Combining ketamine and virtual reality pain control during severe burn wound care: one military and one civilian patient. Pain Med 2011 Apr;12(4):673-678. [doi: 10.1111/j.1526-4637.2011.01091.x] [Medline: 21481162]
44. Maani CV, Hoffman HG, Morrow M, Maiers A, Gaylord K, McGhee LL, et al. Virtual reality pain control during burn wound debridement of combat-related burn injuries using robot-like arm mounted VR goggles. J Trauma 2011 Jul;71(1 Suppl):S125-S130. [doi: 10.1097/TA.0b013e31822192e2] [Medline: 21795888]
45. Hoffman HG, Chambers GT, Meyer WJ, Arceneaux LL, Russell WJ, Seibel EJ, et al. Virtual reality as an adjunctive non-pharmacologic analgesic for acute burn pain during medical procedures. Ann Behav Med 2011 Apr;41(2):183-191. [doi: 10.1007/s12160-010-9248-7] [Medline: 21264690]
46. Patterson DR, Tininenko JR, Schmidt AE, Sharar SR. Virtual reality hypnosis: a case report. Int J Clin Exp Hypn 2004 Jan;52(1):27-38. [doi: 10.1076/iceh.52.1.27.23925] [Medline: 14768967]
47. Patterson DR, Jensen MP, Wiechman SA, Sharar SR. Virtual reality hypnosis for pain associated with recovery from physical trauma. Int J Clin Exp Hypn 2010 Jul;58(3):288-300 [FREE Full text] [doi: 10.1080/00207141003760595] [Medline: 20509069]
48. Patterson DR, Wiechman SA, Jensen M, Sharar SR. Hypnosis delivered through immersive virtual reality for burn pain: A clinical case series. Int J Clin Exp Hypn 2006 Apr;54(2):130-142. [doi: 10.1080/00207140500528182] [Medline: 16581687]
49. Sander Wint S, Eshelman D, Steele J, Guzzetta CE. Effects of distraction using virtual reality glasses during lumbar punctures in adolescents with cancer. Oncol Nurs Forum 2002 Feb;29(1):E8-E15 [FREE Full text] [doi: 10.1188/02.ONF.E8-E15] [Medline: 11845217]
50. Windich-Biermeier A, Sjoberg I, Dale JC, Eshelman D, Guzzetta CE. Effects of distraction on pain, fear, and distress during venous port access and venipuncture in children and adolescents with cancer. J Pediatr Oncol Nurs 2007 Feb;24(1):8-19. [doi: 10.1177/1043454206296018] [Medline: 17185397]
51. Kwekkeboom KL, Cherwin CH, Lee JW, Wanta B. Mind-body treatments for the pain-fatigue-sleep disturbance symptom cluster in persons with cancer. J Pain Symptom Manage 2010 Jan;39(1):126-138 [FREE Full text] [doi: 10.1016/j.jpainsymman.2009.05.022] [Medline: 19900778]
52. Fink R. Pain assessment: the cornerstone to optimal pain management. Proc (Bayl Univ Med Cent) 2000 Jul;13(3):236-239 [FREE Full text] [Medline: 16389388]


## Abbreviations

**2D:** 2-dimensional

**3D:** 3-dimensional

*Edited by G Eysenbach; submitted 08.11.12; peer-reviewed by D Rodriguez, M Themistocleous; comments to author 24.01.13; revised version received 06.02.13; accepted 08.02.13; published 10.04.13.*

*Please cite as:*
*Spyridonis F, Ghinea G, Frank AO*
*Attitudes of Patients Toward Adoption of 3D Technology in Pain Assessment: Qualitative Perspective*
*J Med Internet Res 2013;15(4):e55*
*URL: http://www.jmir.org/2013/4/e55/*
*doi:10.2196/jmir.2427*
*PMID:23575479*

Original Paper

# Preferred Sources of Health Information in Persons With Multiple Sclerosis: Degree of Trust and Information Sought

Ruth Ann Marrie[1], MD, PhD, FRCPC; Amber R Salter[2], MPH; Tuula Tyry[3], PhD; Robert J Fox[4], MD, MS; Gary R Cutter[2], PhD

[1]University of Manitoba, Departments of Internal Medicine & Community Health Sciences, Winnipeg, MB, Canada

[2]University of Alabama at Birmingham, Department of Biostatistics, Birmingham, AL, United States

[3]Barrow Neurological Institute, Division of Neurology, Phoenix, AZ, United States

[4]Neurological Institute, Mellen Center for MS, Cleveland Clinic, Cleveland, OH, United States

**Corresponding Author:**
Ruth Ann Marrie, MD, PhD, FRCPC
University of Manitoba
Departments of Internal Medicine & Community Health Sciences
GF 543 - Health Sciences Centre
820 Sherbrook Street
Winnipeg, MB, R3A 1R9
Canada
Phone: 1 204 787 4951
Fax: 1 204 787 1486
Email: rmarrie@hsc.mb.ca

## Abstract

**Background:** Effective health communication is important for informed decision-making, yet little is known about the range of information sources used by persons with multiple sclerosis (MS), the perceived trust in those information sources, or how this might vary according to patient characteristics.

**Objective:** We aimed to investigate the sources of health information used by persons with MS, their preferences for the source of health information, and levels of trust in those information sources. We also aimed to evaluate how these findings varied according to participant characteristics.

**Methods:** In 2011, participants in the North American Research Committee on Multiple Sclerosis (NARCOMS) Registry were asked about their sources of health information using selected questions adapted from the 2007 Health Information National Trends (HINTS) survey.

**Results:** Of 12,974 eligible participants, 66.18% (8586/12,974) completed the questionnaire. Mass media sources, rather than interpersonal information sources, were the first sources used by 83.22% (5953/7153) of participants for general health topics and by 68.31% (5026/7357) of participants for MS concerns. Specifically, the Internet was the first source of health information for general health issues (5332/7267, 73.40%) and MS (4369/7376, 59.23%). In a logistic regression model, younger age, less disability, and higher annual income were independently associated with increased odds of use of mass media rather than interpersonal sources of information first. The most trusted information source was a physician, with 97.94% (8318/8493) reporting that they trusted a physician some or a lot. Information sought included treatment for MS (4470/5663, 78.93%), general information about MS (3378/5405, 62.50%), paying for medical care (1096/4282, 25.59%), where to get medical care (787/4282, 18.38%), and supports for coping with MS (2775/5031, 55.16%). Nearly 40% (2998/7521) of participants had concerns about the quality of the information they gathered.

**Conclusions:** Although physicians remain the most trusted source of health information for people with MS, the Internet is the first source of health information for most of them. This has important implications for the dissemination of health information.

***(J Med Internet Res 2013;15(4):e67)*** doi:10.2196/jmir.2466

**KEYWORDS**

multiple sclerosis; Internet; social media; trust; health information

## Introduction

When seeking health information, patients use interpersonal sources including peers, and mass media sources such as the Internet. Preferred sources of health information vary by age, ethnicity, socioeconomic status and type of chronic condition [1-5]. Use of the Internet as a health information source, for example, varies from 35% in Canadians with spinal cord injury [2], to 62% in rheumatoid arthritis patients from New Jersey, United States, participating in a commercial health plan [3], to 63-82% in people with multiple sclerosis (MS) from regions in the United States and Israel [4,5].

Effective health communication is important for informed decision-making, yet little is known about the information sources used by people with MS, the perceived trust in those sources, or how this might vary by patient characteristics. Before their first visit at an MS clinic, 82% of people with MS from Ohio, United States, gathered information from the Internet [4]. Another study suggested that more disabled people with MS preferred interaction with health care providers over seeking information electronically; however, that study focused on Internet use in a clinic population and did not evaluate the breadth of information sources used [5].

We aimed to investigate the sources of health information used by people with MS, their preferences for the source of health information, and levels of trust in those information sources. We also aimed to evaluate how these findings varied according to participant characteristics.

## Methods

### North American Research Committee on Multiple Sclerosis Registry

The Consortium of MS Centers developed the North American Research Committee on Multiple Sclerosis (NARCOMS) Registry as a voluntary self-report registry for people with MS [6]. NARCOMS participants agree to the use of their de-identified data for research purposes, and the Registry is approved by the Institutional Review Board at the University of Alabama at Birmingham. Diagnoses of MS were validated in a randomly selected sample of participants [7]. Participants enrolled by mailing in a questionnaire or by completing a questionnaire online [6]. Thereafter, they completed surveys semi-annually, on paper, or online. At enrollment and on the semi-annual surveys, participants provided sociodemographic and clinical information. Disability status was reported using Patient Determined Disease Steps (PDDS) [8], a validated measure that correlates well with a physician-scored Expanded Disability Status Scale (EDSS) [8]. It is scored ordinally from 0 to 8, where a score of 0 approximates an EDSS score of 0, a score of 3 represents early gait disability without needing an assistive device and approximates an EDSS score of 4.0 to 4.5; and scores of 4, 5, and 6 represent EDSS scores of 6 to 6.5.

### Questionnaire

In 2011, participants were asked about sources of health information using questions adapted from the 2007 Health Information National Trends (HINTS) survey [9]. HINTS was developed by the National Cancer Institute to evaluate changing trends in health communication, to assess access to and use of cancer information, and to evaluate perception of cancer risk and health behaviors [9]. The first section of the survey queried information seeking about health topics during the respondents' most recent search, the second inquired about information seeking about cancer, and the third focused on Internet use. Collectively, this survey captured information seeking from interpersonal and mass media sources. We substituted MS for cancer in these sections, and substituted MS organizations (such as the National Multiple Sclerosis Society [NMSS] and Consortium of Multiple Sclerosis Centers for cancer organizations. The adapted questions can be reviewed in Multimedia Appendix 1.

### Statistical Analysis

Because of variability in the availability of health care information sources and in health care systems worldwide, we restricted our analysis to NARCOMS participants living in the United States. Missing responses were not imputed. Note that some participants did not respond to all questions, and some data were missing for non-responders; therefore we report the number of individuals responding to each question throughout. We summarized categorical variables using frequency (percent) and continuous variables using mean (standard deviation) or median (interquartile range) as appropriate.

Use of information sources was tabulated by individual source where only one choice could be selected, and categorized as interpersonal (family, provider, friend, patient advocacy organization, [10]) or mass media sources (books, newspapers, brochures, library, magazines, Internet). Using logistic regression we examined demographic and clinical factors associated with use of mass media versus interpersonal sources of information, factors associated with Internet use (yes vs no), the type of information (yes vs no) sought, including information sought by more than 50% of participants and regarding access to care. We identified health information sources reported by ≥5% of participants as their first information source, and examined the association between sociodemographic and clinical characteristics, and level of trust for those information sources using logistic regression models. For each health information source, trust was dichotomized as more (a lot or some) versus less (a little or not at all) [11].

The independent variables considered for each regression model are described below. Race was categorized as white (reference group), and non-white. Education was included as indicator variables for high school diploma or less (reference group), Associate's Degree or Technical Degree, Bachelor's Degree, and post-graduate degree. Annual household income was included as indicator variables for <$15,000 (reference group), $15,000-29,999, $30,000-49,999, $50,000-100,000, >$100,000, or declined to answer. Insurance status was included as indicator variables for private, public only (reference group), or none. Region of residence was included as indicator variables for West (reference group), Midwest, South, and East as defined by the US Census bureau. Age was categorized as 18-34, 35-49, 50-59, and ≥60 (reference group) years. Disease duration was categorized into quartiles, thus 0-16, 17-24, 25-33 and ≥34

years. Using PDDS, participants were classified as having mild (0-2), moderate (3-4), or severe (5-8, reference group) disability [12]. We also included a variable indicating whether the questionnaire was completed online or on paper.

Assumptions of models were tested using standard methods [13]. For each model, we reported a c-statistic as a measure of discriminating ability (estimate of area under the curve) and the Hosmer Lemeshow test as a measure of goodness of fit. Analyses were performed using SAS V9.2.

## Results

### Respondents

Of 12,974 eligible participants, 8586/12,974 (66.18%) responded. The demographic and clinical characteristics of the responders are summarized in Table 1. As compared to responders, non-responders were more likely to be women (80.90%, 3542/4378 vs 77.56%, 6650/8574, *P*<.0001), and of lower socioeconomic status (annual income at enrollment <$15,000; 458/4219, 10.86% vs 795/8176, 9.72%, *P*=.047). Non-responders had a lower age of symptom onset (mean 30.0, SD 9.7 years) than responders (mean 30.9, SD 10.0 years), and a lower age of diagnosis (mean 37.7, SD 9.8 vs mean 38.5, SD 9.7, both *P*<.0001). Although these differences were statistically significant, the differences in the absolute values were so small that they were unlikely to be clinically significant. Responders and non-responders did not differ with respect to the severity of disability at enrollment.

### Information Sources

Eighty-nine percent of respondents (7512/8439) reported ever seeking information about health or medical topics from any source. Participants who completed the survey online were more likely to answer this question affirmatively (5291/5635, 93.90%) than those who did not (2221/2951, 75.26%, *P*<.0001). Similarly, 88.13% (7459/8464) reported that they had ever sought information about MS specifically, and such information-seeking behavior was greater among participants who completed the survey online (5157/5610, 91.92%) versus paper (2302/2854, 80.66%, *P*<.0001)

Table 2 shows the first source that participants went to for information about general health or medical topics, for the most recent time that they sought information. The Internet was the most common choice, reported by 73.37% (5332/7267) of participants who responded to the question and correctly selected only one choice. The second choice was health care providers (8.48%, 616/7267), followed by the NMSS (400/7267, 5.50%). The Internet was a less common first source of information regarding MS specifically, being reported by 4369/7267 (59.23%) of participants. Health care providers remained the second choice (1127/7376, 15.28%) and the NMSS remained the third choice (962/7376, 13.04%).

For further analysis, information sources were categorized as interpersonal or mass media. Mass media was the first information source used by 83.22% (5953/7153) of participants for general health topics and by 68.31% (5026/7357) of participants for MS concerns. On univariate analysis, characteristics associated with greater mass media use were female sex (*P*=.0048), younger age (*P*<.0001), greater than a high school education (*P*<.0001), higher annual income (*P*<.0001), survey completion online (*P*<.0001), and less disability (*P*<.0001). In a logistic regression model, younger age, less disability, and higher annual income were independently associated with increased odds of use of mass media rather than interpersonal sources of information first (Table 3).

### Health Information Sought

Participants sought a broad range of information regarding MS, including general aspects of MS, treatment, access to health care, and support for coping with MS (Table 4). Using multivariable logistic regression, we evaluated the association between participant characteristics and the types of information sought by more than 50% of participants (Multimedia Appendix 2). Consistently, higher socioeconomic status was associated with lower odds of searching for information. Lower levels of disability were associated with increased odds of searching for information about symptoms, while longer disease duration was associated with decreased odds of searching for information about coping and symptoms.

Information regarding access to care was also sought frequently on topics including paying for medical care (1096/4282, 25.59%), insurance (1008/4393, 22.95%), and where to get medical care (787/4282, 18.38%). Using logistic regression, lower socioeconomic status was consistently associated with searching for information regarding access to care (Multimedia Appendix 3).

### Internet Use

Most participants had accessed the Internet or used email (7292/8469, 86.10%), with more than 60% (4736) indicating that they accessed the Internet several times a day. Participants who completed the questionnaire online used the Internet more frequently than those who completed the questionnaire on paper (linear trend *P*<.0001). Using logistic regression, younger age, higher educational level, higher annual income, and milder disability were independently associated with increased odds of Internet use (Multimedia Appendix 4). Non-white race and public rather than private health insurance was associated with decreased odds of Internet use. Internet use varied somewhat across regions, being lower in the East and Midwest regions than the West.

Participants reported conducting several other activities online including social networking (4606/7528, 61.18%), seeking a health care provider (2595/7516, 34.53%), buying medications or vitamins (2583/7537, 34.27%), communicating with a physician (2408/7536, 31.95%), seeking advice regarding diet, weight, or physical activity (2401/7522, 31.92%), tracking personal health information such as test results or medical appointments (2159/5366, 28.69%), downloading information to a device (1971/7516, 26.21%), using online support groups for people with MS (1557/7526, 20.69%) and blogging (584/7515, 7.77%).

### Satisfaction and Trust

A substantial proportion of respondents had concerns regarding their search for information. Specifically, 2131/7556 (28.20%) felt that it took a lot of effort to get the information required, while 2120/7531 (28.15%) felt frustrated during their search, 2998/7521 (39.86%) were concerned about the quality of information gathered, and 1590/7533 (21.11%) thought that the information obtained was hard to understand.

The degree of trust varied across information sources. The most trusted source of information was a physician, with 97.94% (8318/8493) reporting that they trusted a physician some or a lot (Figure 1). Table 5 shows the results of logistic regression analyses for the association of trust (more vs less) and sociodemographic and clinical characteristics for the three most commonly used information sources for MS-related information (physicians, the Internet, and patient advocacy groups). Although some associations varied according to the information source, general patterns emerged. Higher levels of education and income were associated with greater trust in the three sources, as was mild rather than severe disability. Although age was not associated with trust in doctors, younger age was associated with increased trust in the Internet and patient advocacy groups.

**Figure 1.** Degree of trust in various information sources.

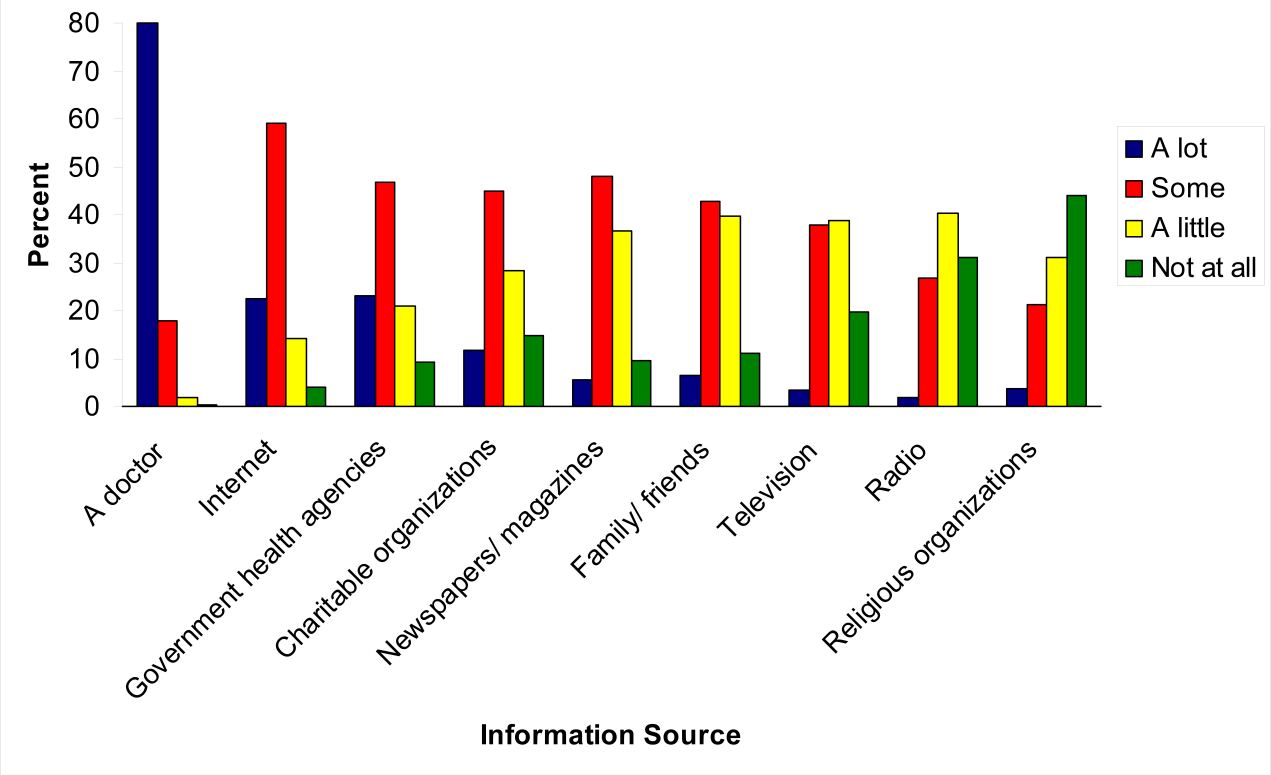


**Table 1.** Demographic and clinical characteristics of responders to the NARCOMS Fall 2011 questionnaire.

| Characteristic | | n/N (%) or mean (SD) |
|---|---|---|
| **Sex** | | |
| | Female | 6649/8573 (77.56) |
| | Male | 1924/8573 (22.44) |
| **Race** | | |
| | White | 7610/7972 (95.46) |
| | Non-White | 362/7972 (4.54) |
| **Education** | | |
| | High school diploma or less | 2777/8505 (32.65) |
| | Associate's or Technical degree | 1457/8505 (17.13) |
| | Bachelor's degree | 2392/8505 (28.12) |
| | Post-graduate degree | 1879/8505 (22.09) |
| **Annual income** | | |
| | <$15,000 | 706/8362 (8.44) |
| | $15,000-29,999 | 1249/8362 (14.94) |
| | $30,000-49,999 | 1347/8362 (16.11) |
| | $50,000-100,000 | 2020/8362 (24.16) |
| | >$100,000 | 1194/8362 (14.28) |
| | I do not wish to answer | 1846/8362 (22.08) |
| **Health insurance** | | |
| | Private | 5612/8384 (66.94) |
| | Public only | 2561/8384 (30.55) |
| | None | 211/8384 (2.52) |
| **Region of residence** | | |
| | East | 1981/8579 (23.09) |
| | Midwest | 2180/8579 (25.41) |
| | South | 2237/8579 (26.08) |
| | West | 2181/8579 (25.42) |
| **Current age (years), mean (SD)** | | |
| | | 56.6 (10.5) |
| **Age of symptom onset (years), mean (SD)** | | |
| | | 30.9 (10.0) |
| **Age of diagnosis (years), mean (SD)** | | |
| | | 38.5 (9.7) |
| **Patient determined disease steps (categorized)** | | |
| | Mild (0-2) | 3009/8507 (35.37) |
| | Moderate (3-4) | 2216/8507 (26.05) |
| | Severe (5-8) | 3282/8507 (38.58) |
| **Questionnaire administration** | | |
| | Online | 5635/8586 (65.63) |
| | Paper | 2951/8586 (34.37) |

**Table 2.** The first source for health information, the last time it was sought (N=7376).

| Source of health information | General health n (%) | Multiple Sclerosis n (%) |
|---|---|---|
| Internet | 5332 (73.40) | 4369 (59.23) |
| Doctor or health care provider | 616 (8.48) | 1127 (15.28) |
| National Multiple Sclerosis Society | 400 (5.50) | 962 (13.04) |
| Books | 267 (3.67) | 175 (2.37) |
| Magazines | 141 (1.94) | 174 (2.36) |
| Other | 115 (1.58) | 116 (1.57) |
| Brochures, pamphlets, etc | 104 (1.43) | 142 (1.93) |
| Family | 78 (1.07) | 44 (0.60) |
| Consortium of MS Centers | 39 (0.54) | 84 (1.14) |
| Telephone information number | 50 (0.69) | 43 (0.58) |
| Newspapers | 36 (0.50) | 22 (0.30) |
| Complementary or alternative practitioner | 35 (0.48) | 35 (0.47) |
| Friend/co-worker | 32 (0.44) | 41 (0.56) |
| Library | 22 (0.30) | 42 (0.57) |

**Table 3.** Characteristics associated with using mass media versus interpersonal information sources as the first source of health information (n=6348).[a]

| Characteristic | | Odds Ratio | 95% CI |
|---|---|---|---|
| **Age group, years** | | | |
| | 18-34 | 1.38 | 0.91, 2.10 |
| | 35-49 | 2.01 | 1.65, 2.44 |
| | 50-59 | 1.58 | 1.37, 1.83 |
| | ≥60 (Reference[b]) | 1.0 | |
| **Annual income** | | | |
| | <$15,000 (Reference[b]) | 1.0 | |
| | $15,000-29,999 | 1.44 | 1.12, 1.86 |
| | $30,000-49,999 | 1.79 | 1.38, 2.31 |
| | $50,000-100,000 | 2.05 | 1.60, 2.62 |
| | >$100,000 | 2.65 | 1.98, 3.54 |
| | Declined to answer | 1.57 | 1.23, 2.00 |
| **Disability** | | | |
| | Mild | 1.44 | 1.12, 1.87 |
| | Moderate | 1.11 | 1.38, 2.31 |
| | Severe (Reference[b]) | 1.0 | |

[a]c-statistic = 0.63; HLGOF $\chi^2_8$= 13.5, *P*=.09

[b] reference group in the regression model

**Table 4.** Type of health information sought about MS at the time of the most recent search.

| Type of health information sought | n/N (%) |
|---|---|
| Treatment for MS | 4470/5663 (78.93) |
| General information about MS | 3378/5405 (62.50) |
| Symptoms of MS | 2810/5032 (55.84) |
| Coping with MS | 2775/5031 (55.16) |
| Complementary and alternative therapies | 2222/4808 (46.21) |
| MS organizations | 1962/4616 (42.50) |
| Cause of MS | 1914/4742 (40.36) |
| Prognosis | 1701/4610 (36.90) |
| Diagnosis of MS | 812/4253 (19.09) |
| Other information | 976/3755 (25.99) |

**Table 5.** Demographic and clinical characteristics associated with trust ("some" or "a lot") versus lack of trust in sources of health information.

| Characteristic | | Information Source | | |
|---|---|---|---|---|
| | | Physician[a] (Nn=7559) | Internet[b] (N=7380) | Patient advocacy group[c] (N=7414) |
| | | OR (95% CI) | OR (95% CI) | OR (95% CI) |
| **Sex** | | | | |
| | Female | 1.0 | 1.0 | 1.0 |
| | Male | 1.29 (0.84, 1.98) | 0.72 (0.62, 0.82) | 1.05 (0.94, 1.18) |
| **Age group, years** | | | | |
| | 18-34 | 0.54 (0.21, 1.40) | 1.04 (0.68, 1.59) | 2.54 (1.75, 3.67) |
| | 35-49 | 1.24 (0.75, 2.05) | 1.42 (1.19, 1.68) | 1.55 (1.35, 1.77) |
| | 50-59 | 1.02 (0.70, 1.48) | 1.50 (1.30, 1.73) | 1.40 (1.26, 1.56) |
| | ≥60 (Reference[d]) | 1.0 | 1.0 | 1.0 |
| **Race** | | | | |
| | White | 1.0 | 1.0 | 1.0 |
| | Other | 1.00 (0.46, 2.18) | 1.05 (0.78, 1.40) | 0.99 (0.79, 1.25) |
| **Education** | | | | |
| | High school or less | 1.0 | 1.0 | 1.0 |
| | Associate's/Technical degree | 1.50 (0.92, 2.44) | 1.31 (1.10, 1.57) | 1.08 (0.94, 1.25) |
| | Bachelor's degree | 1.88 (1.15, 3.00) | 1.37 (1.17, 1.61) | 1.28 (1.13, 1.45) |
| | Graduate degree | 1.38 (0.86, 2.22) | 1.30 (1.10, 1.55) | 1.25 (1.13, 1.45) |
| **Annual income** | | | | |
| | <$15,000 | 1.0 | 1.0 | 1.0 |
| | $15,000-29,999 (Reference[d]) | 1.43 (0.85, 2.42) | 1.41 (1.11, 1.79) | 1.04 (0.85, 1.28) |
| | $30,000-49,999 | 3.45 (1.75, 6.78) | 1.68 (1.31, 2.16) | 1.16 (0.94, 1.43) |
| | $50,000-100,000 | 2.80 (1.51, 5.19) | 1.97 (1.53, 2.52) | 1.33 (1.08, 1.63) |
| | >$100,000 | 2.00 (0.99, 4.03) | 2.17 (1.62, 2.89) | 1.39 (1.10, 1.74) |
| | Declined to answer | 2.02 (1.17, 3.51) | 1.42 (1.12, 1.79) | 0.89 (0.72, 1.08) |
| **Insurance** | | | | |
| | Public only | 1.0 | 1.0 | 1.0 |
| | Private | 1.11 (0.76, 1.63) | 1.04 (0.69, 1.57) | 1.11 (0.81, 1.53) |
| | None | 1.65 (0.51, 5.38) | 0.98 (0.84, 1.13) | 1.10 (0.98, 1.24) |
| **Region** | | | | |
| | West | 1.0 | 1.0 | 1.0 |
| | East | 1.08 (0.84, 1.74) | 0.88 (0.74, 0.82) | 1.04 (0.94, 1.18) |
| | Midwest | 1.17 (0.67, 1.87) | 0.84 (0.71, 1.04) | 1.06 (0.91, 1.21) |
| | South | 0.98 (0.63, 1.54) | 1.07 (0.90, 1.27) | 1.23 (0.93, 1.40) |
| **Disability** | | | | |
| | Mild | 1.34 (0.87, 2.25) | 1.21 (1.04, 1.41) | 1.28 (1.14, 1.44) |
| | Moderate | 1.09 (0.72, 1.75) | 1.07 (0.92, 1.25) | 1.08 (0.96, 1.22) |
| | Severe (Reference[d]) | 1.0 | 1.0 | 1.0 |

[a]c-statistic = 0.66, Hosmer Lemeshow Goodness of Fit (HLGOF) $\chi^2_8 = 7.16$, $P$=.52

[b]c-statistic = 0.61, HLGOF $\chi^2_8 = 8.45$, $P$=.39

[c]c-statistic = 0.61, HLGOF $\chi^2_8$ = 15.3, *P*=.054

[d]reference group in the regression model

## Discussion

### Principal Results

We investigated information sources used by people with MS, satisfaction and trust in those information sources, and the type of information sought about MS. In 2011, more than 85% of our participants reported Internet access. Four years earlier, 69% of American adults who responded to the HINTS survey, which was adapted for our study, reported having access to the Internet [14]. With a frequency of more than 70%, the Internet was the most common first source of general health information reported by study participants at the time of their most recent search, reported 8-fold more often than health care providers, and 13-fold more often than patient advocacy organizations. Although the Internet was a slightly less frequent first choice of information for MS, it was still frequently used (59.23%, 4369/7376). Similarly, one clinic-based study of 96 people with MS from Israel found that 63% used the Internet for MS related searches [5]. In the general population, the Internet is also the first source of cancer information, and this is rising, from just under 50% in 2002-2003 to over 55% in 2008 [15]. In the general population it is also notable that the Internet is less often the first source of information for cancer, than for general health information, similar to our findings for MS [11]. The reasons for these differences are uncertain but the greater complexity of disease-specific information might drive patients to seek more information from their health care providers. Prior studies of the MS population have found that those who used the Internet were younger and more educated than those who did not [5], as have studies of cancer populations [16]. Similarly, we found that higher levels of income, having private health insurance, and white race were associated with more Internet use. We did not evaluate whether this related to computer access, but this should be explored.

Although the Internet was the first source of information reported, 80.02% (6796/8493) of our participants reported a lot of trust in physicians while only 22.63% (1872/8271) reported a lot of trust in the Internet. This is similar to findings in the general population. In the 2008 HINTS survey, trust in physicians was higher with 80% of respondents from the general population reporting the highest trust in information from that source, and only 20% reporting a lot of trust in the Internet [15]. The Internet remained, however, the first information source for more than 55% of respondents.

We found that participants conducted multiple health-related activities online in addition to information seeking (Multimedia Appendix 5). In 2003, the first HINTS survey found that 3.9% of their respondents participated in an online support group and 9.1% bought medicine or supplements online [11]. In 2007, the HINTS survey found that 5% of their respondents participated in an online support group, 14.5% bought medicine online, and 23% used a social networking site [14]. In 2011, NARCOMS participants reported substantially higher use of the Internet for such activities, with 20% using online support groups, 34% buying medication or supplements online, and more than 60% participating in social networking. Although use of the Internet for health reasons is increasing in the general population, these findings suggest it may be higher among people with MS.

The range of information sought regarding MS by NARCOMS participants highlights the varied information needs of people with MS, some of which are likely to change over the disease course. In our study, 18-24% of participants sought information regarding topics related to access to care, including where to get care and how to pay for it. Such information was most often sought by people of lower socioeconomic status. This underscores the economic challenges associated with MS care. In a sample of 2156 people with MS, lower socioeconomic status including lower family income and lack of health insurance, was associated with a lower probability of receiving care from a neurologist [17]. Lower socioeconomic status was also associated with disparities in care for urinary symptoms [18] and mental comorbidity [19].

### Comparison With Prior Work

A mixed methods study of 61 people attending a first visit at an MS clinic found that 82% sought information on the Internet before that appointment [4], highlighting that information seeking often begins even prior to a confirmed diagnosis. These individuals reported doing Internet searches to gather background information, save time during appointments, to verify physician competency, and to find an MS physician and to obtain social support. In an Australian study of 23 people newly diagnosed with MS, most wanted information from the MS Society and MS specialist nurses. They were happy with the amount of information received from those sources, but wanted more information than they currently received from neurologists, family physicians, and education sessions [20]. Thus people with MS frequently seek many types of information from a range of sources, but remain dissatisfied with the amount of information obtained from some sources. This is consistent with the disparity between the first choice of information source and the most trusted information source noted in the present study. Further, despite the readily available information from sources such as the Internet, 39.86% (2998/7521) of participants in our study had concerns regarding information quality, and 21.11% (1590/7553) found the information difficult to comprehend. Participants of lower socioeconomic status reported more difficulty finding and understanding information.

Collectively, these findings have important implications for dissemination of information to people with MS and their families. Health care providers should be aware that their patients are likely to gather considerable information from the Internet, and typically before their patients obtain information from their providers. Further, although patients trust the information obtained from their physicians, they likely want more information from them, and in a more timely fashion. Traditional information sources such as newspapers, television, and radio are likely to be ineffective methods of communication as they are either not used, not trusted, or both. The Internet, including social media, provides a means for rapid dissemination

of health information by health care providers, which can be readily updated. However, it is critical that this information be provided in a way that is readily accessible and comprehensible to people of all socioeconomic backgrounds.

### Limitations

This study has limitations. Our response rate of 66.18% (8586/12,974) was lower than desired, but this may be an underestimate of the true response rate since we could not determine how many participants actually received the questionnaire as changes in home or email addresses and deaths are often identified only long after a participant has failed to respond. The NARCOMS Registry is a voluntary registry, therefore it does not fully represent the MS population in the United States, and non-responders had a lower socioeconomic status than responders. However, the NARCOMS population is large and sociodemographically diverse with characteristics similar to those reported for other MS populations [21]. A key strength of this study was that participants included those who responded online and by mail, avoiding the limitations of other studies about Internet use which have been limited to online only completion. Further, this was the largest study related to this issue that we know of.

### Conclusions

Our work highlights the main sources of health information used by people with MS, and thus has implications for the dissemination of health information, keeping in mind the discordance between the most readily accessible source, the Internet and the most trusted resource. The rise in the use of social networking, and platforms facilitating direct exchange of personal health information between patients [22,23] are dramatically changing patterns of health communication.

### Acknowledgments

This study was supported (in part) by a Don Paty Career Development Award from the MS Society of Canada (to RAM). The NARCOMS Registry is supported (in part) by the Consortium of Multiple Sclerosis Centers and its Foundation. The sponsors had no role in the design or conduct of the study; the interpretation or analysis of the data; or in the decision to submit for publication.

### Conflicts of Interest

Ruth Ann Marrie receives research funding from: Canadian Institutes of Health Research, Public Health Agency of Canada, Manitoba Health Research Council, Health Sciences Centre Foundation, Multiple Sclerosis Society of Canada, Multiple Sclerosis Scientific Foundation, Rx & D Health Research Foundation, has conducted clinical trials funded by Bayer Inc. and Sanofi-Aventis, and serves on the Editorial Boards of Neurology and Multiple Sclerosis Journal. Amber Salter has no conflicts of interest to declare. Tuula Tyry has no conflicts of interest to declare. Robert Fox has received personal consulting fees from Avanir, Biogen Idec, Novartis, Questcor, and Teva Neurosciences; has served on clinical trial advisory committees for Biogen Idec and Novartis; has received research support from the National Multiple Sclerosis Society (RG 4091A3/1; RG 4103A4/2; RC 1004-A-5) and Novartis; and serves on the editorial boards of Neurology and Multiple Sclerosis Journal. Gary Cutter has served on scientific advisory boards for and/or received funding for travel from Alexion, Allozyne, Bayer, Celgene, Consortium of MS Centers, Coronado Biosciences, Diogenix, Klein-Buendel Incorporated, Merck, Novartis, Nuron Biotech, Receptos, Somnus Pharmaceuticals, Spinifex Pharmaceuticals, St. Louis University, Teva pharmaceuticals; receives royalties from publishing Evaluation of Health Promotion and Disease Prevention (The McGraw Hill Companies, 1984); has received honoraria from GlaxoSmithKline, Novartis, Advanced Health Media Inc, Biogen Idec, EMD Serono Inc, EDJ Associates, Inc, the National Heart, Lung, and Blood Institute, National Institute of Neurological Diseases and Stroke, National Marrow Donor Program, Consortium of Multiple Sclerosis Centers; serves as a consultant to Novartis, National Industrial Sand Association, Bayer Pharmaceuticals, and Teva Pharmaceuticals Industries Ltd.; has served on independent data and safety monitoring committees for Apotek, Biogen, Cleveland Clinic, EliLilly, Glaxo Smith Klein Pharmaceuticals, Medivation, Modigenetech, NHLBI, NINDS, NMSS, Ono Pharmaceuticals, Prolor, Sanofi-Aventis, Teva.

### Multimedia Appendix 1

Questions adapted from the Health Information Trends Survey (HINTS).

[PDF File (Adobe PDF File), 49KB - jmir_v15i4e67_app1.pdf ]

### Multimedia Appendix 2

Demographic and clinical characteristics associated with seeking general information regarding MS, using binary logistic regression.

[PDF File (Adobe PDF File), 39KB - jmir_v15i4e67_app2.pdf ]

### Multimedia Appendix 3

Demographic and clinical characteristics associated with seeking information regarding access to care using binary logistic regression.

[PDF File (Adobe PDF File), 39KB - jmir_v15i4e67_app3.pdf ]

### Multimedia Appendix 4

Demographic and clinical characteristics associated with internet use using binary logistic regression (n=7553).

[PDF File (Adobe PDF File), 33KB - jmir_v15i4e67_app4.pdf ]

### Multimedia Appendix 5

Health-related online activities reported by the NARCOMS population and respondents to the HINTS surveys.

[PDF File (Adobe PDF File), 29KB - jmir_v15i4e67_app5.pdf ]

## References


1. Couper MP, Singer E, Levin CA, Fowler FJ, Fagerlin A, Zikmund-Fisher BJ. Use of the Internet and ratings of information sources for medical decisions: results from the DECISIONS survey. Med Decis Making 2010;30(5 Suppl):106S-114S. [doi: 10.1177/0272989X10377661] [Medline: 20881159]
2. Burkell JA, Wolfe DL, Potter PJ, Jutai JW. Information needs and information sources of individuals living with spinal cord injury. Health Info Libr J 2006 Dec;23(4):257-265. [doi: 10.1111/j.1471-1842.2006.00686.x] [Medline: 17177946]
3. Garneau K, Iversen M, Jan S, Parmar K, Tsao P, Solomon DH. Rheumatoid arthritis decision making: many information sources but not all rated as useful. J Clin Rheumatol 2011 Aug;17(5):231-235. [doi: 10.1097/RHU.0b013e318226a220] [Medline: 21778909]
4. Hay MC, Strathmann C, Lieber E, Wick K, Giesser B. Why patients go online: multiple sclerosis, the internet, and physician-patient communication. Neurologist 2008 Nov;14(6):374-381. [doi: 10.1097/NRL.0b013e31817709bb] [Medline: 19008743]
5. Lejbkowicz I, Paperna T, Stein N, Dishon S, Miller A. Internet usage by patients with multiple sclerosis: implications to participatory medicine and personalized healthcare. Mult Scler Int 2010;2010:640749 [FREE Full text] [doi: 10.1155/2010/640749] [Medline: 22096625]
6. NARCOMS Multiple Sclerosis Patient Registry. Birmingham: University of Alabama at Birmingham School of Public Health URL: http://www.narcoms.org/ [accessed 2013-03-08] [WebCite Cache ID 6Eyninh11]
7. Marrie RA, Cutter G, Tyry T, Campagnolo D, Vollmer T. Validation of the NARCOMS registry: diagnosis. Mult Scler 2007 Jul;13(6):770-775. [doi: 10.1177/1352458506075031] [Medline: 17525097]
8. Marrie RA, Goldman M. Validity of performance scales for disability assessment in multiple sclerosis. Mult Scler 2007 Nov;13(9):1176-1182. [doi: 10.1177/1352458507078388] [Medline: 17623733]
9. Nelson DE, Kreps GL, Hesse BW, Croyle RT, Willis G, Arora NK, et al. The Health Information National Trends Survey (HINTS): development, design, and dissemination. J Health Commun 2004;9(5):443-60; discussion 81. [doi: 10.1080/10810730490504233] [Medline: 15513791]
10. Ford BM, Kaphingst KA. Lay interpersonal sources for health information related to beliefs about the modifiability of cancer risk. Cancer Causes Control 2009 Dec;20(10):1975-1983 [FREE Full text] [doi: 10.1007/s10552-009-9392-1] [Medline: 19578935]
11. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Arch Intern Med 2005;165(22):2618-2624. [doi: 10.1001/archinte.165.22.2618] [Medline: 16344419]
12. Marrie RA, Cutter G, Tyry T, Vollmer T, Campagnolo D. Does multiple sclerosis-associated disability differ between races? Neurology 2006 Apr 25;66(8):1235-1240. [doi: 10.1212/01.wnl.0000208505.81912.82] [Medline: 16636241]
13. Hosmer DW, Lemeshow S. Applied Logistic Regression. In: New York: John Wiley & Sons. New York: John Wiley & Sons, Inc; 1989:140-145.
14. Chou WY, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
15. Hesse BW, Moser RP, Rutten LJ. Surveys of physicians and electronic health information. N Engl J Med 2010 Mar 4;362(9):859-860. [doi: 10.1056/NEJMc0909595] [Medline: 20200398]
16. Muusses LD, van Weert JC, van Dulmen S, Jansen J. Chemotherapy and information-seeking behaviour: characteristics of patients using mass-media information sources. Psychooncology 2012 Sep;21(9):993-1002. [doi: 10.1002/pon.1997] [Medline: 21671305]

17. Minden SL, Hoaglin DC, Hadden L, Frankel D, Robbins T, Perloff J. Access to and utilization of neurologists by people with multiple sclerosis. Neurology 2008 Mar 25;70(13 Pt 2):1141-1149. [doi: 10.1212/01.wnl.0000306411.46934.ef] [Medline: 18362274]
18. Marrie RA, Cutter G, Tyry T, Vollmer T, Campagnolo D. Disparities in the management of multiple sclerosis-related bladder symptoms. Neurology 2007 Jun 5;68(23):1971-1978. [doi: 10.1212/01.wnl.0000264416.53077.8b] [Medline: 17548546]
19. Marrie RA, Horwitz R, Cutter G, Tyry T, Campagnolo D, Vollmer T. The burden of mental comorbidity in multiple sclerosis: frequent, underdiagnosed, and undertreated. Mult Scler 2009 Mar;15(3):385-392. [doi: 10.1177/1352458508099477] [Medline: 19153176]
20. Matti AI, McCarl H, Klaer P, Keane MC, Chen CS. Multiple sclerosis: patients' information sources and needs on disease symptoms and management. Patient Prefer Adherence 2010;4:157-161 [FREE Full text] [Medline: 20622916]
21. Jacobs LD, Wende KE, Brownscheidle CM, Apatoff B, Coyle PK, Goodman A, et al. A profile of multiple sclerosis: the New York State Multiple Sclerosis Consortium. Mult Scler 1999 Oct;5(5):369-376. [Medline: 10516782]
22. Frost J, Okun S, Vaughan T, Heywood J, Wicks P. Patient-reported outcomes as a source of evidence in off-label prescribing: analysis of data from PatientsLikeMe. J Med Internet Res 2011;13(1):e6 [FREE Full text] [doi: 10.2196/jmir.1643] [Medline: 21252034]
23. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]

## Abbreviations

**EDSS:** Expanded Disability Status Scale
**HINTS:** Health Information National Trends
**MS:** multiple sclerosis
**NARCOMS:** North American Research Committee on Multiple Sclerosis registry
**NMSS:** National Multiple Sclerosis Society
**PDDS:** Patient Determined Disease Steps

*Edited by G Eysenbach; submitted 02.12.12; peer-reviewed by R Kinkel, B Giesser, P Jongen; comments to author 25.01.13; revised version received 28.01.13; accepted 13.02.13; published 30.04.13.*

*Please cite as:*
*Marrie RA, Salter AR, Tyry T, Fox RJ, Cutter GR*
*Preferred Sources of Health Information in Persons With Multiple Sclerosis: Degree of Trust and Information Sought*
*J Med Internet Res 2013;15(4):e67*
*URL: http://www.jmir.org/2013/4/e67/*
*doi:10.2196/jmir.2466*
*PMID:23635393*

Original Paper

# A Data Encryption Solution for Mobile Health Apps in Cooperation Environments

Bruno M Silva[1], BSC, MSc; Joel JPC Rodrigues[1], BSC, MSc, PhD; Fábio Canelo[1], BSC, Msc Candidate; Ivo C Lopes[1], BSC, MSc; Liang Zhou[2], BSC, MSc, PhD

[1]Instituto de Telecomunicações, University of Beira Interior, Covilhã, Portugal

[2]Nanjing University of Posts and Telecommunications, Nanjing, China

**Corresponding Author:**
Joel JPC Rodrigues, BSC, MSc, PhD
Instituto de Telecomunicações, University of Beira Interior
Rua Marques D'Avila e Bolama
Covilhã, 6201-001
Portugal
Phone: 351 275242081
Fax: 351 275319891
Email: joeljr@ieee.org

## Abstract

**Background:** Mobile Health (mHealth) proposes health care delivering anytime and anywhere. It aims to answer several emerging problems in health services, including the increasing number of chronic diseases, high costs on national health services, and the need to provide direct access to health services, regardless of time and place. mHealth systems include the use of mobile devices and apps that interact with patients and caretakers. However, mobile devices present several constraints, such as processor, energy, and storage resource limitations. The constant mobility and often-required Internet connectivity also exposes and compromises the privacy and confidentiality of health information.

**Objective:** This paper presents a proposal, construction, performance evaluation, and validation of a data encryption solution for mobile health apps (DE4MHA), considering a novel and early-proposed cooperation strategy. The goal was to present a robust solution based on encryption algorithms that guarantee the best confidentiality, integrity, and authenticity of users health information. In this paper, we presented, explained, evaluated the performance, and discussed the cooperation mechanisms and the proposed encryption solution for mHealth apps.

**Methods:** First, we designed and deployed the DE4MHA. Then two studies were performed: (1) study and comparison of symmetric and asymmetric encryption/decryption algorithms in an mHealth app under a cooperation environment, and (2) performance evaluation of the DE4MHA. Its performance was evaluated through a prototype using an mHealth app for obesity prevention and cares, called SapoFit. We then conducted an evaluation study of the mHealth app with cooperation mechanisms and the DE4MHA using real users and a real cooperation scenario. In 5 days, 5 different groups of 7 students selected randomly agreed to use and experiment the SapoFit app using the 7 devices available for trials.

**Results:** There were 35 users of SapoFit that participated in this study. The performance evaluation of the app was done using 7 real mobile devices in 5 different days. The results showed that confidentiality and protection of the users' health information was guaranteed and SapoFit users were able to use the mHealth app with satisfactory quality. Results also showed that the app with the DE4MHA presented nearly the same results as the app without the DE4MHA. The performance evaluation results considered the probability that a request was successfully answered as a function of the number of uncooperative nodes in the network. The service delivery probability decreased with the increase of uncooperative mobile nodes. Using DE4MHA, it was observed that performance presented a slightly worse result. The service average was also slightly worse but practically insignificantly different than with DE4MHA, being considered negligible.

**Conclusions:** This paper proposed a data encryption solution for mobile health apps, called DE4MHA. The data encryption algorithm DE4MHA with cooperation mechanisms in mobile health allow users to safely obtain health information with the data being carried securely. These security mechanisms did not deteriorate the overall network performance and the app, maintaining similar performance levels as without the encryption. More importantly, it offers a robust and reliable increase of privacy, confidentiality, integrity, and authenticity of their health information. Although it was experimented on a specific mHealth app, SapoFit, both DE4MHA and the cooperation strategy can be deployed in other mHealth apps.

*(J Med Internet Res 2013;15(4):e66)* doi:10.2196/jmir.2498

**KEYWORDS**

mobile health; mHealth; mobile computing; eHealth; cooperation; encryption; security

## Introduction

In the last decade, health telematics, also known as electronic health (eHealth), have offered patients major improvements in their lives by providing more accessible and affordable health care solutions [1,2]. This is particularly true for patients that live in remote rural areas, travel constantly, are physically incapacitated, elderly, or chronically ill. Telemedicine assumes the use of medical information, also known as electronic health records (EHRs), exchanged via electronic communications improving the patients' health status [3]. The rapid evolution of information and communication technology (ICT) infrastructures enables and provides rapid access to patient data. The Web 2.0 concept and the emerging Web 3.0 offer opportunities to health care professionals never seen before [4,5]. Now, physicians can perform many tasks through these modern technologies, such as (1) sharing medical videos, photos, and presentations (via YouTube, Flickr, and Slideshare, respectively), (2) use blogs to post medical cases and images, (3) share hospital management information, (4) use social networking to share medical ideas and tasks, and (5) use RSS feeds to keep track of alerts on specific medical interests.

With the advent of mobile communications using smart mobile devices that support 3G and 4G mobile networks for data transport, mobile computing has been the main attraction of research and business communities, thus offering innumerous opportunities to create efficient mobile health solutions. Mobile health (mHealth) is the new edge on health care innovations. It delivers health care anywhere and anytime, surpassing geographical, temporal, and even organizational barriers [6,7]. Laxminarayan and Istepanian defined mobile health for the first time in 2000, as "unwired e-med" [8]. In 2003, the term "mHealth" was defined as the "emerging mobile communications and network technologies for health care systems" [9]. Laxminarayan et al, in 2006, presented a comprehensive study on the impact of mobility on the existing eHealth commercial telemedical systems. They also presented other relevant computing and information technologies that will influence and offer the basis for the next generation of mHealth services [10]. Furthermore, this study served as the basis for future studies on mHealth technologies and services [11]. Several research topics related to health have gathered important findings and contributions using mHealth, such as cardiology [12,13], diabetes [14-16], obesity [17-20], and smoking cessation [21]. More specifically, mHealth apps were applied to health monitoring, disease prevention and detection, basic diagnosis, and in more advanced services. mHealth services are also becoming popular in developing countries where health care facilities are frequently remote and inaccessible [2,22].

Mobile devices and wireless communications present several challenging characteristics and constraints, such as battery and storage capacity, broadcasting constraints, signal interferences, disconnections, noises, limited bandwidths, and network delays. In this sense, cooperation-based approaches are presented as a solution to solve such limitations, focusing on increasing network connectivity, communication rates, and reliability.

In this paper, we present a data encryption solution for mHealth apps (DE4MHA) in cooperative environments guaranteeing data confidentiality, integrity, and authenticity. This novel and early-proposed cooperation strategy [23] for mHealth apps focuses on forwarding and retrieving data to and from nodes that have no direct connection to an mHealth service. In this way, devices without Internet connectivity can use mHealth apps without problems. This cooperation approach presents a reputation-based strategy where a Web service manages the access control and the cooperation among nodes along with their reputation. It considers the following three main components: a *node control message*, a *requester control message*, and a *cooperative Web service* (CWS). Both control messages are used to manage a local cooperation between two or more nodes. The CWS includes a reputation table for all the nodes and decides which nodes can have access to the requested services. The cooperation strategy and the DE4MHA was deployed and evaluated in an mHealth app for obesity prevention and control, called SapoFit [24-26]. To the best of our knowledge, there are no cooperative solutions thus far for mHealth services and apps considering this network scenario with constant network disconnection. DE4MHA uses symmetric and asymmetric encryption and decryption techniques. We used the Rivest, Shamir, Adleman (RSA) algorithm [27] for asymmetric encryption/decryption to ensure key exchange confidentiality, and the Advanced Encryption Standard (AES) [28] algorithm for symmetric encryption/decryption for data confidentiality. To ensure data integrity, we have created a message digest that creates a hash of transmitted data. For data authenticity, we used a digital signature. We encrypted the hash message with the RSA private key. To secure the communication with the SapoFit Web service (WS), we used the Hypertext Transfer Protocol Secure (HTTPS) protocol.

In this paper we report two studies that were performed to design and construct the DE4MHA algorithms: (1) a direct evaluation and comparison of several encryption algorithms, and (2) a series of trials evolving 35 people and 7 different mobile devices with SapoFit. The first study revealed what algorithms performed best in an mHealth app in cooperation environments. Overall, this study evaluated the performance of the DE4MHA over the cooperation mechanisms for mHealth apps. The second study revealed that real users experimenting on the SapoFit app trusted DE4MHA. More relevant, this study concluded that the performance of the app used was not affected by the inclusion of DE4MHA.

## Methods

### Overview

This study used an existing mHealth app, called SapoFit, to deploy, evaluate, and validate the proposed solution. This app uses a cooperation strategy that addresses two related limitations to mHealth apps with service-oriented architectures, namely the network infrastructure and Internet connectivity dependency. It follows a reputation-based approach as an incentive method for cooperation, which includes a Web service to manage all the network cooperation. It is responsible for verifying the cooperation status of neighbor nodes and to provide relay nodes the required data in order to perform a full data request.

### Cooperation Strategy for mHealth Apps

The cooperation strategy for mHealth apps with service oriented architectures (SOAs) is based on the following two mobile modules and one remote module, respectively: (1) the *node control message*, (2) the *requester control message*, and (3) the *CWS*.

The mobile *nodes control messages* aim to provide an awareness of the relay node status, that is, if the node is willing to cooperate and in what conditions. It contains the established node unique identifier, the battery state, the Internet connectivity status, and the cooperation status (ie, if it is cooperative or not).

The *requester control message* is sent by the initial requester node first (the mobile device with mHealth app requesting health data), and it comprises the following five main components: (1) the requester ID, the node unique identifier, (2) the service request, that is, what the node is specifically requesting (eg, the login token or its health profile), (3) the neighbors list, (4) the reputation list, and (5) the achieved cooperation time (ACT).

The *CWS* is responsible for performing a fair access control to data. Thus, according to the received reputation information, the Web service holds the final reputation list in order to decide if a requester node should have access to the mHealth app Web service or not. The reputation list contains all registered network nodes with their identifier and their corresponding reputation value.

Figure 1 presents a user scenario of the mHealth cooperation approach. *User A* has network connectivity and cooperates. *User B* has network connectivity and does not cooperate. The status value is according to the battery status. Then, the status value will suffer a negative impact according to the battery status. *Users C* and *D* do not have network connectivity. *User C* queries *User A* for cooperation and receives a positive response and all the requested data. *User D* queries *User B* for cooperation and receives a negative response. Then, *User D* requests data from *User C* that answers this request, getting positive status by cooperating.

### SapoFit App

SapoFit is a weight control mobile app that allows users to keep track of weight in a healthier and more practical way. SapoFit allows users to control their weight, body mass index (BMI), basal metabolic rate (BMR), sports activity, and the possibility to follow food plans based on their needed calories. In this mHealth app, all the users must be registered in a Web service. Figure 2 presents screenshots of three main activities of the SapoFit app: *Login*, *Plans*, and *User Profile.*

Cooperating nodes have a better reputation, and have priority over selfish nodes to access the mHealth app services.

### Data Encryption Algorithm for Mobile Health Apps (DE4MHA)

The process begins with a mobile node (a person using SapoFit) trying to access the SapoFit Web Service that contains the user profile, weight measures, fitness, and diet indications.

A SapoFit user (mobile requester node) without network connectivity and therefore without access to the SapoFit WS obtains the required health information through cooperation. Another SapoFit user with network connectivity (mobile requested node) will forward the requested health information from the SapoFit Web service. Both the requested and requester nodes will create a pair of RSA keys and send public keys to both the requested and requester node through Bluetooth. After the public key exchange, the requested node creates an AES session key.

The next step is the creation of the digest message and its encryption using the private key. The Message Digest 5 (MD5) algorithm was used to create a 128-bit hash. For data authenticity, we used a digital signature. A digital signature is created for the message containing requested health information. This digital signature allows any node to verify that the message is the original one. By decrypting the digital signature with the public key, the original digest message is obtained. The receiver node then creates a new hash of the received message and compares it to the decrypted digest message to guarantee authenticity. The digital signature is then added to the message. When the message containing the session key is received, if its integrity and authenticity is verified, the requester node then sends an acknowledgement (*ack*) to the requested node. This method guaranties safe communication between nodes; if the integrity and authenticity is not verified, the communication between nodes is ended.

A mobile node with network connectivity will access the cooperative WS to obtain the required health information. To secure all communication with the WS the Secure Socket Layer (SSL) over the HTTP (also known as HTTPS) is used. Therefore, it grants confidentiality, integrity, and authenticity of all retrieved health data from the Web service.

Two studies were performed: (1) a study evaluating which symmetric and asymmetric algorithm present the best performance in SapoFit in cooperation environment, and (2) a series of trials involving 35 people and 7 different mobile devices with SapoFit. This study evaluated the performance of the DE4MHA over the cooperation mechanisms.

**Figure 1.** Illustration of the interaction for an mHealth app with the proposed cooperation approach for 4 users.

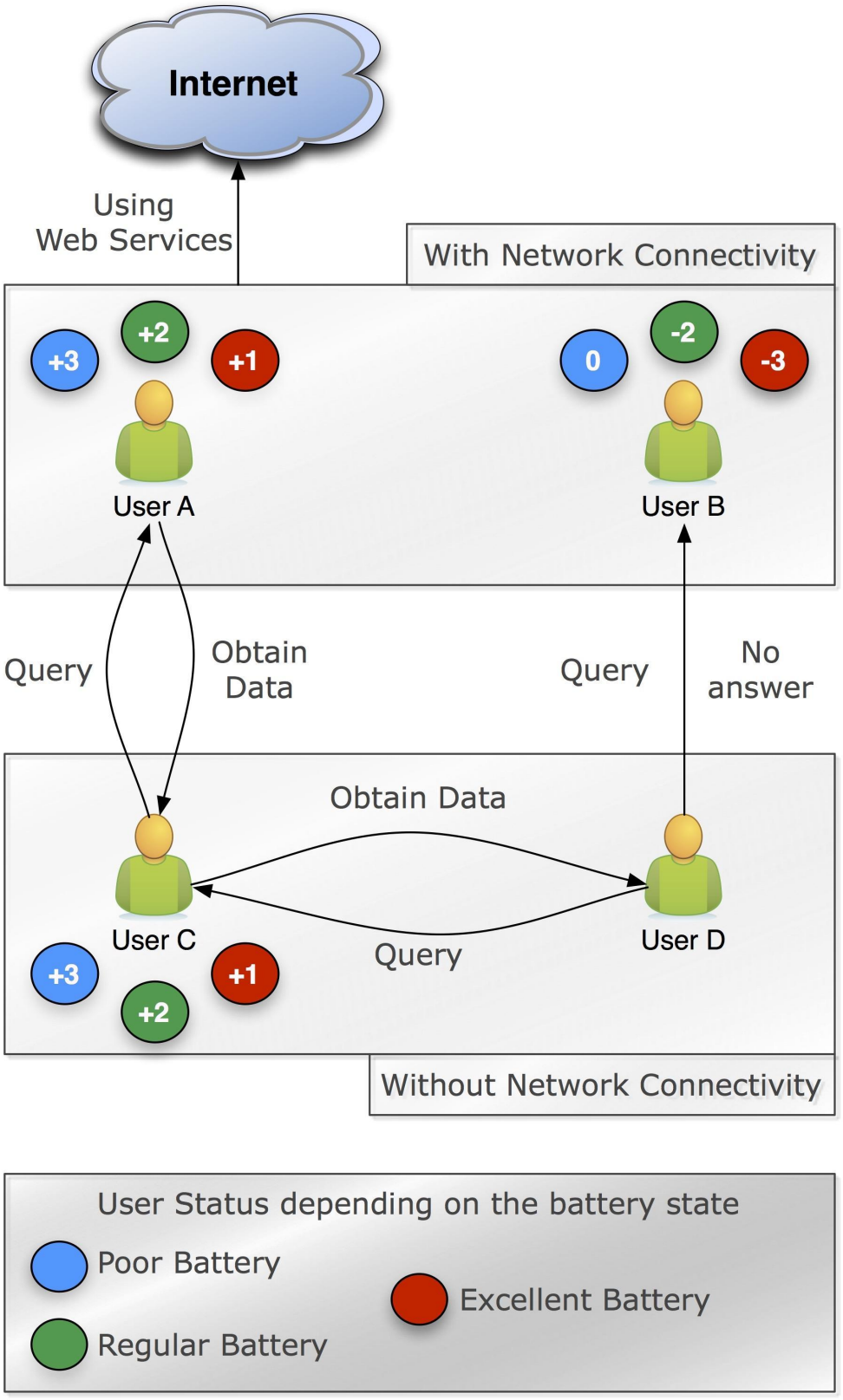
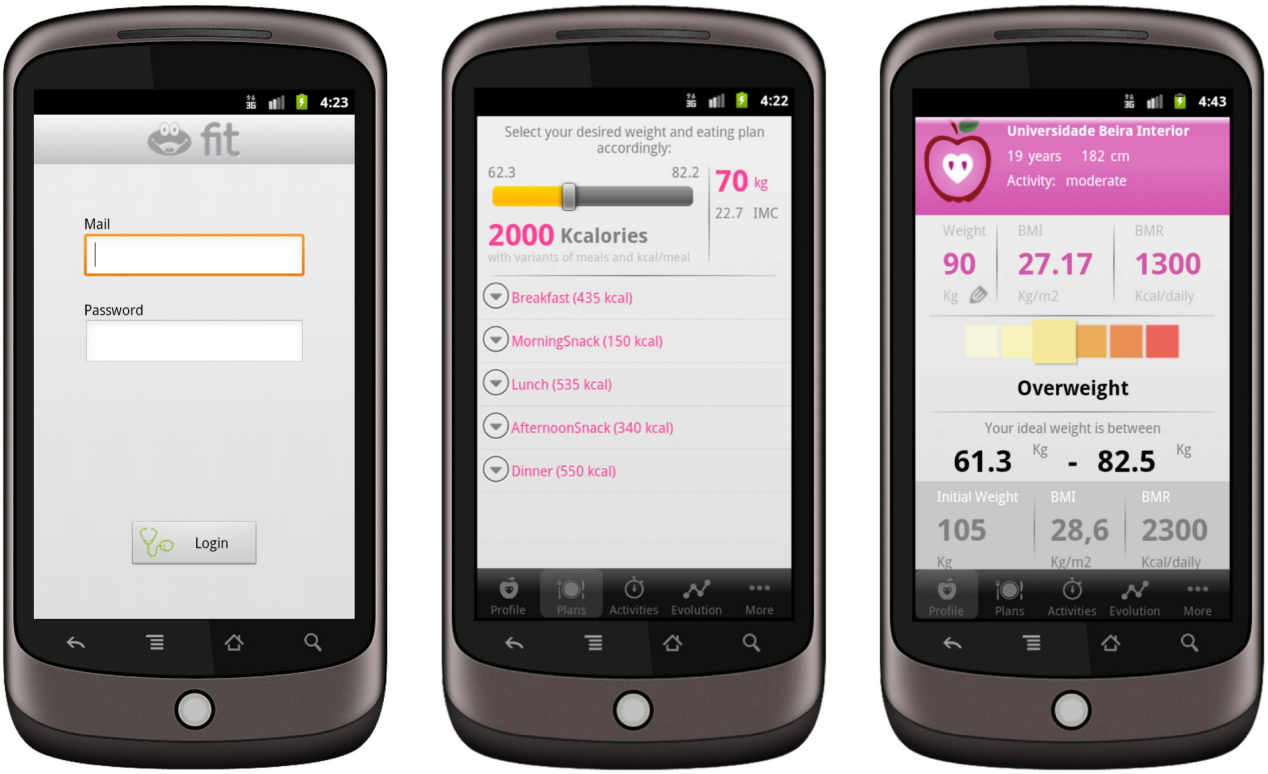


**Figure 2.** Screenshots of the three main activities of SapoFit app: Login, Plans, and User Profile.




## Study 1: Study of Cryptography Algorithms in an mHealth App Under a Cooperation Environment

### *Symmetric Algorithm*

In order to choose the best suited symmetric encryption algorithm for the SapoFit app, performance experiments were conducted using 4 different encryption algorithms, namely AES, Triple Data Encryption Standard (3DES) [29], Rivest Cypher 4 (RC4) [30], and Blowfish [31], using the data size encryption as a performance metric.

### *Asymmetric Algorithm*

Two options were considered in selecting an asymmetric algorithm to exchange session keys between mobile nodes. We tested RSA, which enables encrypting the session key before sending it, and Diffie-Hellman [32], which allows users to first share a secret, then generating a session key based on the shared secret. For our network scenario, these were the most suitable algorithms. Other encryption algorithms were considered in this study, such as the Elliptic Curve Cryptography (ECC) algorithm [33].

## Study 2: Performance evaluation of the DE4MHA

### *Performance Evaluation*

The performance evaluation study was carried out using 7 real mobile devices, which were used 7 times in 5 different days with a total of 35 different users who successfully tested the app. Figure 3 presents the 7 different mobile devices with different hardware and software used with the SapoFit mHealth app.

Cooperative nodes without network connection cooperated with each other through Bluetooth. The communication with the CWS was obtained through the Wi-Fi or Edge/3.5G/4G modules of the device. The CWS was developed with the help of Java Server Pages (JSP) technology, using the Representational State Transfer (REST) architecture. To serve the WS, the Apache Tomcat Web Server was used.

Non-cooperative cases were controlled and measured to a maximum of 4 to guarantee the minimum service performance in order to guarantee cooperation among nodes. Through cooperation, all the devices could use the mHealth app. However, uncooperative nodes directly affect the service delivery probability, service average delay, and the overall network performance. Performance metrics considered in this study were the service delivery probability (as percentages) and the average service delay (in seconds). We present a comparison of the performance of the mHealth app with and without the DE4MHA.

### *User Trials Evaluation of DE4MHA in Cooperative Environments*

User trials were conducted within the University informatics department using 7 devices available for trials. Within 5 days, 5 different groups of 7 students selected randomly agreed to use and experiment the SapoFit app using the 7 devices available for trials. Users were constantly moving far away from each other. This mobility was necessary to test the network scenario, forcing network delays and disconnections. The cooperative strategy and the DE4MHA was ubiquitous to its user. Throughout the trials, users only experienced and used the obesity prevention services that SapoFit offered without any

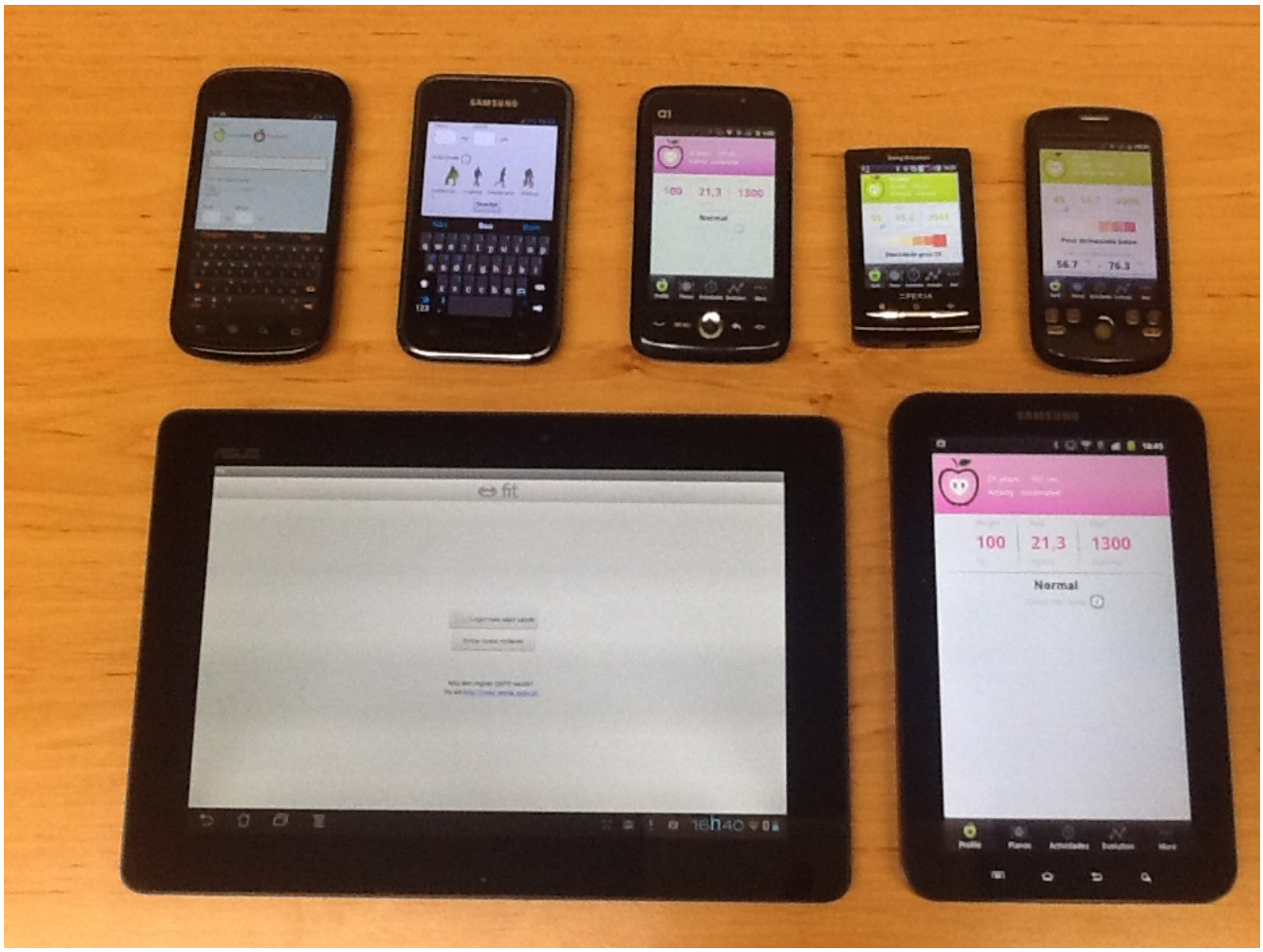
constraints or perception of any cooperation mechanism or encryption algorithm that was embedded in the mHealth app.

While conducting the experiments, almost every users asked if their information was being kept secure or not, clearly showing that they did not wanted their health information to be available or disclosed to unauthorized people, revealing privacy concerns. Another frequently raised question was the need to share Internet connectivity to other users.

We explained and justified that sharing and cooperating with other users was essential to obtain a better reputation to gain cooperation privileges over other nodes with worse reputation. Furthermore we demonstrated to them that SapoFit was not intrusive with other users' personal data on the mobile device and only requested for SapoFit services.

After the experiments, the users completed a survey evaluating their experience. The questions are listed in Textbox 1 and the results in Figure 4.

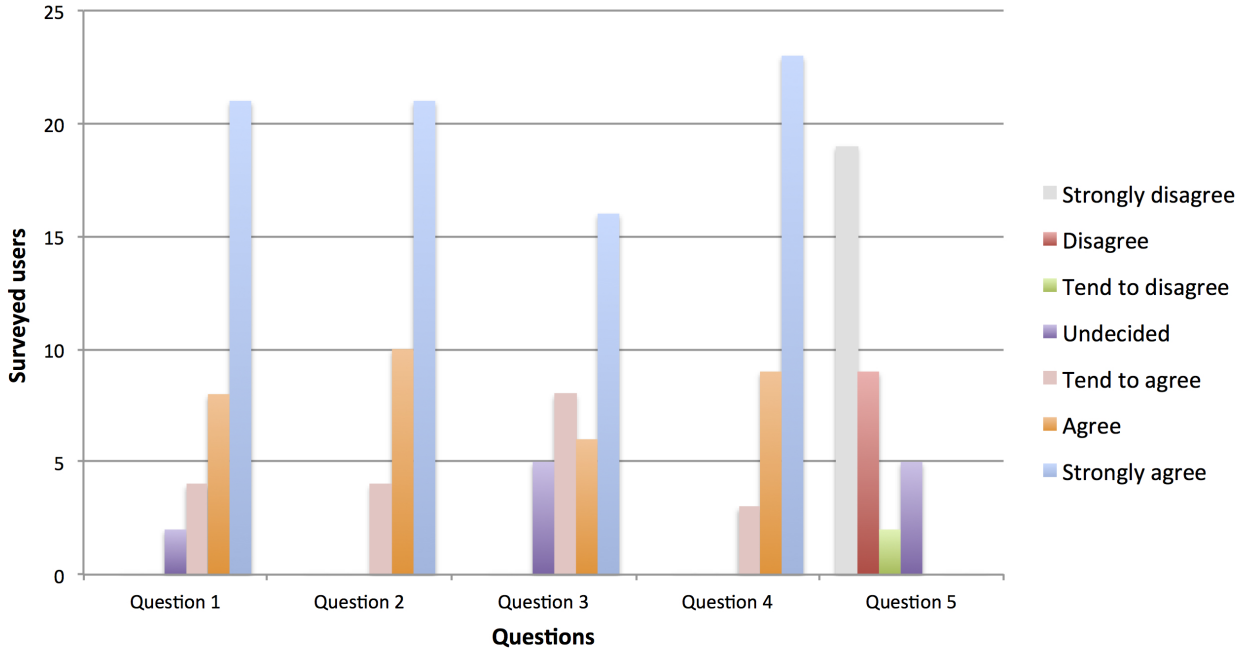
**Figure 3.** Mobile devices used for trials with the SapoFit mHealth app.




**Textbox 1.** Survey questions.

- Without network connectivity, do the user always gets the required information?
- Without network connectivity, does the required information get delivered in a comfortable time?
- Understanding the implemented cooperation strategy and its benefits, are you willing to cooperate and share the device/network resources with other users?
- With the encryption strategy applied to SapoFit, do you trust that your personal health information is secure?
- Was the mobile device affected by app cooperation and encryption mechanisms (eg, broadband, battery, etc)?

**Figure 4.** Results of the survey evaluating the main questions about the performance of the mHealth app with the proposed cooperation strategy and encryption solution.




## Results

### Symmetric Algorithm

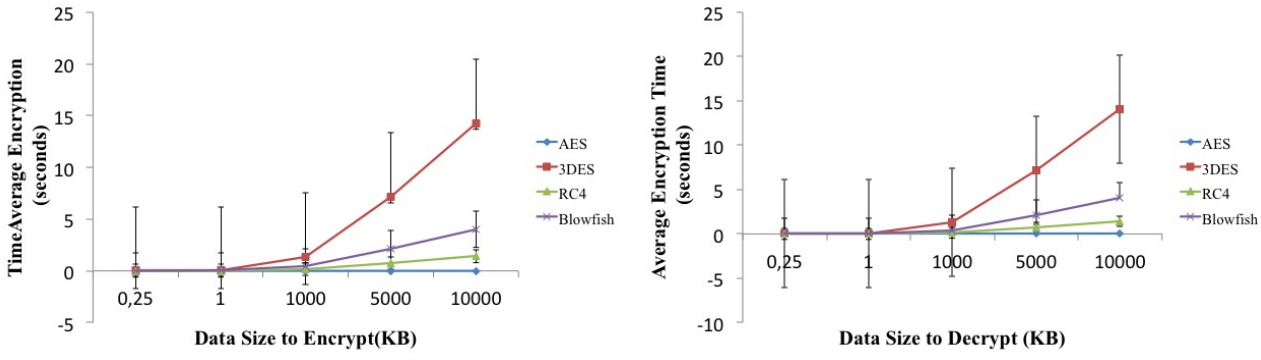
As seen in Figure 5, results showed that when the amount of data that needed to be encrypted increased, the encryption time (in seconds) also increased, as expected. When comparing small amounts of data, all 4 algorithms presented similar results. However, the AES algorithm presented better results, as there was a slower overall increase in encryption time in response to increased amount of data. The encryption time of the other three experimented algorithms grew exponentially when encryption data size exceeded 1000 kilobytes. The 3DES algorithm presented the worst encryption rate, encrypting 10,000KB of data in, on average, 14.3 seconds, presenting an average time of 4.58 seconds for multiple data size inputs (SD 6.17). With the same amount of data, the AES encryption time was only 0.0045 seconds, resulting in an average of 0.0035 seconds for the given data set (SD 0.00061 seconds). Decryption gave similar results (Figure 5). The average AES algorithm decryption time was 0.0038 seconds for 10,000KB of data and 0.0025 seconds on average for the working data-set (SD 0.001 seconds). Overall, the AES algorithm with a 128 bits key encryption was the most efficient algorithm for these network scenarios, when handling with both small and large amounts of data.

### Asymmetric Algorithm

Two options were considered for an asymmetric algorithm in order to exchange session keys between mobile nodes—the RSA and the Diffie-Hellman algorithms. The RSA encrypts the session key first before it gets sent, and Diffie-Hellman allows users to first share a secret, then generates a session key based on the shared secret.

Several experiments were performed with both algorithms with RSA presenting better encryption times than Diffie-Hellman, due to the high amount of computation needed by Diffie-Hellman and the low processing capacity of mobile devices.

**Figure 5.** Comparison of encryption and decryption of symmetric algorithms (AES, 3DES, RC4, and Blowfish).

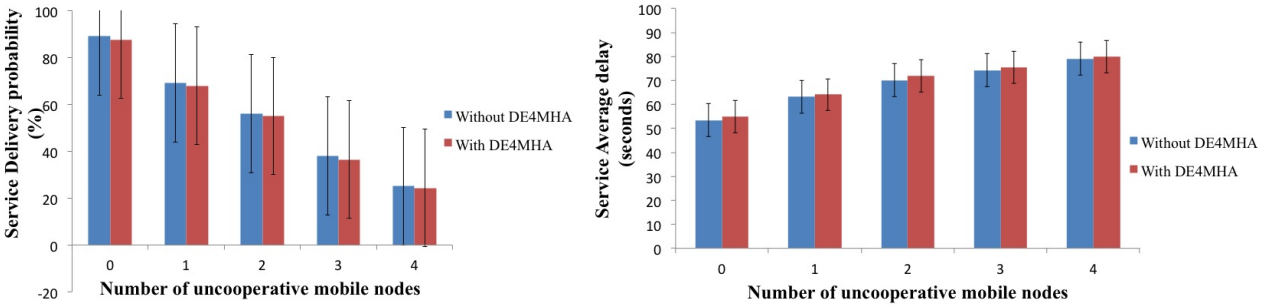


### DE4MHA Performance Evaluation Results

In the presented scenario, all the devices had Bluetooth class 2 modules, but only 3 devices had Internet connectivity. Users without Internet connectivity had to use the integrated cooperation mechanisms in order to obtain the requested health information. When the number of uncooperative nodes increased, the service delivery probability decreased. The average service delay was also affected in the same mannor, as expected. Increased number of uncooperative nodes increased the average service delay.

Figure 6 shows the results of the service delivery probability and the average service delay as a function of the number of uncooperative mobile nodes with and without the DE4MHA. The probability that a request was successfully answered as a function of uncooperative nodes in the network decreased with the increase of uncooperative mobile nodes. Taking into account both approaches, with and without DE4MHA, it was observed that DE4MHA presented a slightly worse result. The average service delay also grew when the number of uncooperative mobile nodes increased, as expected. The results of the DE4MHA were also slightly worse but practically insignificant. As can be observed, with 4 uncooperative nodes, the service average delay, with and without DE4MHA, was about 30.78 and 29.77 seconds, respectively.

**Figure 6.** Service delivery probability and average service delay as a function of the number of uncooperative mobile nodes with and without the DE4MHA.




## Discussion

### Findings

Our main goal was to propose and construct a security encryption/decryption based solution in a cooperative environment for mHealth apps. We aimed to ensure data confidentiality, integrity, and authenticity. Privacy is a top priority issue in mHealth services and apps that deal with user sensitive information. On mHealth apps, several security issues must be considered, such as personal information management, secondary use of personal information, improper use of personal information, and errors with stored personal information. Therefore, cryptographic mechanisms can be seen as a solution to guarantied data confidentiality and protection [34].

In a mobility and cooperative environment with constant health data being forwarded and retrieved, studying and developing security mechanisms become crucial. Several experiments were conducted, involving 35 different users, to check if they could distinguish the app running with and without the DE4MHA embedded. Through the trials, we concluded that users could not tell which app had the DE4MHA embedded mainly because the time response taken to obtain the user health information was nearly the same as without DE4MHA. The DE4MHA was implemented in a ubiquitous manner so users were able to keep using the app without noticing changes or presence of any cryptography mechanisms.

### Limitations

There were several limitations to the study. The main limitation was applying security on mobile devices due to the low processor capacity compared with personal computers (PCs), though tremendous improvements in this area have been made, with a few mobile devices capable of competing with traditional PCs. This improvement allowed us to test several security algorithms to address the issues of confidentiality (AES, RC4, 3DES, and Blowfish), integrity (MD5 and SHA1), and authenticity (RSA with MD5 and DSA with SHA1) in order to verify which combination had a better performance in a mobile environment.

During the experiments, some users without Internet connectivity who wanted to obtain health information were in constantly moving further away from other users. Although the cooperation mechanism was embedded, users that were beyond the range of 10 meters (the maximum range for Bluetooth class 2 modules) could not obtain the desired health information. Another limitation, though not related to security, was with regard to the number of uncooperative nodes (mobile nodes that may not want to cooperate or they may not have cooperation mechanisms embedded), compromising service response probability.

### Conclusion

This paper proposed a data encryption solution for mobile health apps, called DE4MHA. The data encryption algorithm DE4MHA with cooperation mechanisms in mobile health allow users to safely obtain health information with the data being carried securely. These security mechanisms did not deteriorate the overall network performance and the app, maintaining similar performance levels as without the encryption. More importantly, it offers a robust and reliable increase of privacy, confidentiality, integrity, and authenticity of their health information. Although it was experimented on a specific mHealth app, SapoFit, both DE4MHA and the cooperation strategy can be deployed in other mHealth apps.

## Acknowledgments

This study was partially supported by Instituto de Telecomunicacões, Next Generation Networks and Applications Group (NetGNA) Portugal, National Funding from the FCT—Fundacão para a Ciência e a Tecnologia through the PEst-OE/EEI/LA0008/2011 Project, National Natural Science Foundation of China (Grant No. 61201165 and No. 61271240), and the AAL4ALL (Ambient Assisted Living for All), project co-funded by COMPETE under FEDER via QREN Programme.

## Conflicts of Interest

None declared.

## References

1. Le Moullee B, Ray P. Issues in E-Health Cost Impact Assessment. In: IFMBE Proceedings. 2009 Presented at: World Congress on Medical Physics and Biomedical Engineering; September 9-12, 2009; Munich, Germany p. 223-226. [doi: 10.1007/978-3-642-03893-8_63]
2. Akter S, D'Ambra J, Ray P. User Perceived Service Quality of mHealth Services in Developing Countries. 2010 Jun 09 Presented at: European Conference on Information Systems; June 6-9, 2010; Pretoria, South Africa.
3. United Nations Foundation, Vodafone Group Foundation, Telemedicine Society of India. mHealth and Mobile Telemedicine - an Overview. 2008 Aug 01. URL: http://ehealth-connection.org/content/mhealth-and-mobile-telemedicine-an-overview [accessed 2013-03-06] [WebCite Cache ID 6EvqsNn7K]
4. Subramoniam S, Sadi S. Healthcare 2.0. IT Professional 2010;12(6):46-51. [doi: 10.1109/MITP.2010.66]
5. epSOS – the European eHealth Project. URL: http://www.epsos.eu/ [accessed 2013-02-28] [WebCite Cache ID 6E0ipZcb2]
6. Akter S, Ray P. mHealth - an Ultimate Platform to Serve the Unserved. Yearb Med Inform 2010:94-100. [Medline: 20938579]
7. Tachakra S, Wang XH, Istepanian RS, Song YH. Mobile e-health: the unwired evolution of telemedicine. Telemed J E Health 2003;9(3):247-257. [doi: 10.1089/153056203322502632] [Medline: 14611692]
8. Laxminarayan S, Istepanian RS. UNWIRED E-MED: the next generation of wireless and internet telemedicine systems. IEEE Trans Inf Technol Biomed 2000 Sep;4(3):189-194. [Medline: 11026588]
9. Istepanian RSH, Lacal JC. Emerging mobile communication technologies for health: some imperative notes on m-health. : IEEE; 2003 Presented at: 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; September 17-21; Cancun, Mexico p. 1414-1416. [doi: 10.1109/IEMBS.2003.1279581]
10. Istepanian RSH, Laxminarayan S, Pattichis CS. M-Health: Emerging Mobile Health Systems. In: M-health: emerging mobile health systems. New York, NY: Springer; 2006.
11. Paré G, Moqadem K, Pineau G, St-Hilaire C. Clinical effects of home telemonitoring in the context of diabetes, asthma, heart failure and hypertension: a systematic review. J Med Internet Res 2010 Jun;12(2):e21 [FREE Full text] [doi: 10.2196/jmir.1357] [Medline: 20554500]
12. Fayn J, Rubel P. Toward a personal health society in cardiology. IEEE Trans Inf Technol Biomed 2010 Mar;14(2):401-409. [doi: 10.1109/TITB.2009.2037616] [Medline: 20007033]
13. Lin CT, Chang KC, Lin CL, Chiang CC, Lu SW, Chang SS, et al. An intelligent telecardiology system using a wearable and wireless ECG to detect atrial fibrillation. IEEE Trans Inf Technol Biomed 2010 May;14(3):726-733. [doi: 10.1109/TITB.2010.2047401] [Medline: 20371411]
14. Kollmann A, Riedl M, Kastner P, Schreier G, Ludvik B. Feasibility of a mobile phone-based data service for functional insulin treatment of type 1 diabetes mellitus patients. J Med Internet Res 2007 Dec;9(5):e36 [FREE Full text] [doi: 10.2196/jmir.9.5.e36] [Medline: 18166525]
15. Mougiakakou SG, Bartsocas CS, Bozas E, Chaniotakis N, Iliopoulou D, Kouris I, et al. SMARTDIAB: a communication and information technology approach for the intelligent monitoring, management and follow-up of type 1 diabetes patients. IEEE Trans Inf Technol Biomed 2010 May;14(3):622-633. [doi: 10.1109/TITB.2009.2039711] [Medline: 20123578]
16. Jara AJ, Zamora MA, Skarmeta AFG. An internet of things–based personal device for diabetes therapy management in ambient assisted living (AAL). Pers Ubiquit Comput 2011 Jan 2011;15(4):431-440. [doi: 10.1007/s00779-010-0353-1]
17. Zhu F, Bosch M, Woo I, Kim S, Boushey CJ, Ebert DS, et al. The Use of Mobile Devices in Aiding Dietary Assessment and Evaluation. IEEE J Sel Top Signal Process 2010 Aug;4(4):756-766 [FREE Full text] [doi: 10.1109/JSTSP.2010.2051471] [Medline: 20862266]
18. Pollak J, Gay G, Byrne S, Wagner E, Retelny D, Humphreys L. It's Time to Eat! Using Mobile Games to Promote Healthy Eating. IEEE Pervasive Comput 2010 Jul 2010;9(3):21-27. [doi: 10.1109/MPRV.2010.41]
19. Patrick K, Raab F, Adams MA, Dillon L, Zabinski M, Rock CL, et al. A text message-based intervention for weight loss: randomized controlled trial. J Med Internet Res 2009;11(1):e1 [FREE Full text] [doi: 10.2196/jmir.1100] [Medline: 19141433]

20. Watson A, Bickmore T, Cange A, Kulshreshtha A, Kvedar J. An internet-based virtual coach to promote physical activity adherence in overweight adults: randomized controlled trial. J Med Internet Res 2012;14(1):e1 [FREE Full text] [doi: 10.2196/jmir.1629] [Medline: 22281837]
21. Whittaker R, Dorey E, Bramley D, Bullen C, Denny S, Elley CR, et al. A theory-based video messaging mobile phone intervention for smoking cessation: randomized controlled trial. J Med Internet Res 2011;13(1):e10 [FREE Full text] [doi: 10.2196/jmir.1553] [Medline: 21371991]
22. Déglise C, Suggs LS, Odermatt P. Short message service (SMS) applications for disease prevention in developing countries. J Med Internet Res 2012;14(1):e3 [FREE Full text] [doi: 10.2196/jmir.1823] [Medline: 22262730]
23. Silva BMC, Rodrigues JJPC, Lopes IMC, Machado TMF, Zhou L. A Novel Cooperation Strategy for Mobile Health Applications. IEEE Journal on Selected Areas in Communications Special Issue on Emerging Technologies in Communications - eHealth, IEEE Communications Society 2013:1-9 (forthcoming).
24. Silva BMC, Lopes IMC, Rodrigues JJPC, Ray P. SapoFitness: A mobile health application for dietary evaluation. : IEEE; 2011 Presented at: 13th IEEE International Conference on e-Health Networking Applications and Services (IEEE HEALTHCOM 2011); June 13-15, 2011; Columbia, MO, USA p. 13-20. [doi: 10.1109/HEALTH.2011.6026782]
25. Rodrigues JJ, Lopes IM, Silva BM, de la Torre I. A new mobile ubiquitous computing application to control obesity: SapoFit. Inform Health Soc Care 2013 Jan;38(1):37-53. [doi: 10.3109/17538157.2012.674586] [Medline: 22657250]
26. SapoFit. URL: https://itunes.apple.com/pt/app/sapo-fit/id438487775?mt=8 [accessed 2013-02-15] [WebCite Cache ID 6Cw2BHsOA]
27. Jonsson J, Kaliski B. Public-Key Cryptography Standards (PKCS) #1: RSA Cryptography Specifications Version 2.1. URL: http://www.rfc-editor.org/ [accessed 2013-02-15] [WebCite Cache ID 6Cw66XuiI]
28. Raeburn K. Advanced Encryption Standard (AES) Encryption for Kerberos 5. URL: http://www.rfc-editor.org/ [accessed 2013-02-15] [WebCite Cache ID 6Cw4IKbXi]
29. Housley R. Triple-DES and RC2 Key Wrapping. URL: http://www.rfc-editor.org/ [accessed 2013-02-15] [WebCite Cache ID 6Cw4ZXBdH]
30. Jaganathan K, Zhu L, Brezak J. The RC4-HMAC Kerberos Encryption Types. URL: http://www.rfc-editor.org/ [accessed 2013-02-15] [WebCite Cache ID 6Cw5Lwx1B]
31. Shirey R. Internet Security Glossary, Version 2. URL: http://www.rfc-editor.org/ [accessed 2013-02-15] [WebCite Cache ID 6Cw5pWnxS]
32. Rescorla E. Diffie-Hellman Key Agreement Method. URL: http://www.rfc-editor.org/ [accessed 2013-02-15] [WebCite Cache ID 6Cw6I1jk5]
33. McGrew D, Igoe M, Salter M. Fundamental Elliptic Curve Cryptography Algorithms. URL: http://www.rfc-editor.org/rfc/rfc6090.txt [accessed 2013-02-28] [WebCite Cache ID 6E0ktAxqb]
34. Raychaudhuri K, Ray P. Privacy Challenges in the Use of eHealth Systems for Public Health Management. International Journal of e-Health and Medical Communications, IGI-Global 2010;1(2):12-23. [doi: 10.4018/jehmc.2010040102]


## Abbreviations

**3DES:** Triple Data Encryption Standard
**ACT:** achieved cooperation time
**AES:** Advanced Encryption Standard
**BMI:** body mass index
**BMR:** basal metabolic rate
**CWS:** cooperative Web service
**DE4MHA:** data encryption solution for mobile health apps
**EHR:** electronic health record
**HTTPS:** Hypertext Transfer Protocol Secure
**ICT:** information and communication technologies
**JSP:** Java Server Pages
**RC4:** Rivest Cipher 4
**REST:** Representational State Transfer
**RSA:** Rivest, Shamir, Adleman
**SOAs:** service oriented architectures
**SSL:** Secure Socket Layer
**WS:** Web service

*Edited by G Eysenbach, A Jara, S Koch, P Ray; submitted 15.12.12; peer-reviewed by S Zeadally, C Verikoukis, A Vinel, P Lorenz, A Jara; comments to author 10.01.13; revised version received 22.01.13; accepted 09.02.13; published 25.04.13.*

*Please cite as:*
*Silva BM, Rodrigues JJPC, Canelo F, Lopes IC, Zhou L*
*A Data Encryption Solution for Mobile Health Apps in Cooperation Environments*
*J Med Internet Res 2013;15(4):e66*
*URL: http://www.jmir.org/2013/4/e66/*
*doi:10.2196/jmir.2498*
*PMID:23624056*

Original Paper

# Examining the Differences in Format and Characteristics of Zoonotic Virus Surveillance Data on State Agency Websites

Matthew Scotch[1,2], PhD, MPH; Brittany Baarson[1,2]; Rachel Beard[1,2]; Robert Lauder[1,2]; Aarthi Varman[1,2]; Rolf U Halden[1], PhD

[1]Center for Environmental Security, Biodesign Institute and Security and Defense Systems Initiative, Arizona State University, Tempe, AZ, United States

[2]Department of Biomedical Informatics, Arizona State University, Scottsdale, AZ, United States

**Corresponding Author:**
Matthew Scotch, PhD, MPH
Center for Environmental Security
Biodesign Institute and Security and Defense Systems Initiative
Arizona State University
P.O. Box 875904
Tempe, AZ, 85287-5904
United States
Phone: 1 480 727 2985
Fax: 1 480 965 6603
Email: matthew.scotch@asu.edu

## Abstract

**Background:** Zoonotic viruses are infectious organisms transmittable between animals and humans. Agencies of public health, agriculture, and wildlife conduct surveillance of zoonotic viruses and often report data on their websites. However, the format and characteristics of these data are not known.

**Objective:** To describe and compare the format and characteristics of statistics of zoonotic viruses on state public health, agriculture, and wildlife agency websites.

**Methods:** For each state, we considered the websites of that state's public health, agriculture, and wildlife agency. For each website, we noted the presence of any statistics for zoonotic viruses from 2000-2012. We analyzed the data using numerous categories including type of statistic, temporal and geographic level of detail, and format. We prioritized our analysis within each category based on assumptions of individuals' preferences for extracting and analyzing data from websites. Thus, if two types of data (such as city and state-level) were present for a given virus in a given year, we counted the one with higher priority (city). External links from agency sites to other websites were not considered.

**Results:** From 2000-2012, state health departments had the most extensive virus data, followed by agriculture, and then wildlife. We focused on the seven viruses that were common across the three agencies. These included rabies, West Nile virus, eastern equine encephalitis, St. Louis encephalitis, western equine encephalitis, influenza, and dengue fever. Simple numerical totals were most often used to report the data (89% for public health, 81% for agriculture, and 82% for wildlife), and proportions were not different (chi-square $P$=.15). Public health data were most often presented yearly (66%), while agriculture and wildlife agencies often described cases as they occurred (Fisher's Exact test $P$<.001). Regarding format, public health agencies had more downloadable PDF files (68%), while agriculture (61%) and wildlife agencies (46%) presented data directly in the text of the HTML webpage (Fisher's Exact test $P$<.001). Demographics and other information including age, gender, and host were limited. Finally, a Fisher's Exact test showed no association between geography data and agency type ($P$=.08). However, it was noted that agriculture department data was often at the county level (63%), while public health was mixed between county (38%) and state (35%).

**Conclusions:** This study focused on the format and characteristics of statistics of zoonotic viruses on websites of state public health, wildlife, and agriculture agencies in the context of population health surveillance. Data on zoonotic viruses varied across agencies presenting challenges for researchers needing to integrate animal and human data from different websites.

***(J Med Internet Res 2013;15(4):e90)*** doi:10.2196/jmir.2487

**KEYWORDS**

public health; zoonoses; World Wide Web; epidemiology; data analysis

## Background

Data that are freely available on health agency websites can be used for surveillance, including monitoring infection rates over time and identifying outbreaks. This includes data on zoonotic viruses that are transmittable between animals and humans [1]. These viruses represent a significant population health concern [2], and their control is vital for reducing human and animal morbidity and mortality. In the United States, surveillance is done at the local, state, and federal levels [3]. Every state maintains departments of public health, agriculture, and wildlife. Often, these agencies rely on clinical and laboratory reporting of disease cases from predetermined lists [4] including those involving animals [5]. Depending on the state, reporting of animal cases can be separate from public health reporting [2]. A survey by Kahn [2] examined the reporting by practicing veterinarians to their state-appointed veterinarian who often work for agriculture agencies. Kahn found that 18% of state veterinarians responded that individual case reports of zoonotic viruses in animals are sent to public health agencies rather than to agriculture agencies [2]. As an explanation, Kahn notes the history of agriculture agencies was often to focus on agriculture and not necessarily surveillance of animal disease [2]. In a separate study by Kahn, a survey of over 1000 practicing veterinarians found that 30% notified their state agriculture agency in the event of an unusual infectious disease, while 23% notified their state public health agency [2]. On the public health side, a survey by M'ikanatha et al [6] asked state and local infectious disease epidemiologists what data sources they accessed for surveillance. Only 54% considered state agency websites their main resource of online investigation of public health data [6]. In another study, Staes et al [7] found that only 35% of clinicians during the 2009 influenza A H1N1 epidemic accessed state agency websites at least once a week while 50% never accessed them at all. In contrast, over half of the participants visited the CDC's influenza website (53%) at least once a week [7]. These studies suggest inconsistent data reporting structures and utilization patterns across both human and animal health agencies.

One of the few Internet studies on zoonoses was a 2008 study by Pappas et al [8], which examined online resources for scientific information on zoonoses. The authors performed a Google and Yahoo! search for any content related to zoonoses or zoonotic infections [8]. The websites that were found included those sponsored by agencies at the international, country, state, and local levels [8]. Despite the global threat of zoonotic infections and the burdens that they have on developing countries, the authors found that the majority of the sites were from agencies or academic institutions within the United States [8]. Many of these sites were from state health agencies—a finding that is relevant to the scope of this paper and demonstrates their potential as a resource for zoonotic surveillance.

While there have been a limited number of studies focused on zoonoses, agency website data have been shown to be useful in other areas of health-related research and surveillance. A 2011 paper by Aswani et al examined state differences in reporting of central line-associated bloodstream infections (CLABSIs) [9]. The authors utilized data from agency websites for hospital acquired infection and CLABSI reports [9]. Searches from state websites included terms such as "health care quality, data, and/or statistical reports" ([9] pg. 388) or more specific terms related to CLABSIs or hospital acquired infection. Overall, the authors found variation in the data across the 15 states that had publicly available data; 86% of websites published data using infection rates while others did not provide a rate [9]. In addition, certain states adjusted their data by standardized infection ratios while others used device utilization ratios [9]. There were also differences in how the data were aggregated as some were done without consideration of the unit within the hospital [9]. The wide variation of data on agency websites makes state comparisons difficult and highlights the need for better standards in reporting.

At the federal level, the ArboNET system [10] is a collaborative effort to compile data on certain infectious diseases, most notably West Nile virus (WNV). It has been used in many surveillance research studies (such as [11-13]). It also includes data on St. Louis encephalitis, eastern equine encephalitis, La Crosse encephalitis, Powassan virus, and dengue fever. A study by O'Leary et al [12] used ArboNET to examine the severity of WNV post 1999. Cases between January 1, 2002, and March 15, 2003, were included in the study. The authors analyzed information such as date of illness and county of residence [12]. In total, over 4000 cases were included with 54% classified as *confirmed* and 46% as *probable* [12]; 84% of cases were from 11 states with the most cases from Michigan, Illinois, and Ohio [12]. Mississippi had the highest incidence when factoring in population size [12].

Federal surveillance initiatives such as ArboNET and the CDC's influenza program [14] exist for some viruses; however, there are often many limitations with the data. Agencies follow their own procedures in terms of reporting [15] requiring careful consideration before aggregation with other states. There is often a loss of granularity as data are reported from states to the federal level. For example, CDC's influenza surveillance program aggregates statistics by Health and Human Services (HHS)-defined geographic regions. This eliminates the ability to compare influenza statistics by state. In addition, ArboNet provides tables showing cumulative data by county per year, rather than providing monthly summaries. This eliminates the ability to compare statistics by month. They do provide more detailed data in the form of graphs; however, exact numbers can be difficult to interpret. Finally, important data such as gender and age of individuals might get removed as it is reported to the federal level prohibiting it from being used as a dimension for comparison. This eliminates the ability to compare differences by sex, race, and age groups. As data get aggregated, it becomes more and more difficult to uncover temporal, geographic, and demographic relationships of viral infection.

As an alternative, state agency data has the potential to be a valuable resource for epidemiologists and clinicians looking to analyze zoonotic viruses data among animals and humans. Current popular resources like ArboNET and HealthMap [16-18] have generated immense interest in the surveillance community and shown that integration of animal and human data can be of great value for monitoring of zoonotic viruses. However, less is known about the format and characteristics of state agency website data, despite the large amount of data that are collected. The purpose of this study is to examine websites of state agencies of public health, agriculture, and wildlife to characterize a subset of publicly available data for surveillance of zoonotic viruses. We explored issues beyond the mere presence or absence of data for a given virus and considered the quality, format, and completeness of the data for downloading and subsequent utilization for research purposes, including integration of human and animal data.

## Methods

We generated a list of 63 zoonotic viruses (Multimedia Appendix 1) from a review of Krauss [1]. We decided a priori, not to limit our focus to viruses that were endemic/enzootic to the United States, enabling rare events to be considered in our study. Owing to resource limitations and a desire to focus on the most recent data, the study was limited to the years 2000-2012 and to zoonotic viruses, excluding bacteria, fungi, and parasites.

For each of the 50 states, our search criteria consisted of the following steps:

1. Search on Google for *[state name]+ public health department* or *[state name] + wildlife department* or *[state name] + department of agriculture*. Two of the authors (RL and AV) examined public health websites, and two examined agriculture and wildlife websites relating to animal health (BB and RB).
2. Identify the appropriate website link from the Google search results.
3. Examine homepage for links related to surveillance data such as *epidemiological data*, *disease statistics*, *zoonotic diseases*, *animal health*, etc.
4. Use the website's search function (if available) to find any data that might have been missed through navigation of the links.
5. Collect the following data related to a virus on the a priori list: the type of the statistic (including totals and averages), the years data are available (from 2000-2012), the infected host, and whether the data are in an HTML table, a free-text paragraph, or downloadable as a PDF file, Word document, or spreadsheet. We also considered the frequency of the data, the level of geographic detail, and the presence of any demographic-related data.
6. Summarize results for each category in #5 (except for *demographics*), prioritizing based on our estimation of the following: If researchers need to extract data from a website for analysis purposes, what might be their preferences of data format, level of geography, how it was presented, etc. For *geography*, the highest level of granularity is likely preferred for research compared to lower levels such as state totals. Thus, we chose *city* as the best scenario for the *geography* category. For this analysis, if virus data were provided at different levels for the same year, such as city and state, we considered it *city*. A *mixed* result was only indicated if for a given virus and year, it had one type of data (eg, as city-level data) and then for the same virus for a different year, had another type of data (eg, county-level data). This procedure was done in order to avoid binning everything into *mixed*.
7. Exclude from our results any links to external sites; thus, a website with a link to an outside source, such as another state agency or a federal agency, was not factored into our work.

Descriptive statistics and charts were done in Microsoft Excel. Hypothesis testing for comparison of proportions within public health, agriculture, and wildlife was done using SAS v.9.2. For this, we considered either the chi-square or Fisher's Exact test.

## Results

All 50 states had websites for public health, wildlife, and agriculture agencies, although naming conventions did vary. From 2000-2012, state health departments, followed by agriculture, and then wildlife, most often had virus data. In total, seven viruses were common among all three groups. These included rabies, West Nile virus (WNV), eastern equine encephalitis (EEE), St. Louis encephalitis (SLE), western equine encephalitis (WEE), influenza, and dengue fever. Figure 1 shows the number of states with data for these viruses from 2000-2012. Rabies and WNV data were the most abundant. Rabies data were different from the other zoonotic viruses. Here, many public health agencies provided data on cases of infection in animals instead of agriculture or wildlife agencies. Since human cases of rabies in the United States are extremely rare, due to postexposure prophylaxis, there is often nothing to report. Thus, the 45 state public health agencies providing rabies data during this time span detailed animal cases and the rare exception of human cases. Concurrently, some wildlife or agriculture agencies did provide some rabies data, and thus there is a possibility of an overlap of rabies data. However, this occurred infrequently.

Figure 2 shows a temporal analysis of public health website data for all 50 US states, considering the seven common viruses (Figure 1). As of this writing, the year 2012 is not complete and since some websites might wait until the year is completed to provide statistics, this value is likely an underestimate.

Tables 1-3 compare by agency, the number of states that provided zoonotic viruses data for 2000-2012. If data were provided on a website for any year during 2000-2012, the value "1" was added to the column. The number of years was not a factor in this comparison, and thus 10 years of rabies data or 5 years still resulted in adding "1" to the column.

Table 1 considers the type of the statistic such as a total, rate, or percentage. We set *totals* as the priority for this category, believing that this statistic is the easiest to analyze across groups versus a rate or ratio where the at-risk population, especially

for animals, is not usually known. Here, the majority of the website data of all three agency types were provided as totals. Thus rates, percentages, other types of statistics, and agencies that provided a mix of formats were grouped into the *other* group. We also considered the level of geographic granularity, as this impacts conclusions that can be drawn from a population-level analysis. As mentioned, *city* was set as the priority with the remaining order based on granularity (ie, county then state, etc). A Fisher's Exact test showed no association between geography data and agency type ($P$=.08). However, public health agencies, which mostly provide data on human cases, had a proportion of 38% at the county level versus agriculture agencies that provided 63% of their data at the county level. This could be related to the concern over issues of confidentiality among small geographic sample sizes. The *mixed* variable represents an agency that provided virus data at different levels of granularity for different years, such as one year at the county level and one year at the state level. Finally, the analysis of demographics suggested that when data were provided, they usually did not include age, gender, or race of humans. Lack of data on demographics limits comparisons across datasets and the ability to identify the most at-risk groups. For example, if agriculture and wildlife agencies do not provide the infected host (eg, bats, skunks) when providing statistics on virus cases, comparisons across different host populations are limited.

We also considered the format of the data (Table 2). Here, we set *spreadsheet* such as XLS or CSV files as the priority since these files can easily be downloaded and formatted and are already in an application that supports analysis and comparison. HTML was second since content on a webpage can often be copied and pasted as needed. The format of the data impacts the feasibility of data integration. For public health agencies, 68% of the data was provided through links to PDF documents, which often requires manual data entry into a researcher's database or spreadsheet. A higher percentage of agriculture website data was available in HTML format than public health (61% versus 19%). Finally, we also analyzed the manner in which the data were presented. Tabular data (the priority) is potentially easier to load into a database or spreadsheet, while graphs and charts are visually informative yet sometimes the actual numbers are not provided, thereby forcing the researcher to infer the true numbers. Public health departments had a much greater percentage of data in tables (57%), while agriculture agencies mostly provided their virus statistics within text (63%). This presents challenges for researchers relying on automated updates (data dumps) or "Web crawlers" to perform Natural Language Processing (NLP) on the free-text in order to extract the statistics.

Table 3 considers the frequency in which data are provided as well as the number of years that the data are available. Public health departments most often provided data as annual aggregates (66%) as opposed to weekly or monthly trends. Conversely, agriculture or wildlife agencies more often provided data *as identified* (58%) and typically in a paragraph of free-text such as describing a recent case of WNV in a red-tailed hawk.

We also considered the time-span of reporting since gaps over time limit the ability for researchers to identify true temporal fluctuations in the data. An agency might provide data for cases from 2000-2005, but for a variety of reasons (loss of Webmaster or epidemiologist, perceived lack of interest in the community, etc) not include the data after that period. Here, the average number of years that data for a given virus were available on their website was greater among public health departments than departments of agriculture and wildlife. This is likely associated with the higher proportion of *as identified* data among the animal agencies (agriculture and wildlife).

For public health agencies, it was interesting that highly prevalent viruses such as influenza had the smallest number of years available (7.1). This might be related to the fact that the CDC and their surveillance program [14] offer historical data on a weekly basis, and thus many agency websites contain a link to the CDC for this information. Since we did not consider external links in this study, this was not counted in our totals.

We also mapped total virus data by the different agencies during 2000-2012 (Figures 3 and 4) and considered all 63 viruses from our list (Multimedia Appendix 1). The presence of the virus data even once during the time-span was counted towards the total (hence the number of years was not a factor). Thus, the theoretical maximum for each state was 63 since that was the size of our a priori list. Since many of the viruses on our list are not found in the United States, the numbers were much lower. In addition, we did not consider any links to external sites in our results, thus a state department of agriculture that links all of their data to the US Department of Agriculture (USDA) would seem to have less data in our results than a site that did not link out. Geographic influence appeared to occur in some areas of the country, as states in the middle had similar numbers, yet areas in the northeast such as Ohio and Pennsylvania or southwest such as Arizona and New Mexico were different from one another. The map of the two animal agencies (Figure 4) is greatly impacted by the number of links to external data sources. This is evident with a state like Texas, which did not provide any statistics on their sites, instead providing links to other sources such as the state public health agency or the USDA.

**Table 1.** Comparison by agency type of the number of states that provided zoonotic viruses data on websites by type of statistic used, level of geography, and inclusion of demographics, 2000-2012 (data combined for each of virus in Figure 1; *P* values determined by chi-square for Statistic and Fisher's Exact test for Geography; percentages may not add to 100% due to rounding).

| | | Agencies | | | |
|---|---|---|---|---|---|
| | | Public Health (N=187) | Agriculture (N=58) | Wildlife (N=11) | *P* value |
| **Statistic** [a] | | | | | .15 |
| | Totals | 167 (89%) | 47 (81%) | 9 (82%) | |
| | Other[b] | 20 (11%) | 11 (19%) | 2 (18%) | |
| **Geography** | | | | | .08 |
| | City | 13 (7%) | 3 (5%) | 0 | |
| | County | 72 (38%) | 37 (63%) | 4 (31%) | |
| | State | 66 (35%) | 13 (22%) | 7 (54%) | |
| | Mixed | 31 (16%) | 5 (8%) | 2 (15%) | |
| | Other[c] | 5 (3%) | 1 (2%) | 0 | |
| **Demographics** [d] | | | | | |
| | Age | 51 | 5 | 0 | |
| | Gender | 32 | 2 | 0 | |
| | Race | 13 | 0 | 0 | |
| | Animal host | 80 | 54 | 11 | |
| | None | 84 | 3 | 2 | |

[a]Additional *none* variable not shown. There were three instances where a website indicated that "no cases" were reported.

[b]Includes *rates*, *percentages*, *mixed*, or *other*.

[c]Includes *zipcodes*, *region*, *town*, *district*, *jurisdiction.* There was one instance of zipcode-level data.

[d]Percentages not used since categories are not mutually exclusive. Base "N" for each agency type will not add up since categories are not mutually exclusive. Removed *other,* which was zero for all agencies.

**Table 2.** Comparison by agency type of the number of states that provided zoonotic viruses data on websites by format and presentation, 2000-2012 (*P* values determined by Fisher's Exact test; percentages might not add to 100% due to rounding).

| | | Agencies | | | |
|---|---|---|---|---|---|
| | | Public Health | Agriculture | Wildlife | *P* value |
| **Format** | | | | | .001[a] |
| | Spreadsheet | 3 (1%) | 0 | 0 | |
| | HTML | 36 (19%) | 36 (61%) | 6 (46%) | |
| | PDF | 128 (68%) | 20 (33%) | 5 (38%) | |
| | Mixed | 20 (11%) | 3 (5%) | 2 (15%) | |
| **Presentation** | | | | | |
| | Table | 108 (57%) | 5 (8%) | 2 (15%) | |
| | Graph | 16 (9%) | 0 | 1 (8%) | |
| | Map | 6 (3%) | 6 (10%) | 0 | |
| | Text | 13 (7%) | 37 (63%) | 5 (38%) | |
| | Mixed | 44 (23%) | 11 (18%) | 5 (38%) | |

[a]Computed by removing *other* variable with all zeros.

**Table 3.** Comparison by agency type: The number of states that provided zoonotic viruses data on websites by frequency as well as the average duration of years the data is provided by virus, 2000-2012 (viruses chosen from Figure 1; frequency *P* value determined by Fisher's Exact test and virus *P* value determined by ANOVA; percentages might not add to 100% due to rounding).

| | | Agencies | | | |
|---|---|---|---|---|---|
| | | Public Health | Agriculture | Wildlife | *P* value |
| **Frequency** | | | | | <.001[a] |
| | Weekly | 25 (13%) | 3 (5%) | 0 | |
| | Monthly | 8 (4%) | 2 (3%) | 0 | |
| | Yearly | 124 (66%) | 18 (30%) | 3 (23%) | |
| | Mixed | 24 (13%) | 2 (3%) | 1 (8%) | |
| | As identified | 6 (3%) | 34 (58%) | 9 (69%) | |
| **Virus (average # years)** [b] | | | | | <.001 |
| | Rabies | 9.3 | 3.4 | 4.2 | |
| | WNV | 10.0 | 3.5 | 7 | |
| | EEE | 8.5 | 2.6 | 1 | |
| | SLE | 8.2 | 5 | 0 | |
| | WEE | 6 | 9 | 0 | |
| | Influenza | 7.1 | 1.5 | 2 | |
| | Dengue | 7.7 | 1 | 0 | |

[a]Computed by removing *other* variable with all zeros.

[b]Only includes states that reported data for a given virus.

**Figure 1.** Common zoonotic viruses data provided by agencies of public health, agriculture, or wildlife on their websites.

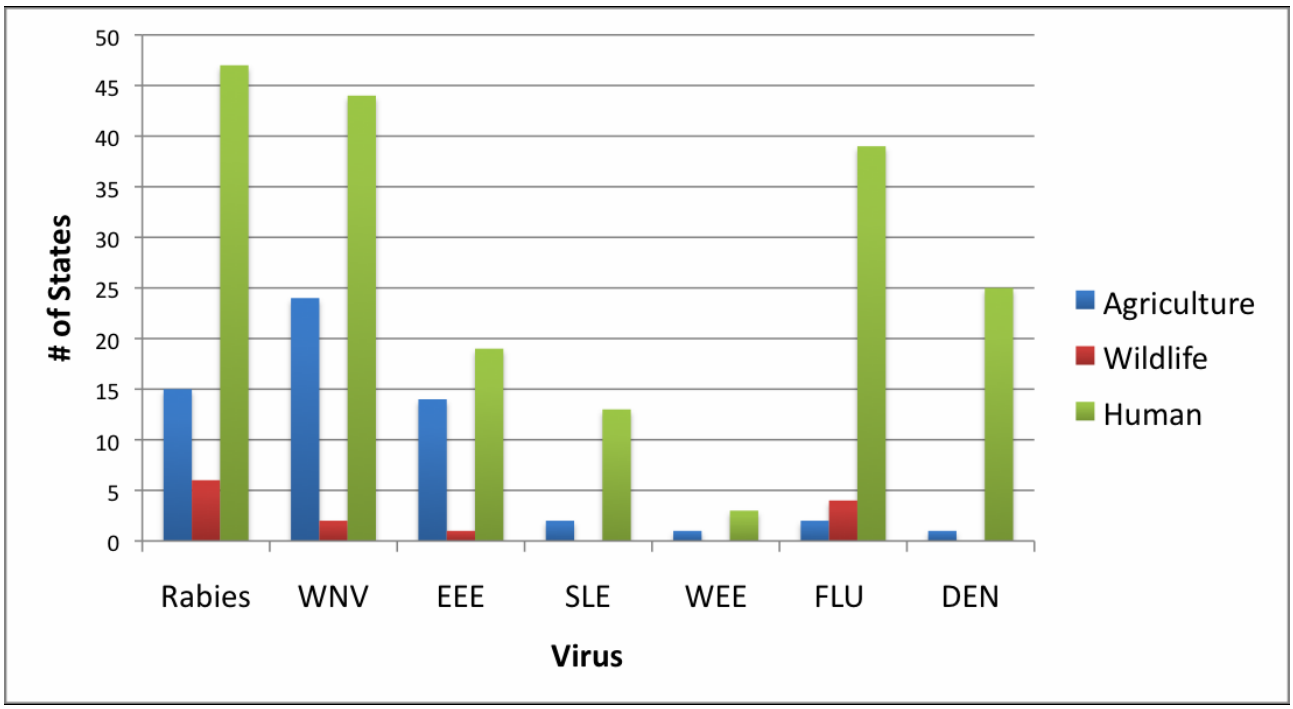


**Figure 2.** Cumulative number of state public health websites with surveillance data, by virus and year.

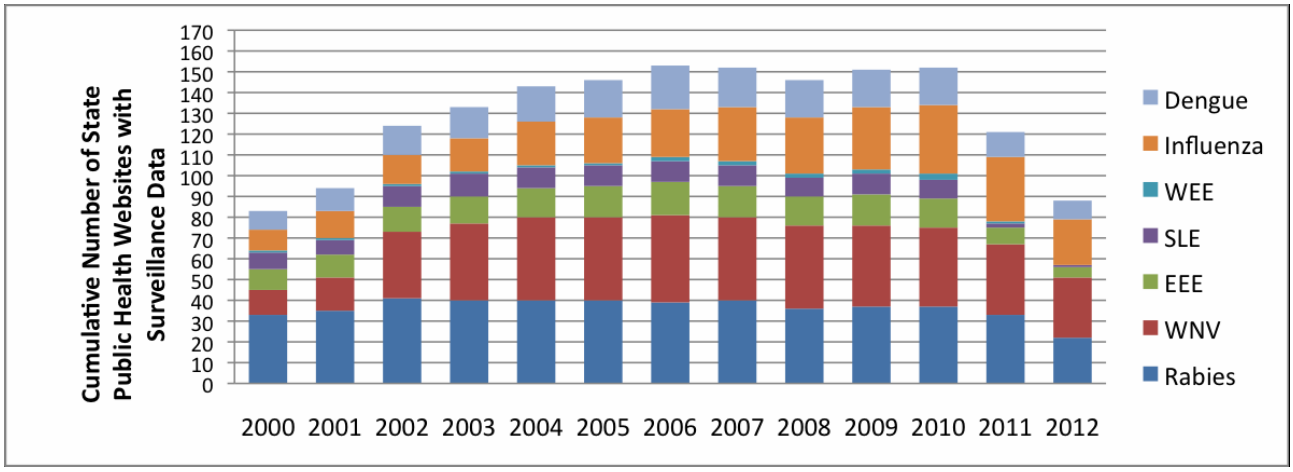



**Figure 3.** Map of the number of online virus data at public health departments (theoretical maximum: 63 viruses).

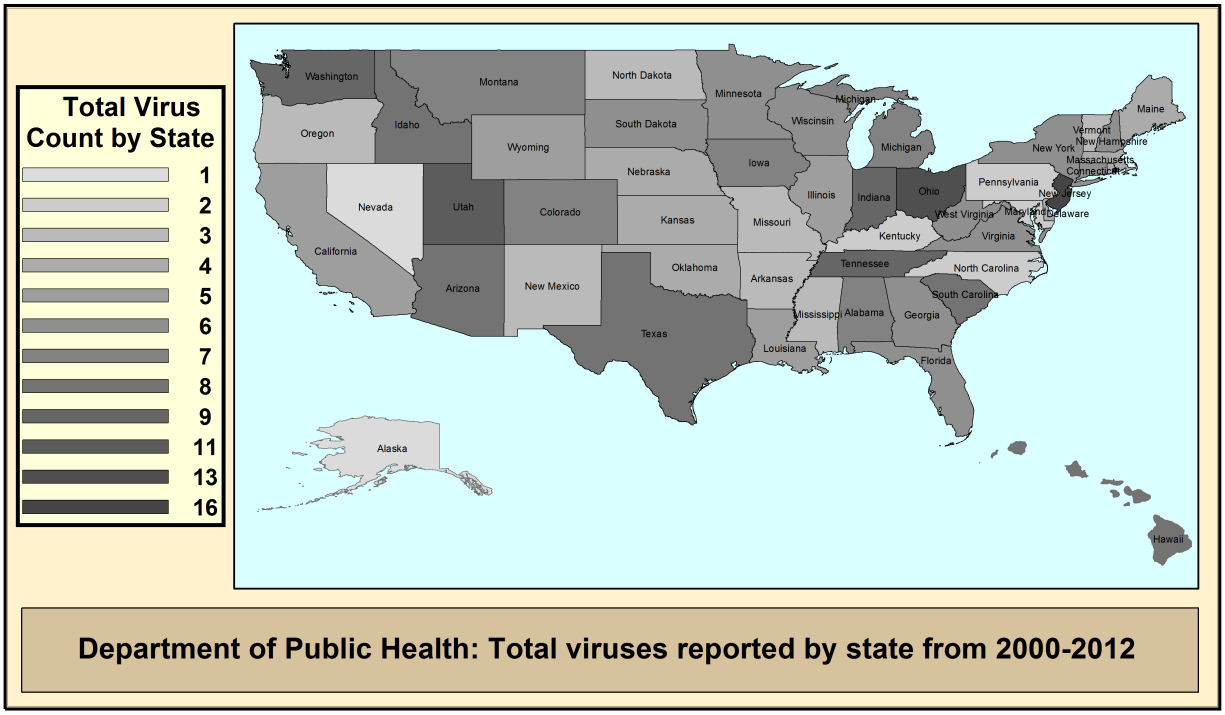
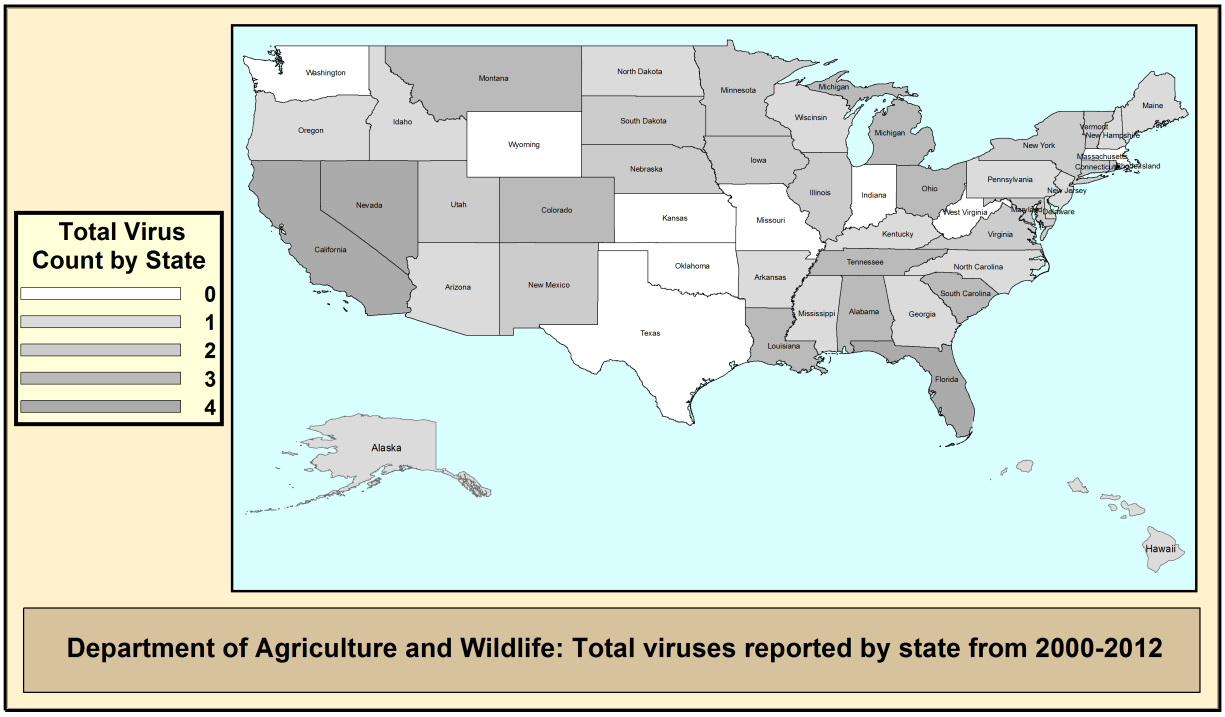


**Figure 4.** Map of the number of online virus data at agriculture and wildlife agencies.




## Discussion

Zoonotic viruses data varied across agencies of public health, agriculture, and wildlife, in relation to many of the different factors. Considering the amount of data available, wildlife agencies clearly had the least. This limits the ability to make informed decisions regarding the impact of zoonotic viruses among wildlife populations. The larger amount of data for public health and agriculture agencies are likely due to the fact that they receive provider-oriented data from clinicians and laboratories. Meanwhile, wildlife agencies rely on mortality events reported by the public, as the case with dead crow sightings for WNV [2] or initiatives like hunter-harvest programs.

Pennsylvania's WNV data [19] was one of the examples that we considered as having excellent data characteristics. It has specific dates and county-level data from 2001-2012. They also include demographic data for age and gender. While it does not have Excel format, the counts are in HTML tables with tables and maps (which can easily be copied into a spreadsheet). Finally, it had human, bird, mosquito, and veterinarian data enabling for easy comparison. Despite the larger amount of data for public health and agriculture agencies, differences exist that provide challenges for integration and analysis of zoonotic viruses. We now discuss the implications of our findings related to geography, format, presentation style, and frequency of the data.

### Geographic Mismatches Between Datasets

Results suggested that public health agencies provided fewer data at higher levels of geographic granularity than agricultural departments; however, no statistical difference was found ($P$=.08). As mentioned, this might be related to issues of sensitivity and confidentiality of data, especially in small geographic areas. When integrating disparate datasets for a given virus, a common practice is to use the highest matching level of granularity. Given that public health data are often provided at higher geographic levels, this will result in a loss of granularity for animal health data. One potential approach is to use data mining and classification algorithms, such as decision trees, to predict the data at a higher level of granularity; however, the accuracy of this technique for geographic inference is not known.

### Format and Presentation Styles

Of the public health data, 68% were provided in PDF file format. This impedes processing and analysis of the data, as manual database entry is often required. Web 2.0 and 3.0 technologies can partially address this problem including the use of software that can convert PDFs to text, thereby enabling easier modification. Even data in HTML format can be a challenge. Our findings showed that the majority of agriculture agencies had their data embedded in the webpage. For this, applications such as screen scrapers like Yahoo!'s Dapper [20] enable content to be extracted and automatically placed into an electronic format. Prior work by Yang et al [21] developed a framework for screen scraping climate data from websites, and work from Moumtzidou [22] considered environmental health data.

Agencies differed on their presentation style, with public health departments using more tables to present their data. Tables often lend themselves to easier extraction as they can be copied and pasted into a spreadsheet for processing. In addition, screen scrapers tend to work well with tables, although this can vary depending on the layout of the table. Conversely, agriculture and wildlife agencies presented their data within text of the

webpage. This likely corresponded to their tendency to provide data *as identified* as a written description. This presents challenges for researchers as they are forced to scan the text and manually identify relevant data. As previously mentioned, one solution is to use Natural Language Processing (NLP) to automatically extract relevant information from the text. Much of the work in the NLP community has focused on extraction of health information from the social media (such as Twitter) as opposed to health content sites [23]. One example by Doing-Harris et al [24] uses NLP for an online consumer health website. Another focus of NLP has been on mining and characterization of public health reports. Examples include research groups employing EpiSPIDER [25] and Stewart et al [26] who performed text mining classification work on ProMed reports [27,28]. While promise has been shown in these areas, more work needs to be done by the NLP community for applying methods and approaches to extract surveillance data from health agency websites.

### Frequency and Gaps in Data

Our results indicate that temporal gaps do exist regardless of agency. While our decision not to consider external links may have resulted in a potential underestimation (see Limitations), gaps in providing data do exist and need to be considered when analyzing disparate data. Our findings showed that agriculture and wildlife agencies had very low averages for total years of data. As indicated, this might be related to their tendency to provide data *as needed*. However, the researcher must determine if these gaps are due to absence of cases or missing information. Consistency in reporting data on websites is important for accurate assessment of public health needs. Gaps or missing data due to inconsistent reporting can bias parameter estimates [29] and lead to underestimation of virus infection in the population. A solution is to compare state agency website data with other sources of information, including federal health sites, news reports (ProMed, HealthMap), and online public databases, such as the Food and Agriculture Organization's (FAO) EMPRES-i [30], or electronic health record (EHR) and hospital utilization data. One resource for this is the Semantic Web, which was designed to promote interoperability of online resources and concepts, such as mash-ups facilitating integration of data across the Web [31]. Examples of public health applications that have used advanced and open-source Web technology include EpiVue [32] and SMCP-Aedes for dengue fever surveillance [33]. However, these systems do not address issues related to routine sharing and integration of animal and human health data.

### Recent Trends in Zoonotic Viruses Data on Agency Websites

Examining how characteristics in zoonotic viruses data changed over time across the three agencies will highlight trends for future surveillance efforts. For all of the categories measured in Tables 1-3, we graphed annual trends for each agency (Multimedia Appendix 2). For public health agencies, data in the year 2011 sharply decreased across all categories (in Multimedia Appendix 2, see Figures S1, S4, S7, S10, S13, S16). In Multimedia Appendix 2, Figure S1, the frequency of data was 145 in 2010, but only 123 in 2011. Similarly, geography of data was 150 in 2010 then declined to 122 in 2011 (Multimedia Appendix 2, Figure S4). Meanwhile, agriculture agencies had an increase across all categories for 2012, despite the year being incomplete. For example, the graph of frequency of data was 17 in 2011, while rising to 26 in 2012 (Multimedia Appendix 2, Figure S2). In addition, presentation of data was 17 in 2011, but 28 in 2012 (Multimedia Appendix 2, Figure S11).

There are many explanations for these findings. The sharp decrease in public health data for 2011 could be because more states might have increased their use of external links to data collected by the CDC rather than reporting it on their own websites. Since our study did not include these in our results, this would suggest a decrease in data. Another possibility is lack of time and resources dedicated to adding data to websites. Since many state public health budgets are small, there is a possibility that these departments have decided to allocate their time away from website reporting. Surprisingly, while public health agencies showed a decrease in 2011, agriculture agencies increased their data in 2012. There is a possibility that agriculture agencies have made more of an effort to post their data on their own websites rather than providing external links to CDC or USDA. These agencies might feel an obligation to keep the public informed as more people become aware of zoonotic viruses and the relationship between animal and human transmission.

Since one or two years does not provide enough evidence of a changing trend, additional work should focus on studying 2013 and beyond. Also, for zoonotic viruses, more research needs to focus on the potential of integrating animal and human data across state websites. Our study found seven common diseases in the United States. This provides for a great opportunity to utilize both animal and human sources for surveillance. In addition, for certain viruses, state-level detail might be used to augment more aggregated results at the federal level. Understanding how these different data sources can be utilized together might enable for more robust and elaborate surveillance systems.

Despite these challenges, agency website data offer great potential for virus surveillance by both clinicians and public health professionals. A survey by Gesteland found that 30% of clinicians in Utah access their states health department website [34]. However, the authors found that the main reason for not accessing the site was lack of awareness [34]. Thus informing clinicians about the potential benefits of public health data might increase utilization and provide valuable resources for clinical care.

### Limitations

The authors recognize several limitations with this work. First, we decided to prioritize variables (based on anticipated preference among researchers) within categories in order to reduce binning everything into a *mixed* category. Thus, lower priority variables such as *county* or *state* were omitted in the presence of higher priority variables such as *city*. Thus, our results underestimate the proportion of lower priority characteristics associated with surveillance data. However, our method was consistent across the three different agency types;

thus, we feel that our results still provide informative findings into the characteristics of zoonotic viruses data among the different types of agencies.

Second, we did not consider external links as a data source in this study, such as state health department websites that send visitors to other agencies (such as the CDC or USDA). Thus, a state like Texas that does not provide animal health data directly on their site, received lower scores (Figure 4). This likely resulted in an underestimation of the amount of data available on these types of sites. In other variables, such as *frequency*, we likely overestimated the amount of gaps over time (if an agency decided to link to another source). However, the purpose of our study was to focus completely on the content in the website itself, and not as a data portal.

We considered a website as providing data if it was done at least once during our study period (2000-2012). We did not explore data that was received by agencies but not indicated on their website during the study period. Finally, due to resources and time constraints, we limited our work to zoonotic viruses. This implies that more work needs to be done to understand the availability and utility of health data relating to animal-borne bacteria, fungi, and parasites, as well as non-zoonotic viruses.

## Conclusions

This study focused on the format and characteristics of zoonotic virus statistics on websites of state public health, wildlife, and agriculture agencies in the context of secondary sources of surveillance and research data. Zoonotic viruses data varied across agencies presenting challenges for researchers needing to integrate animal and human data from different agency websites. Advanced Web technologies can partially address this, but more effort is needed from the biomedical informatics community to work with public health, agriculture, and wildlife agencies to address online data access, quality, and consistency in order to promote and facilitate integration of animal and human data for surveillance of zoonotic viruses.

Federal initiatives such as ArboNET and the CDC influenza program have limitations for granular-level data analysis including regional preparedness efforts. Geographic, temporal, and demographic information might be available through a state health department but become lost as it is aggregated for federal reporting. This makes it more difficult to uncover hidden drivers of viral infection in the population. Data on state websites has often been overshadowed by more popular federal initiatives but offer the potential to be a valuable and rich resource for zoonotic disease surveillance.

## Acknowledgments

The project described was supported by award number R00LM009825 from the National Library of Medicine to Matthew Scotch. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Library of Medicine or the National Institutes of Health.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

List of 63 zoonotic viruses considered in this study.

[PDF File (Adobe PDF File), 188KB - jmir_v15i4e90_app1.pdf ]

## Multimedia Appendix 2

Temporal analysis by agency type.

[PDF File (Adobe PDF File), 697KB - jmir_v15i4e90_app2.pdf ]

## References

1. Krauss H. Zoonoses: infectious diseases transmissible from animals to humans, 3rd ed. Washington, DC: ASM Press; 2003.
2. Kahn LH. Confronting zoonoses, linking human and veterinary medicine. Emerg Infect Dis 2006 Apr;12(4):556-561 [FREE Full text] [doi: 10.3201/eid1204.050956] [Medline: 16704801]
3. US Government Accountability Office. Emerging infectious diseases review of state and federal disease surveillance efforts : Report to the Chairman, Permanent Subcommittee on Investigations, Committee on Governmental Affairs, US Senate. Washington, DC: GAO; 2004.
4. Doyle TJ, Ma H, Groseclose SL, Hopkins RS. PHSkb: a knowledge base to support notifiable disease surveillance. BMC Med Inform Decis Mak 2005;5:27 [FREE Full text] [doi: 10.1186/1472-6947-5-27] [Medline: 16105177]
5. Allen H. Reportable animal diseases in the United States. Zoonoses Public Health 2012 Feb;59(1):44-51. [doi: 10.1111/j.1863-2378.2011.01417.x] [Medline: 21824370]
6. M'ikanatha NM, Rohn DD, Robertson C, Tan CG, Holmes JH, Kunselman AR, et al. Use of the internet to enhance infectious disease surveillance and outbreak investigation. Biosecur Bioterror 2006;4(3):293-300. [doi: 10.1089/bsp.2006.4.293] [Medline: 16999590]

7. Staes CJ, Wuthrich A, Gesteland P, Allison MA, Leecaster M, Shakib JH, et al. Public health communication with frontline clinicians during the first wave of the 2009 influenza pandemic. J Public Health Manag Pract 2011;17(1):36-44 [FREE Full text] [doi: 10.1097/PHH.0b013e3181ee9b29] [Medline: 21135659]
8. Pappas G, Fragoulis KN, Falagas ME. World Wide Web resources on zoonotic infections: a subjective overview. Trans R Soc Trop Med Hyg 2008 Dec;102(12):1181-1188. [doi: 10.1016/j.trstmh.2008.04.006] [Medline: 18513766]
9. Aswani MS, Reagan J, Jin L, Pronovost PJ, Goeschel C. Variation in public reporting of central line-associated bloodstream infections by state. Am J Med Qual 2011;26(5):387-395. [doi: 10.1177/1062860611399116] [Medline: 21825038]
10. Marfin AA, Petersen LR, Eidson M, Miller J, Hadler J, Farello C, ArboNET Cooperative Surveillance Group. Widespread West Nile virus activity, eastern United States, 2000. Emerg Infect Dis 2001;7(4):730-735 [FREE Full text] [doi: 10.3201/eid0704.010423] [Medline: 11585539]
11. Crockett RK, Burkhalter K, Mead D, Kelly R, Brown J, Varnado W, et al. Culex flavivirus and West Nile virus in Culex quinquefasciatus populations in the southeastern United States. J Med Entomol 2012 Jan;49(1):165-174. [Medline: 22308785]
12. O'Leary DR, Marfin AA, Montgomery SP, Kipp AM, Lehman JA, Biggerstaff BJ, et al. The epidemic of West Nile virus in the United States, 2002. Vector Borne Zoonotic Dis 2004;4(1):61-70. [doi: 10.1089/153036604773083004] [Medline: 15018774]
13. Marfin AA, Petersen LR, Eidson M, Miller J, Hadler J, Farello C, ArboNET Cooperative Surveillance Group. Widespread West Nile virus activity, eastern United States, 2000. Emerg Infect Dis 2001;7(4):730-735 [FREE Full text] [doi: 10.3201/eid0704.010423] [Medline: 11585539]
14. CDC. FluView URL: http://www.cdc.gov/flu/weekly/ [WebCite Cache ID 6Cq4fWynq]
15. Abernethy NF, DeRimer K, Small PM. Methods to identify standard data elements in clinical and public health forms. In: AMIA Annu Symp Proc. 2011 Presented at: AMIA Annual Symposium; Oct. 22-26, 2011; Washington, DC p. 19-27 URL: http://europepmc.org/abstract/MED/22195051
16. Brownstein JS, Freifeld CC, Reis BY, Mandl KD. Surveillance Sans Frontières: Internet-based emerging infectious disease intelligence and the HealthMap project. PLoS Med 2008 Jul 8;5(7):e151 [FREE Full text] [doi: 10.1371/journal.pmed.0050151] [Medline: 18613747]
17. Brownstein JS, Freifeld CC. HealthMap: the development of automated real-time internet surveillance for epidemic intelligence. Euro Surveill 2007 Nov;12(11):E071129.5 [FREE Full text] [Medline: 18053570]
18. Brownstein JS. HealthMap. 2012. URL: http://healthmap.org/en/ [accessed 2012-12-11] [WebCite Cache ID 6CqXX9x84]
19. Pennsylvania's West Nile Virus Control Program. 2013. URL: http://www.westnile.state.pa.us/surv_2012.htm [accessed 2013-04-17] [WebCite Cache ID 6FwwO6gKx]
20. Yahoo! Dapper. URL: http://open.dapper.net/ [accessed 2012-12-11] [WebCite Cache ID 6Cq4phkmP]
21. Yang Y, Wilson LT, Wang JY. Development of an automated climatic data scraping, filtering and display system. Computers and Electronics in Agriculture 2010;71(1):77-87.
22. Moumtzidou A. Environmental data extraction from multimedia resources. In: Proceedings of the 1st ACM International Workshop on Multimedia Analysis for Ecological Data. 2012 Presented at: 1st ACM International Workshop on Multimedia Analysis for Ecological Data; Oct. 29-Nov. 2, 2012; Nara, Japan p. 13-18.
23. Fernandez-Luque L, Karlsen R, Bonander J. Review of extracting information from the Social Web for health personalization. J Med Internet Res 2011;13(1):e15 [FREE Full text] [doi: 10.2196/jmir.1432] [Medline: 21278049]
24. Doing-Harris KM, Zeng-Treitler Q. Computer-assisted update of a consumer health vocabulary through mining of social network data. J Med Internet Res 2011;13(2):e37 [FREE Full text] [doi: 10.2196/jmir.1636] [Medline: 21586386]
25. Tolentino H. Scanning the Emerging Infectious Diseases Horizon - Visualizing ProMED Emails Using EpiSPIDER. Advances in Disease Surveillance 2007;2:169.
26. Stewart A. A transfer approach to detecting disease reporting events in blog social media. In: Proceedings of the 22nd ACM Conference on Hypertext and Hypermedia. Eindhoven, The Netherlands; 2011 Presented at: 22nd ACM Conference on Hypertext and Hypermedia; June 6-9, 2011; Eindhoven, the Netherlands p. 271-280.
27. Chase V. ProMED: a global early warning system for disease. Environ Health Perspect 1996 Jul;104(7):699 [FREE Full text] [Medline: 8841752]
28. ISID. ProMed. 2012. URL: http://www.promedmail.org/ [accessed 2012-12-11] [WebCite Cache ID 6Cq548wqa]
29. Graham JW. Missing data analysis: making it work in the real world. Annu Rev Psychol 2009;60:549-576. [doi: 10.1146/annurev.psych.58.110405.085530] [Medline: 18652544]
30. FAO. EMPRES-i: Global Animal Disease Information System. 2012. URL: http://empres-i.fao.org/eipws3g/ [accessed 2012-12-11] [WebCite Cache ID 6Cq5CITEk]
31. Boulos MN, Scotch M, Cheung KH, Burden D. Web GIS in practice VI: a demo playlist of geo-mashups for public health neogeographers. Int J Health Geogr 2008;7:38 [FREE Full text] [doi: 10.1186/1476-072X-7-38] [Medline: 18638385]
32. Yi Q, Hoskins RE, Hillringhouse EA, Sorensen SS, Oberle MW, Fuller SS, et al. Integrating open-source technologies to build low-cost information systems for improved access to public health data. Int J Health Geogr 2008;7:29 [FREE Full text] [doi: 10.1186/1476-072X-7-29] [Medline: 18541035]

33. Regis L, Souza WV, Furtado AF, Fonseca CD, Silveira Jr JC, Ribeiro Jr PJ, et al. An entomological surveillance system based on open spatial information for participative dengue control. An Acad Bras Cienc 2009 Dec;81(4):655-662 [FREE Full text] [Medline: 19893891]
34. Gesteland PH. Clinician Use and Acceptance of Population-Based Data about Respiratory Pathogens: Implications for Enhancing Population-Based Clinical Practice. In: AMIA Annu Symp Proc. 2008 Presented at: AMIA Annual Symposium; Nov. 8-12, 2008; Washington, DC p. 232-236.

## Abbreviations

**ANOVA:** analysis of variance
**CDC:** Centers for Disease Control and Prevention
**CLABSI:** central line–associated bloodstream infection
**CSV:** comma-separated values
**EEE:** Eastern Equine Encephalitis
**EMPRES:** Emergency Prevention System
**FAO:** Food and Agriculture Organization
**HHS:** Health and Human Services
**HTML:** HyperText Markup Language
**NLP:** Natural Language Processing
**PDF:** Portable Document Format
**SLE:** St. Louis encephalitis
**USDA:** United States Department of Agriculture
**WEE:** Western equine encephalitis
**WNV:** West Nile virus

*Edited by G Eysenbach; submitted 13.12.12; peer-reviewed by A Turner, C Staes, C Freifeld, F Oprescu, D Uglow; comments to author 04.01.13; revised version received 25.01.13; accepted 12.04.13; published 29.04.13.*

*Please cite as:*
*Scotch M, Baarson B, Beard R, Lauder R, Varman A, Halden RU*
*Examining the Differences in Format and Characteristics of Zoonotic Virus Surveillance Data on State Agency Websites*
*J Med Internet Res 2013;15(4):e90*
*URL: http://www.jmir.org/2013/4/e90/*
*doi:10.2196/jmir.2487*
*PMID:23628771*

Original Paper

# Health-Related Effects Reported by Electronic Cigarette Users in Online Forums

My Hua[1]; Mina Alfi[1], BS; Prue Talbot[1], BS, MS, PhD

University of California, Department of Cell Biology and Neuroscience, University of California, Riverside, CA, United States

**Corresponding Author:**
Prue Talbot, BS, MS, PhD
University of California
Department of Cell Biology and Neuroscience
University of California
900 Univeristy Avenue
Riverside, CA, 92521
United States
Phone: 1 950 827 3768
Fax: 1 950 827 4286
Email: talbot@ucr.edu

## Abstract

**Background:** The health effects caused by electronic cigarette (e-cigarette) use are not well understood.

**Objective:** Our purpose was to document the positive and negative short-term health effects produced by e-cigarette use through an analysis of original posts from three online e-cigarettes forums.

**Methods:** Data were collected into Microsoft Access databases and analyzed using Cytoscape association graphics, frequency distributions, and interactomes to determine the number and type of health effects reported, the organ systems affected the frequency of specific effects, and systems interactions.

**Results:** A total of 405 different symptoms due to e-cigarette use were reported from three forums. Of these, 78 were positive, 326 were negative, and one was neutral. While the reported health effects were similar in all three forums, the forum with the most posts was analyzed in detail. Effects, which were reported for 12 organ systems/anatomical regions, occurred most often in the mouth and throat and in the respiratory, neurological, sensory, and digestive systems. Users with negative symptoms often reported more than one symptom, and in these cases interactions were often seen between systems, such as the circulatory and neurological systems. Positive effects usually occurred singly and most frequently affected the respiratory system.

**Conclusions:** This is the first compilation and analysis of the health effects reported by e-cigarette users in online forums. These data show that e-cigarette use can have wide ranging positive and negative effects and that online forums provide a useful resource for examining how e-cigarette use affects health.

***(J Med Internet Res 2013;15(4):e59)*** doi:10.2196/jmir.2324

**KEYWORDS**

Electronic cigarettes; e-cigarettes; electronic nicotine delivery devices; ENDS; health effects; nicotine; harm reduction; symptoms; Internet

## Introduction

New nicotine delivery products, often advertised as “safer” or with “harm reduction”, are designed to reduce health problems caused by cigarette smoking [1,2]. Electronic cigarettes, also called e-cigarettes or electronic nicotine delivery systems, are one such product that has rapidly gained popularity worldwide [3,4]. E-cigarettes produce aerosol by heating a humectant (usually propylene glycol or vegetable glycerol) containing nicotine and flavorings without actually burning tobacco. The aerosol, when inhaled, delivers nicotine to the user [5,6]. Users can purchase either prepackaged cartridges/cartomizers with varying amounts of nicotine or bottles of fluid for refilling empty cartridges/cartomizers [7]. E-cigarettes vary widely in their performance with respect to pressure drop, air flow rate required to activate the battery, aerosol delivery, number of puffs produced per cartridge, and in the amount of nicotine/puff [7-9]. Puff duration data mined from YouTube videos showed that e-cigarette users on average take puffs that are over twice as

long as conventional smokers [10]. Because e-cigarette production is currently not regulated, quality control during manufacture has been questioned [11-14].

While e-cigarettes may facilitate smoking cessation [15-17], users are concerned about product safety and toxicity [5]. Most knowledge regarding the health effects of e-cigarette use comes from studies on the emissions in their aerosol [14,18]. Because e-cigarettes deliver fewer total chemicals and fewer carcinogens than conventional tobacco-burning cigarettes, they are sometimes considered safer products [18,19]. However, e-cigarette cartridge fluids and their emissions are not yet well characterized and may vary among products. In a recent study, e-cigarette refill fluids varied significantly in their cytotoxicity when tested *in vitro* with embryonic and adult cells [20]. One cartridge analyzed by the US Food and Drug Administration contained diethylene glycol, a known toxicant [21], and errors in labeling of nicotine concentrations have been reported [12,14]. Moreover, performance of e-cigarettes is highly variable both between and within brands. A recent study of e-cigarette users showed that 5 minutes of puffing adversely affected lung physiology [22,23], indicating that a better understanding of the health effects related to e-cigarette use is needed.

The Internet has become a useful resource for consumers seeking health information online. Moreover, Internet sites can be mined to acquire data relevant to issues dealing with public health, a science referred to as infodemiology [24-26]. Several prior infodemiological-type studies have produced informative data related specifically to tobacco products [3,10,27]. E-cigarette forums have an increasingly popular presence on the Internet where e-cigarette users can post data relevant to the health effects they experience from using these products. In this study, we used an infodemiological approach to acquire and analyze self-reported health effects posted on Internet forums by e-cigarette users. While medical and lifestyle histories are not known for the individuals in this study, these Internet data indicate that a broad spectrum of symptoms may accompany e-cigarette use and that further studies are needed in this area.

## Methods

A Google Internet search was performed using the words “electronic cigarette forum” to identify online e-cigarette forums with “health and safety sections” that allowed posts on the health effects experienced when using e-cigarettes. The three websites with the highest number of posts in health and safety sections were selected for study (Electronic Cigarette Forum posts = 543, Vapers Forum posts = 34, and Vapor Talk posts = 55). Data were collected from posts on these websites through July 15, 2011, and the Electronic Cigarette Forum, which had the most entries, was analyzed in detail. Only data reported directly by an e-cigarette user were included.

Databases were created using Microsoft Access to record basic information (age, gender, location) and positive and negative health effects. These data were recorded for persons who reported their own health effects and excluded those that reported health effects of others. Basic user information was gathered by accessing the individual’s profile pages upon entering their posts. Data were also recorded for users who visited a physician/dentist or emergency room and self-reported their diagnosis on a forum website. These entries were recorded as “signs”.

To ensure health reports were not duplicated or over-reported, individual symptoms that users reported were grouped under their coded user name. Any duplicate reports of symptoms were omitted and counted only once.

Data were analyzed iteratively. In the first analysis, all health-related effects reported by e-cigarette users were grouped according to the organ system/anatomical region, which we define as systems. Some effects, such as improved overall health, could not be categorized by system and were kept separately. When users described their effects with synonyms (eg, fatigue and lethargy), the effects were combined using one term, such as fatigue. Sometimes symptoms were characterized by degree, such as severe stomach cramps or stomach cramps, in which case the symptoms were grouped in the category “stomach cramps”. When a symptom could have been associated with more than one system, the effect was assigned to the system for which it had the strongest fit (eg, improved sense of taste was assigned to sensory but could have been mouth/throat). An association graph was created using Cytoscape software (an interactome creation and analysis program) to group the health effects by systems.

Following the initial analysis, health-related effects were consolidated within a system to make the data more manageable. For example, there were many symptoms involving the tongue, such as swollen tongue, red tongue, and bumps on tongue. These were all grouped under “tongue” for further analysis. Frequency distributions for the grouped data in each system were plotted using MS Excel.

Information pertaining to individual users and the systems affected by e-cigarette use was transferred from Access to Excel spreadsheets and uploaded to Cytoscape software. Interactomes were then created using the edge-weighted spring embedded or force directed views.

## Results

### Demographics

The Electronic Cigarette Forum had the largest number of users (Table 1). Because some of these users posted multiple times, the total number of health-related posts attributed to e-cigarettes (N=543) was greater than the total number of users (N=481). Most users posted symptoms only (n=492) (ie, health effects perceived by the individual), while some reported both symptoms and signs (n=20) (ie, effects diagnosed by a physician then self-reported in the forum by the user), or signs only (n=31). Users’ ages ranged from 18-71 years with the highest number being in the 26-35 year age bracket. For the few users that provided gender on the Electronic Cigarette Forum, more females self-reported health effects than males. Most users were from North America. Demographics for the two smaller forums (Vapers Forum and Vapor Talk) are also provided in Table 1.

**Table 1.** Demographics of forum users.

| Characteristics | | | ECF[a] | VF[b] | VT[c] |
|---|---|---|---|---|---|
| **Self-reported ages in years** | | | | | |
| | 18-25 | | 46 | N/A | N/A |
| | 26-35 | | 107 | N/A | 8 |
| | 36-45 | | 60 | 1 | 8 |
| | 46-55 | | 31 | 1 | 4 |
| | >56 | | 17 | N/A | 1 |
| | Did not state | | 220 | 29 | 27 |
| **Gender** | | | | | |
| | Male | | 10 | N/A | 24 |
| | Female | | 24 | N/A | 14 |
| | Did not state | | 447 | 31 | 10 |
| **Location** | | | | | |
| | North America | | | | |
| | | US Midwest | 60 | 4 | 8 |
| | | US Northeast | 70 | 2 | 7 |
| | | US Southeast | 84 | 5 | 10 |
| | | US Southwest | 30 | N/A | 4 |
| | | US West | 69 | 5 | 11 |
| | | US territories | 2 | N/A | N/A |
| | | US unspecified | 18 | N/A | N/A |
| | | Canada | 14 | 2 | 3 |
| | Europe | | | | |
| | | UK | 13 | N/A | N/A |
| | | Ireland | 3 | N/A | N/A |
| | | Other | 6 | N/A | N/A |
| | Other regions | | | | |
| | | Africa | 4 | N/A | N/A |
| | | Australia, New Zealand | 7 | 1 | N/A |
| | | Asia | 6 | N/A | N/A |
| | | Central America | 1 | N/A | N/A |
| | Did not state | | 94 | 12 | 5 |
| **Total number users evaluated** | | | 481 | 31 | 48 |
| **Total number of posts** | | | 543 | 34 | 55 |

[a] ECF = Electronic Cigarette Forum

[b] VF = Vapers Forum

[c] VT = Vapor Talk

## Health Effect-System Associations

A total of 388 different symptoms were reported by e-cigarette users on the Electronic Cigarette Forum (Figure 1). Most health effects were broadly distributed across 12 different categories. These categories included 10 organ systems (eg, respiratory, neurological) and two anatomical regions (chest and mouth/throat), which we collectively refer to as systems. Respiratory, mouth/throat, neurological, and sensory had the most symptoms associated with them. Mouth and throat had more negative symptoms than any other group. A significant number of health effects appeared in the digestive, muscular/skeletal, and integumentary systems, while the urogenital, immune, and endocrinological systems had relatively

few symptoms. Seventeen reported symptoms were not associated with a system. (The legend for Figure 1 is as follows: IMPVD = improved; Elim = eliminated; REC = recurring; PRSRT = persistent; UNS = unspecified; SENS = sensation; OCD = obsessive compulsive disorder; COPD = chronic obstructive pulmonary disease; COMP = complications; PO = post-operative; w/ = with.)

Of the reported effects, 318 were negative, 69 were positive, and one was neutral (Figure 1). All systems had both positive and negative symptoms, except for urogenital, which had only negative effects. Negative symptoms were often described as persistent, worsened, or increasing. In contrast, positive effects were often described as decreased, improved, or eliminated. Negative and positive effects were sometimes opposites, eg, improved cough and worsened cough were reported by different individuals. Some symptoms (17) were not assigned to a category. Negative symptoms in this group included “swelling” and “dehydration”, while positive included “improved stamina” and “improved overall health”. Some symptoms occurred during e-cigarette use, such as “metal taste in mouth”, while others occurred after use, such as “choking after use”. Of the symptoms not associated with a system, the most frequently reported were: no symptoms [4], improved exercise endurance (6), and dehydration (7).

## Symptom Frequency

To analyze the frequency of reports for various symptoms, the data in Figure 1 were condensed by combining all health effects into structural or physiological groups. For example, red tongue, swollen tongue, and bumps on tongue were all combined into the structural group “tongue”, and constipation, indigestion, and frequent bowel movements were combined into the physiological group “intestine/digestion”. The frequency of positive and negative reports in each structural/physiological group was graphed for each system (Figure 2). In Figure 2, n=the total number of health effects (positive and negative) for each system, and data in each column are plotted as a percentage of the total number of effects reported for each system. Urogenital and endocrinological data are not shown because there were few reports in these categories. For each system in Figure 2, the negative effects outnumbered the positive effects. In most graphs, there were one or two negative structural/physiological groups that were dominant. For example, the bronchi/lungs were frequently affected in “respiratory”, and sight was frequently affected in “sensory”. Some structural/physiological groups had only negative effects, such as acne in “integumentary”, tongue in “mouth and throat”, and aches/pain in “muscular/skeletal”. In respiratory, digestive, sensory, and immune, each condensed group had both positive and negative symptoms. One physiological group, smell in “sensory”, had only positive reports. The neurological system had the largest number of different groups, which included headaches, dizziness, and temperature regulation.

**Figure 1.** Summary of all positive (green), negative (red),and neutral (white) symptoms and their associated systems as reported by e-cigarette users on the Electronic Cigarette Forum.

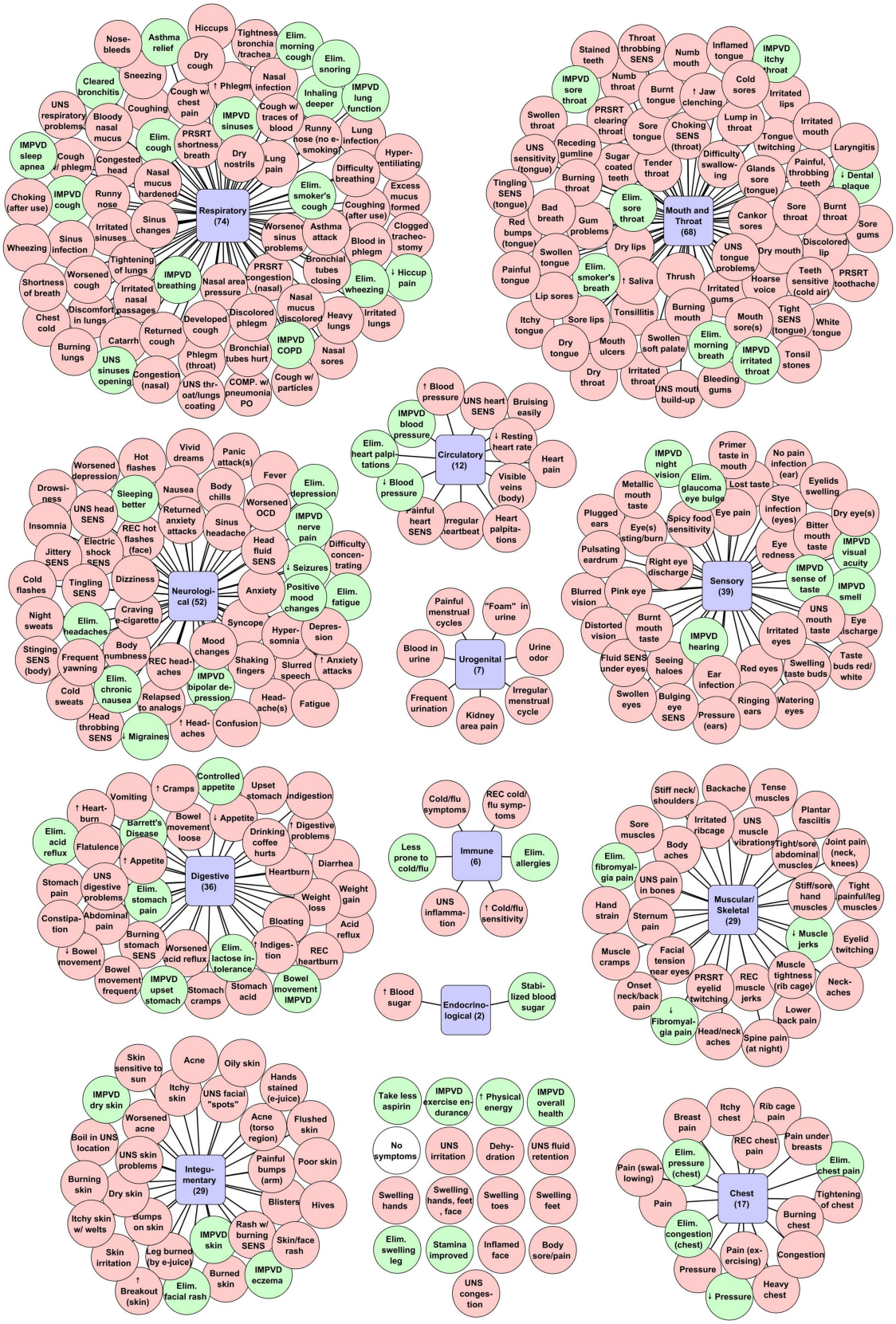
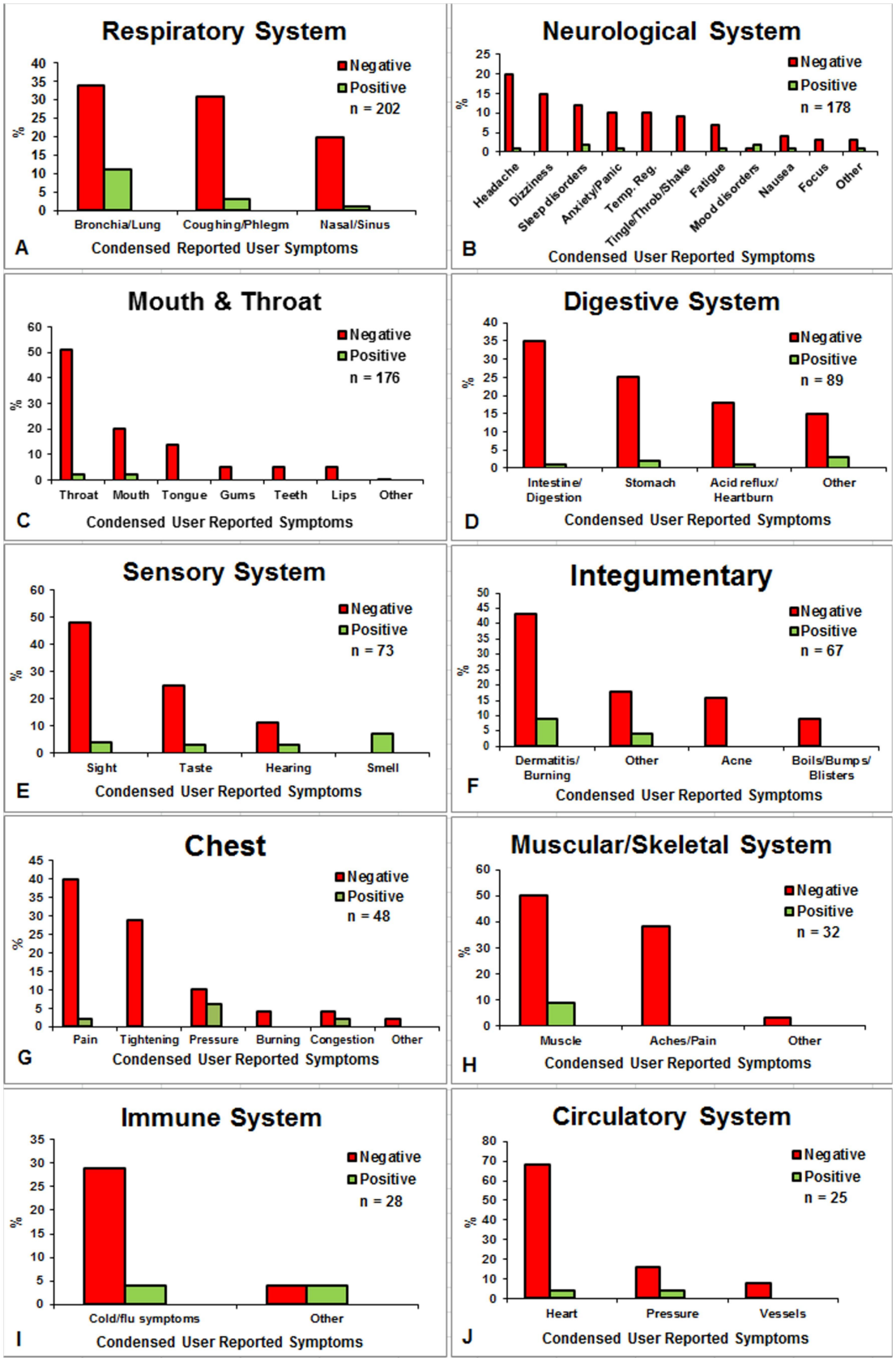


**Figure 2.** Frequency of positive and negative symptoms for 10 systems.




## Interactomes

Electronic Cigarette Forum data were further analyzed by creating an interactome showing the relationship between the negative symptoms reported by each individual and the systems with which they interacted (Figure 3). Individuals in the interactome are represented by small numbered circles, while systems appear as rectangular purple nodes. Those reporting symptom(s) for one system (n=276, red circles) are generally at the outer edges of the interactome. Individuals reporting two

systems (n=85, teal circles) are grouped more toward the center of the interactome or between adjacent systems with which they interact. Those reporting health effects in three (n=34, purple circles), four (n=12, pink circles), five (n=5, light blue circles), or six (n=2, green circles) systems are grouped in the interior of the interactome.

For most users, a single system was affected. However, 34% of the individuals had negative effects in more than one system. For individuals reporting effects in two systems, clear interactions were seen between: mouth/throat and respiratory; respiratory and chest; mouth/throat and sensory; sensory and neurological; mouth/throat and neurological; neurological and respiratory; digestive and neurological, and neurological and circulatory. Although a number of individuals reported symptoms in immune, there was no clear interaction between immune and other systems. There was relatively little interaction with urogenital and no interaction with endocrinological.

The corresponding positive interactome for the Electronic Cigarette Forum had relatively few users (n=58) (Figure 4). Two-way interactions were seen mainly between respiratory and sensory, respiratory and mouth/throat, and respiratory and chest. For most users, a single system was affected, although 2 users reported positive effects in as many as four different systems.

The color-coding in Figure 4 shows the number of systems affected for each user (red = users reporting symptom(s) for one system; teal = users reporting symptom(s) for two systems; purple = users reporting symptom(s) for three systems; pink = users reporting symptom(s) for four systems).

**Figure 3.** Interactome showing relationship between users who reported negative symptoms and the systems that were affected.

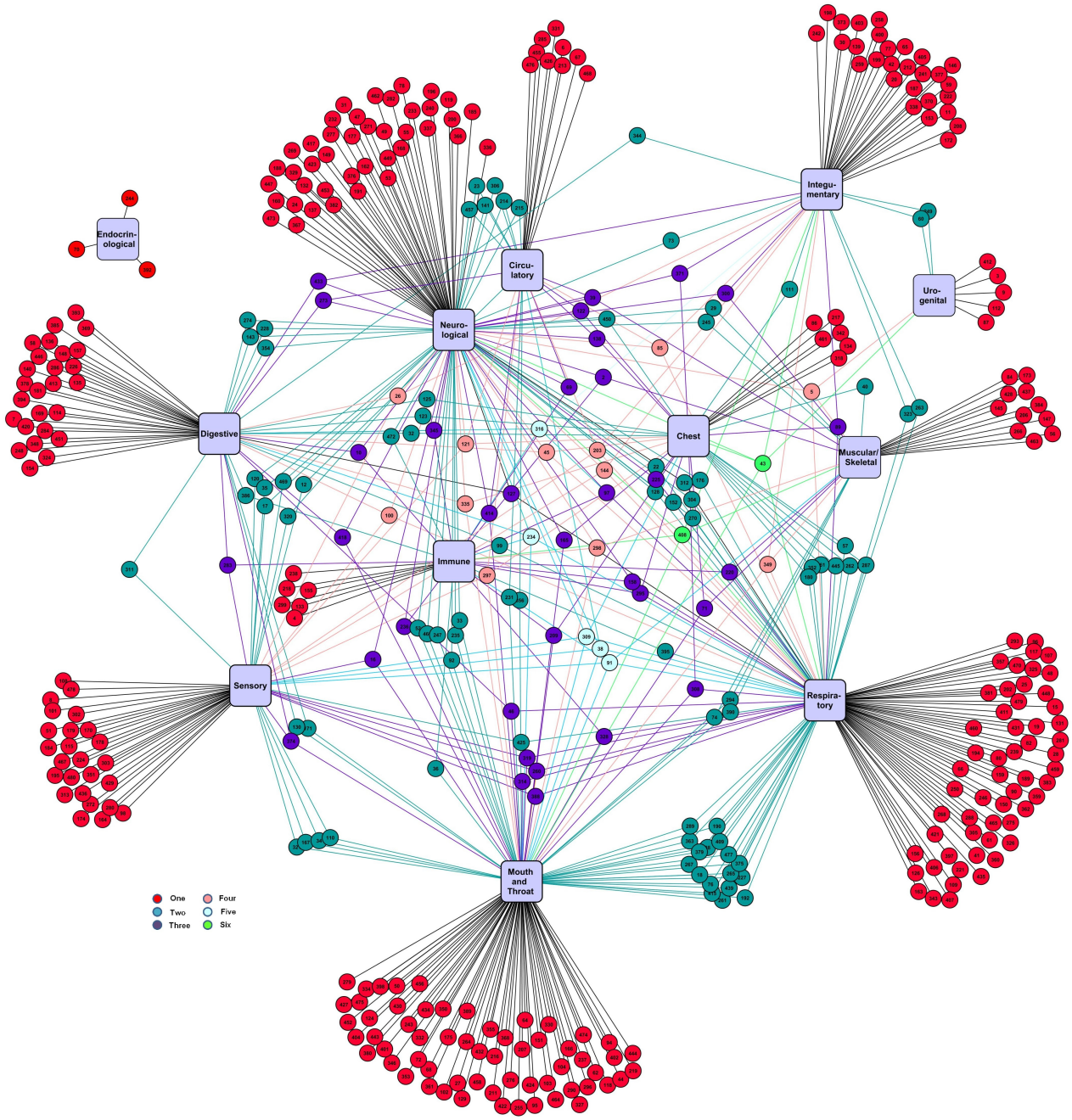
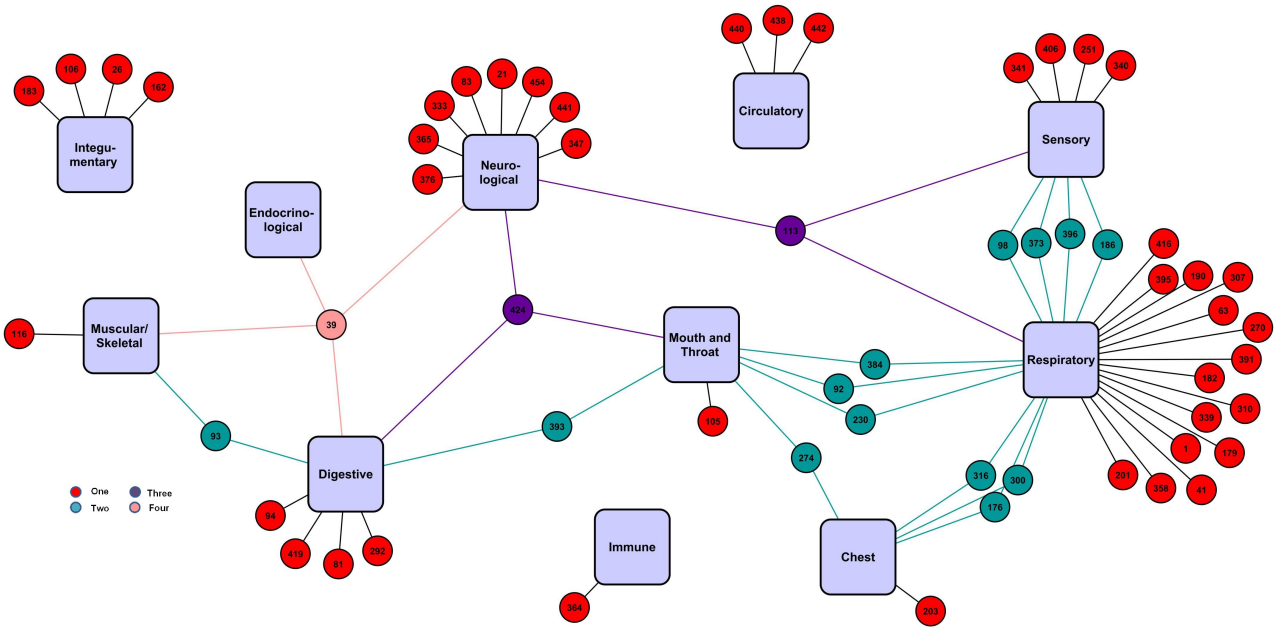


**Figure 4.** Positive interactome for the Electronic Cigarette Forum.




## Time Frame

In the Electronic Cigarette Forum, 66 users reported a time frame for the appearance of their symptoms. Forty-seven users stated that symptoms occurred 1 week or less after use began. The remaining 19 reported that symptoms occurred more than 1 week after use began.

## Signs

Some e-cigarette users reported diagnoses (signs) made by physicians or dentists (Multimedia Appendix 1). Most signs occurred in the circulatory system, respiratory system, or mouth/throat. Several negative health effects that did not appear in the self-reported data were diagnosed by physicians and dentists, such as periodontitis, rhinitis, parethesia, cataract development, and anemia. Some positive health effects that were diagnosed but did not appear in user-reported symptoms (Figure 1) were whiter teeth, improved gum health, and improved spirometry test results.

## Other Forums

Interactomes for the Vapor Talk (Figure 5) and Vapers Forums (Figure 6) show the relationship between users (small circles), negative symptoms (light pink diamonds) and positive symptoms (light green diamonds), and systems (purple rectangles). Data for these smaller forums were similar to the Electronic Cigarette Forum. In both small forums, more negative than positive symptoms were posted, and most effects were reported for “respiratory”, “mouth/throat”, and “neurological”. Most users reported only one effect that generally appeared near the outside edge of the interactome. Those reporting two and or more effects are grouped toward the center of the interactome. There were not enough individuals in the two smaller forums to observe interactions. Very few individuals in the two smaller forums reported signs (Multimedia Appendix 1). The color-coding in Figures 5 and 6 shows the number of systems affected for each user (red circles = users reporting symptom(s) for one system; teal circles = users reporting symptom(s) for two systems; purple circles = users reporting symptom(s) for three systems; light blue = users reporting symptom(s) for five systems).

**Figure 5.** Interactome for Vapers Forum.

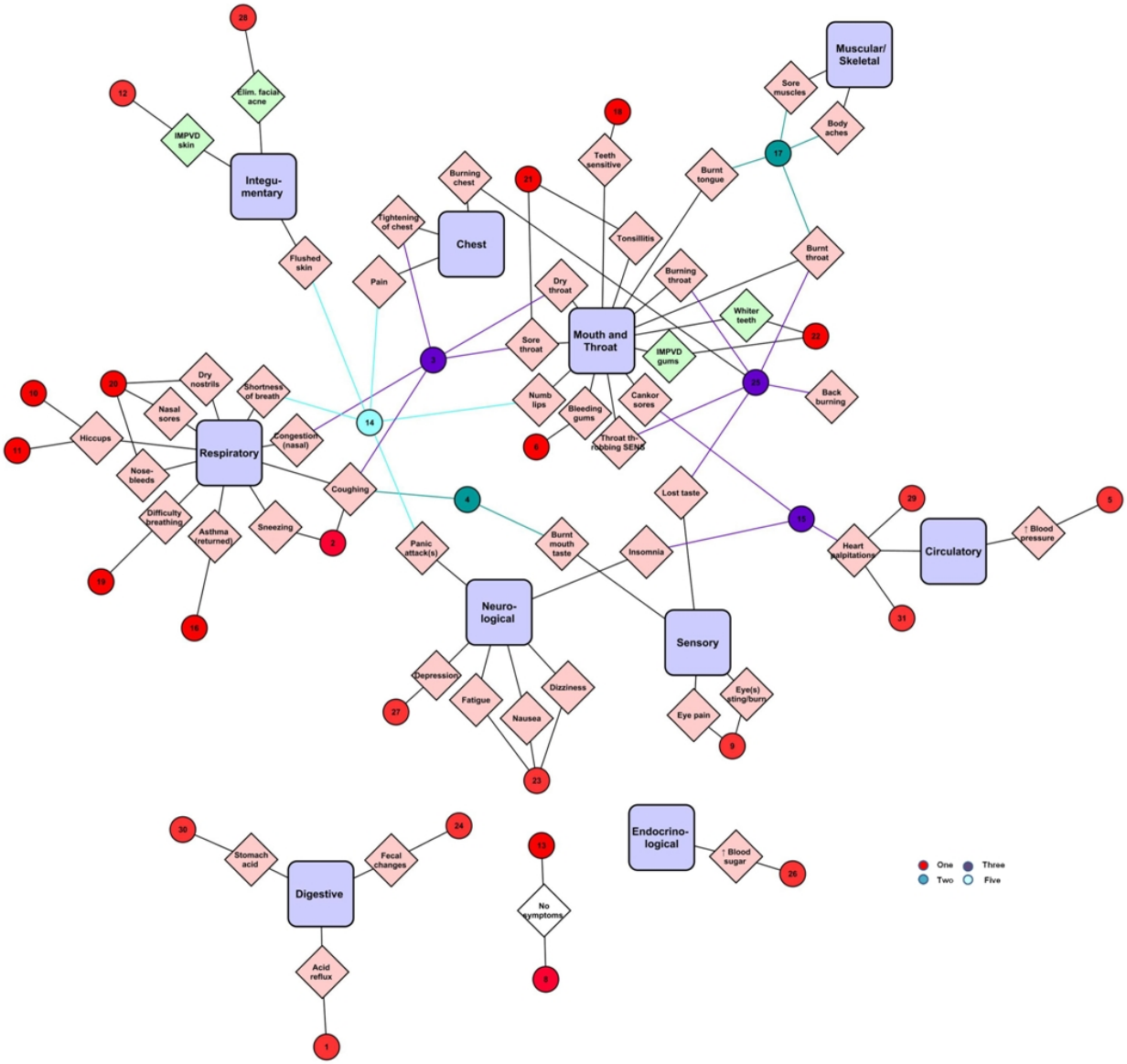
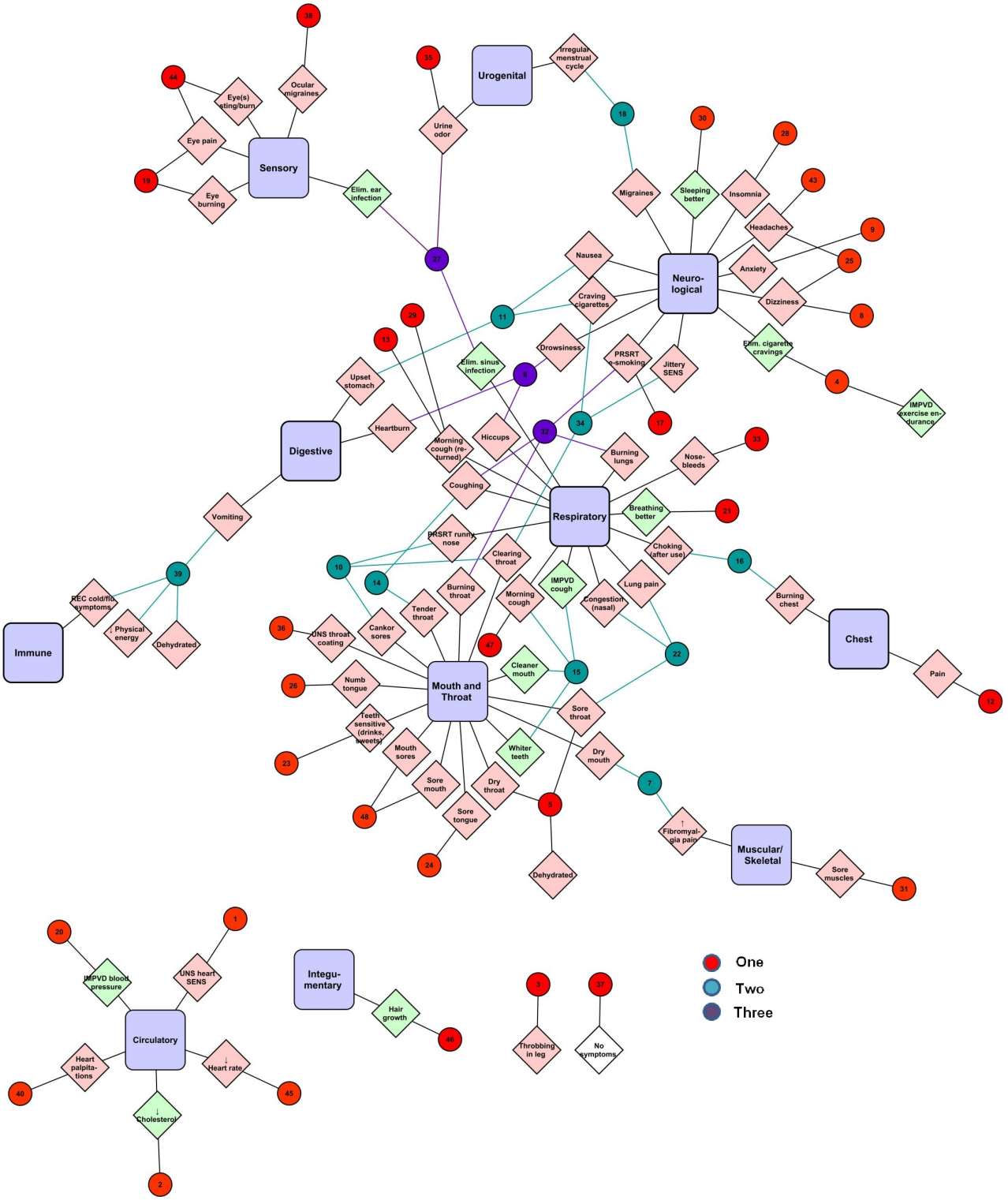


**Figure 6.** Interactome for Vapor Talk.




## Discussion

Information on the health effects related to e-cigarette use is important to users, health professionals, and regulatory agencies. In this study, we have taken an infodemiological approach (ie, used information on the Internet for public health research) [26] to evaluate the health effects reported by e-cigarette users. Our data, while preliminary, show that a broad spectrum of positive and negative symptoms attributed to e-cigarette use has been self-reported in online forums. Our data are in general agreement with prior studies and extend earlier work by demonstrating the breadth of e-cigarette related health effects, determining the frequency of occurrence of specific effects for 12 systems, and demonstrating interactions between systems.

Two prior studies deal with the effects of e-cigarette use on blood pressure and heart rate. When 32 participants consumed one e-cigarette cartridge/day for 4 weeks, no abnormal changes in blood pressure were observed [28]. In contrast, we found,

based on a larger number of participants, that blood pressure changes were reported by 3.5% of e-cigarette users (5 individuals with symptoms and 12 individuals with signs), and increased blood pressure was the sign most frequently diagnosed in e-cigarette users by physicians (n=9). Increased heart rate was observed in a study in which increased levels of plasma nicotine were verified during e-cigarette use [6]. This agrees with our study in which some e-cigarette users reported changes in heart rate and palpitations. Although we found that e-cigarette users reported relatively few different symptoms for the circulatory system, some, such as increased blood pressure, may not be perceived by the user and could have significant health impacts if left untreated.

Several prior health-related studies on e-cigarettes deal with the respiratory system and mouth/throat. In our study, these systems had the largest number of different symptoms associated with them. In an Internet survey of e-cigarette users, positive effects were reported to be improved breathing/respiration, less coughing, fewer sore throats, improved fitness, and reduced bad breath [5], similar to the main positive effects found in our study. We found additional positive effects such as “improved COPD [chronic obstructive pulmonary disease]”, “asthma relief”, and “cleared bronchitis”. However for most positive effects, there were corresponding negative reports, such as “developed cough” or “cough worsened”, suggesting that individual responses to e-cigarette aerosol can vary and be opposite. In a study of smoking reduction, mouth and throat irritation as well as dry cough were common, but diminished by 24 weeks [29]. These symptoms were reported in our study with high frequency. We found that some users had immediate reaction to e-cigarettes, such as coughing or choking, or same-day reaction (insomnia), while others reported that symptoms appeared at least a week after beginning e-cigarette use. When respiratory physiology was monitored following 5 minutes of ad libitum use of e-cigarettes by healthy non-smokers, a significant decrease in exhaled nitric oxide and an increase in pulmonary resistance were observed, similar to effects seen during use of conventional cigarettes [22,23]. We did find reports of “tightening of the lungs” and “difficulty breathing”, which suggest increased pulmonary resistance. In all the above studies, the reported health effects are short term.

In general, the negative short-term effects reported by e-cigarette users appear relatively minor compared to more serious long-term conditions (eg, cancer and stroke) that occur in conventional smokers [30]. Since some symptoms associated with e-cigarette use appear to be minor, such as sneezing and dry skin, and may not impact overall health and life style, e-cigarette users may consider these effects to be insignificant compared to the health consequences of conventional smoking. Other effects correlated with e-cigarette use may forewarn of significant health problems. For example, dizziness could have a major impact on a user’s life style if it were serious enough to prevent driving or working, as was reported to us by an e-cigarette user. Users need to be aware that e-cigarettes are not free of health consequences and that numerous negative effects have been reported. For individuals who do not smoke conventional cigarettes and who do not currently use e-cigarettes, knowledge of these health effects may be important in helping them decide if they wish to begin e-cigarette use.

Many e-cigarette users reported multiple positive or negative symptoms that often affected more than one system, as shown in Figures 3 and 4. The negative interactome for the Electronic Cigarette Forum demonstrated that when multiple systems are affected in individual e-cigarette users, there is often interaction between systems, such as respiratory and chest, respiratory and mouth/throat, and digestive and neurological. This information may be useful to physicians treating patients who are using e-cigarettes and/or have a prior history of smoking. The data may also be used to help diagnose health problems in patients for whom smoking/e-cigarette history is not known. Although the positive interactome for the Electronic Cigarette Forum had fewer users, interactions between systems were similar to those observed in the negative interactome.

In the negative Electronic Cigarette Forum interactome, many symptoms involved the neurological and sensory systems, and these may have been caused by an overdose or withdrawal of nicotine, which activates nicotinic cholinergic receptors in the brain [31]. Some symptoms reported by e-cigarette users, such as vomiting, nausea, changes in heart rhythm, confusion, dizziness, and fatigue, would be consistent with a nicotine overdose [32]. Because nicotine can be added to cartomizers/cartridges, the potential to overdose exists. It is also possible that some reported symptoms, such as anxiety and depression, which are characteristic of nicotine withdrawal [33], were caused by insufficient delivery of nicotine to some e-cigarette users. As previously shown, inexperienced users of e-cigarettes may not receive adequate levels of nicotine [34].

While our study demonstrates the value of an infodemiological approach, it has several limitations. The online forums may be biased toward negative reporting, and the positive effects are probably under-reported in our analysis. In the current study, users posted effects that they observed after starting e-cigarette use and that they therefore attribute the effects to e-cigarettes. However, these results do not take into account past smoking and lifestyle history, pre-existing conditions, or other medical problems the users may have had before e-cigarette use began, any of which may have affected or even aggravated their response to e-cigarette aerosol. In addition, user posts cannot be validated, and it is possible that inaccurate information appears in the forums. Future work should be undertaken using methods that would acquire additional medical and lifestyle history of each user and that draw participants from a random sample of e-cigarette users.

Additionally, our data do not address the effects that e-cigarette use may have on prenatal development, a period in the human life cycle that is particularly sensitive to environmental chemicals [35,36]. Our recent study found that some e-cigarette refill fluids were highly toxic to human embryonic stem cells [20], indicating the need for further work in this area.

Our data will be helpful to e-cigarette users who may experience similar effects, to health care professionals advising individuals on e-cigarette usage, and to policy makers and legislators who regulate sales, use, and marketing of e-cigarettes. Interactome data may help health care workers identify problems caused by

e-cigarette use, eg, problems with the respiratory system are often linked to problems with the neurological system. While it can be argued that e-cigarettes are safer than conventional tobacco-burning products, the data in this study demonstrate that e-cigarette users are not free of negative health effects.

### Conclusions

This study provides a preliminary synopsis of the short-term health effects related to e-cigarette use as reported in online forums. Effects occurred most often in the respiratory and neurological systems and in the mouth and throat with a total of 12 systems being affected. Within each system, certain categories of health related-effects were dominant, such as throat problems and headache. As more individuals adopt e-cigarettes and use them for longer periods of time, additional positive and negative effects on human health will likely be reported. This study is a step in understanding these issues; however, it will be many years before the long-term health consequences of e-cigarette use are known.

### Acknowledgments

This work was supported by the Tobacco-Related Disease Research Program of California grant No. 20XT-0118, University of California, Riverside, Medical Scholar's Summer Research Program (MSSRP), Chancellor's Research Fellowship, and the Academic Senate. We are grateful to Vinay Seshachellam and Taylor Gipson for their help saving the data and to all members of the lab for their suggestions on the manuscript.

### Conflicts of Interest

None declared.

### Multimedia Appendix 1

Supplementary table.

[PDF File (Adobe PDF File), 98KB - jmir_v15i4e59_app1.pdf ]

### References

1. Hatsukami DK, Henningfield JE, Kotlyar M. Harm reduction approaches to reducing tobacco-related mortality. Annu Rev Public Health 2004;25:377-395. [doi: 10.1146/annurev.publhealth.25.102802.124406] [Medline: 15015926]
2. Warner KE. Will the next generation of "safer" cigarettes be safer? J Pediatr Hematol Oncol 2005 Oct;27(10):543-550. [Medline: 16217258]
3. Ayers JW, Ribisl KM, Brownstein JS. Tracking the rise in popularity of electronic nicotine delivery systems (electronic cigarettes) using search query surveillance. Am J Prev Med 2011 Apr;40(4):448-453. [doi: 10.1016/j.amepre.2010.12.007] [Medline: 21406279]
4. Yamin CK, Bitton A, Bates DW. E-cigarettes: a rapidly growing Internet phenomenon. Ann Intern Med 2010 Nov 2;153(9):607-609. [doi: 10.1059/0003-4819-153-9-201011020-00011] [Medline: 21041581]
5. Etter JF. Electronic cigarettes: a survey of users. BMC Public Health 2010;10:231 [FREE Full text] [doi: 10.1186/1471-2458-10-231] [Medline: 20441579]
6. Vansickel A, Eissenberg T. Electronic cigarettes: effective nicotine delivery after acute administration. Nicotine Tob Res 2013 Jan;15(1):267-270. [doi: 10.1093/ntr/ntr316] [Medline: 22311962]
7. Williams M, Talbot P. Variability among electronic cigarettes in the pressure drop, airflow rate, and aerosol production. Nicotine Tob Res 2011 Dec;13(12):1276-1283. [doi: 10.1093/ntr/ntr164] [Medline: 21994335]
8. Trtchounian A, Williams M, Talbot P. Conventional and electronic cigarettes (e-cigarettes) have different smoking characteristics. Nicotine Tob Res 2010 Sep;12(9):905-912. [doi: 10.1093/ntr/ntq114] [Medline: 20644205]
9. Goniewicz M, Kuma T, Gawron M, Knysak J, Kosmider L. Nicotine levels in electronic cigarettes. Nicotine Tob Res 2013 Jan;15(1):158-166. [doi: 10.1093/ntr/nts103] [Medline: 22529223]
10. Hua M, Yip H, Talbot P. Mining data on usage of electronic nicotine delivery systems (ENDS) from YouTube videos. Tob Control 2013 Mar;22(2):103-106. [doi: 10.1136/tobaccocontrol-2011-050226] [Medline: 22116832]
11. Hadwiger ME, Trehy ML, Ye W, Moore T, Allgire J, Westenberger B. Identification of amino-tadalafil and rimonabant in electronic cigarette products using high pressure liquid chromatography with diode array and tandem mass spectrometric detection. J Chromatogr A 2010 Nov 26;1217(48):7547-7555. [doi: 10.1016/j.chroma.2010.10.018] [Medline: 20980012]
12. Trehy M, Ye W, Hadwiger ME, Moore TW, Allgire JF, Woodruff JT, et al. Analysis of electronic cigarette cartridges, refill solutions, and smoke for nicotine and nicotine related impurities. Journal of Liquid Chromatography & Related Technologies 2011 Aug 2011;34(14):1442-1458. [doi: 10.1080/10826076.2011.572213]
13. Trtchounian A, Talbot P. Electronic nicotine delivery systems: is there a need for regulation? Tob Control 2011 Jan;20(1):47-52. [doi: 10.1136/tc.2010.037259] [Medline: 21139013]
14. Westenberger BJ. Evaluation of e-cigarettes. St Louis, MO: Food and Drug Administration; 2009:1-8.

15. Caponnetto, Polosa R, Russo C, Leotta C, Campagna D. Successful smoking cessation with electronic cigarettes in smokers with a documented history of recurring relapses: a case series. J Med Case Rep 2011;5(1):585 [FREE Full text] [doi: 10.1186/1752-1947-5-585] [Medline: 22185668]
16. Dawkins L, Turner J, Hasna S, Soar K. The electronic-cigarette: effects on desire to smoke, withdrawal symptoms and cognition. Addict Behav 2012 Aug;37(8):970-973. [doi: 10.1016/j.addbeh.2012.03.004] [Medline: 22503574]
17. Etter JF, Bullen C. Electronic cigarette: users profile, utilization, satisfaction and perceived efficacy. Addiction 2011 Nov;106(11):2017-2028. [doi: 10.1111/j.1360-0443.2011.03505.x] [Medline: 21592253]
18. Laugesen M. Safety report on the Ruyan e-cigarette cartridge and inhaled aerosol. New Zealand: Health New Zealand; 2008:1-22.
19. Cahn Z, Siegel M. Electronic cigarettes as a harm reduction strategy for tobacco control: a step forward or a repeat of past mistakes? J Public Health Policy 2011 Feb;32(1):16-31. [doi: 10.1057/jphp.2010.41] [Medline: 21150942]
20. Bahl V, Lin S, Xu N, Davis B, Wang YH, Talbot P. Comparison of electronic cigarette refill fluid cytotoxicity using embryonic and adult models. Reprod Toxicol 2012 Dec;34(4):529-537. [doi: 10.1016/j.reprotox.2012.08.001] [Medline: 22989551]
21. FDA. FDA warns of health risks posed by e-cigarettes.: Food and Drug Adminstration; 2009. URL: http://www.fda.gov/forconsumers/consumerupdates/ucm173401.htm [accessed 2013-02-28] [WebCite Cache ID 6Elnapi2K]
22. Vardavas C, Anagnostopoulos N, Kougias M, Evangelopoulou V, Connolly GN, Behrakis PK. Short-term pulmonary effects of using an electronic cigarette: impact on respiratory flow resistance, impedance, and exhaled nitric oxide. Chest 2012 Jun;141(6):1400-1406. [doi: 10.1378/chest.11-2443] [Medline: 22194587]
23. Vardavas CI, Anagnostopoulos N, Kougias M, Evangelopoulou V, Connolly GN, Behrakis PK. Short-term pulmonary effects of using an electronic cigarette: impact on respiratory flow resistance, impedance, and exhaled nitric oxide. Chest 2012 Jun;141(6):1400-1406. [doi: 10.1378/chest.11-2443] [Medline: 22194587]
24. Eysenbach G. Infodemiology: The epidemiology of (mis)information. Am J Med 2002 Dec 15;113(9):763-765. [Medline: 12517369]
25. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
26. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
27. Ayers JW, Althouse BM, Allem JP, Ford DE, Ribisl KM, Cohen JE. A novel evaluation of World No Tobacco day in Latin America. J Med Internet Res 2012;14(3):e77 [FREE Full text] [doi: 10.2196/jmir.2148] [Medline: 22634568]
28. Miura K, Kikukawa Y, Nakao T, Tokai H, Izumi Y, Fujii H, et al. Safety assessment of electronic cigarettes in smokers. Seikatsu Eisei 2011;55(1):59-64.
29. Polosa R, Caponnetto P, Morjaria JB, Papale G, Campagna D, Russo C. Effect of an electronic nicotine delivery device (e-Cigarette) on smoking reduction and cessation: a prospective 6-month pilot study. BMC Public Health 2011;11:786 [FREE Full text] [doi: 10.1186/1471-2458-11-786] [Medline: 21989407]
30. USDHHS. How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease. In: A Report of the Surgeon General. Atlanta: U.S. D.o.H.a.H Services; 2010.
31. Benowitz NL. Pharmacology of nicotine: addiction, smoking-induced disease, and therapeutics. Annu Rev Pharmacol Toxicol 2009;49:57-71 [FREE Full text] [doi: 10.1146/annurev.pharmtox.48.113006.094742] [Medline: 18834313]
32. Schep LJ, Slaughter RJ, Beasley DM. Nicotinic plant poisoning. Clin Toxicol (Phila) 2009 Sep;47(8):771-781. [doi: 10.1080/15563650903252186] [Medline: 19778187]
33. Hughes JR, Higgins ST, Bickel WK. Nicotine withdrawal versus other drug withdrawal syndromes: similarities and dissimilarities. Addiction 1994 Nov;89(11):1461-1470. [Medline: 7841857]
34. Eissenberg T. Electronic nicotine delivery devices: ineffective nicotine delivery and craving suppression after acute administration. Tob Control 2010 Feb;19(1):87-88 [FREE Full text] [doi: 10.1136/tc.2009.033498] [Medline: 20154061]
35. Grandjean P, Bellinger D, Bergman A, Cordier S, Davey-Smith G, Eskenazi B, et al. The faroes statement: human health effects of developmental exposure to chemicals in our environment. Basic Clin Pharmacol Toxicol 2008 Feb;102(2):73-75. [doi: 10.1111/j.1742-7843.2007.00114.x] [Medline: 18226057]
36. Talbot P, Lin S. Cigarette Smoke's Effect on Fertilization and Pre-implantation Development: Assessment Using Animal Models, Clinical Data, and Stem Cells. Journal of Biological Research 2011;44:189-194.

*Edited by G Eysenbach; submitted 25.08.12; peer-reviewed by J Brigham; comments to author 08.10.12; revised version received 14.11.12; accepted 05.02.13; published 08.04.13.*

*Please cite as:*
*Hua M, Alfi M, Talbot P*
*Health-Related Effects Reported by Electronic Cigarette Users in Online Forums*
*J Med Internet Res 2013;15(4):e59*
*URL: http://www.jmir.org/2013/4/e59/*
*doi:10.2196/jmir.2324*
*PMID:23567935*

Original Paper

# Computing Health Quality Measures Using Informatics for Integrating Biology and the Bedside

Jeffrey G Klann[1,2,3], MEng, PhD; Shawn N Murphy[1,2,3], MD, PhD

[1]Laboratory of Computer Science, Massachusetts General Hospital, Boston, MA, United States

[2]Research Computing, Partners Healthcare System, Inc., Boston, MA, United States

[3]Harvard Medical School, Boston, MA, United States

**Corresponding Author:**
Jeffrey G Klann, MEng, PhD
Laboratory of Computer Science
Massachusetts General Hospital
One Constitution Center
Boston, MA, 02129
United States
Phone: 1 6176435879
Fax: 1 6176435280
Email: jklann@partners.org

## Abstract

**Background:** The Health Quality Measures Format (HQMF) is a Health Level 7 (HL7) standard for expressing computable Clinical Quality Measures (CQMs). Creating tools to process HQMF queries in clinical databases will become increasingly important as the United States moves forward with its Health Information Technology Strategic Plan to Stages 2 and 3 of the Meaningful Use incentive program (MU2 and MU3). Informatics for Integrating Biology and the Bedside (i2b2) is one of the analytical databases used as part of the Office of the National Coordinator (ONC)'s Query Health platform to move toward this goal.

**Objective:** Our goal is to integrate i2b2 with the Query Health HQMF architecture, to prepare for other HQMF use-cases (such as MU2 and MU3), and to articulate the functional overlap between i2b2 and HQMF. Therefore, we analyze the structure of HQMF, and then we apply this understanding to HQMF computation on the i2b2 clinical analytical database platform. Specifically, we develop a translator between two query languages, HQMF and i2b2, so that the i2b2 platform can compute HQMF queries.

**Methods:** We use the HQMF structure of queries for aggregate reporting, which define clinical data elements and the temporal and logical relationships between them. We use the i2b2 XML format, which allows flexible querying of a complex clinical data repository in an easy-to-understand domain-specific language.

**Results:** The translator can represent nearly any i2b2-XML query as HQMF and execute in i2b2 nearly any HQMF query expressible in i2b2-XML. This translator is part of the freely available reference implementation of the QueryHealth initiative. We analyze limitations of the conversion and find it covers many, but not all, of the complex temporal and logical operators required by quality measures.

**Conclusions:** HQMF is an expressive language for defining quality measures, and it will be important to understand and implement for CQM computation, in both meaningful use and population health. However, its current form might allow complexity that is intractable for current database systems (both in terms of implementation and computation). Our translator, which supports the subset of HQMF currently expressible in i2b2-XML, may represent the beginnings of a practical compromise. It is being pilot-tested in two Query Health demonstration projects, and it can be further expanded to balance computational tractability with the advanced features needed by measure developers.

***(J Med Internet Res 2013;15(4):e75)*** doi:10.2196/jmir.2493

**KEYWORDS**

medical informatics; healthcare quality assessment; reimbursement; incentive; systems integration; database management systems

## Introduction

### Background and Significance

In 2010, the US Congress enacted a 10-year, $27 billion dollar incentive program to promote adoption and meaningful usage of electronic health record systems (EHRs) [1]. The government's vision is that this will lead to a "learning health system" in which health care data can be dynamically aggregated and analyzed for applications such as research, population health measurement, and disease surveillance [2]. The Meaningful Use (MU) program is being rolled out in three stages, each of which expands the definition of what is considered meaningful usage. Stage 1 of Meaningful Use (MU1) focused on capturing structured, coded data, but Stages 2 and 3 (MU2 and MU3) move toward the larger goal of a learning health system, through a focus on population health improvement enabled by Clinical Quality Measures (CQMs)[3].

MU1 involved a small number of computationally simple CQMs [4]. The final rules for MU2 [5,6] require between 9 and 24 CQMs from a menu of 93 [5], many of which are more complex than the MU1 measures, involving test results, deeper computation, and many more elements. The MU3 Request for Comment (RFC) discusses CQMs as the basis of population health management dashboards and introduces the idea that an EHR might need to respond to arbitrary (not pre-defined) CQMs [7]. The goal appears to be dynamic, distributed population health queries [8], enabled by standards and technologies from the Query Health initiative [9-11].

Query Health was convened by the Office of the National Coordinator for Health Information Technology (ONC) in 2011 to develop a standard approach for distributed population health queries. Query Health has defined a standards-based methodology using the Health Level 7 (HL7) Health Quality Measures Format (HQMF) for queries and the Quality Reporting Document Architecture (QRDA) for responses. Additionally, Query Health has developed a reference implementation using their selection of three best-of-breed technological components [9,10].

Query Health has three current pilots, which each use parts of this reference implementation [12]. Two of the pilots are in cooperation with a Department of Health (New York City and Massachusetts) for disease monitoring and surveillance. In the third, the FDA will use Query Health for medication safety surveillance as part of their Mini-Sentinel project. Mini-Sentinel has predominantly used administrative data, and the FDA is evaluating the increased utility of using clinical information.

Two of the pilots use the Informatics for Integrating Biology and the Bedside (i2b2) analytical platform for processing HQMF. i2b2 is a flexible, componentized clinical analytics and data warehousing platform that now enjoys widespread adoption as a research data repository, warehousing clinical data alongside the EHR at over 80 sites nationwide. It is part of an NIH-funded center charged with developing a national computational infrastructure for biomedical computing [13]. i2b2 has underpinned many studies that have discovered new knowledge about disease and its genetic bases (eg, [14]). As part of the Query Health Reference implementation, i2b2 now needs to support composition and consumption of HQMF.

HQMF is becoming the lingua franca for defining population health queries. A draft standard of HQMF was released by HL7 in 2010 [15] and was adopted by the National Quality Forum (NQF) and Query Health. In 2011, the NQF converted 113 of its CQMs to HQMF format (called eMeasures) [16], and now they, with the Centers for Medicare and Medicaid Services (CMS), have published the 93 CQMs used in MU2 as eMeasures [17]. The first HQMF draft standard expired in March 2012, and a second draft standard has been developed jointly by Query Health and HL7 and is in ballot at the time of this writing [18]. This release (HQMFr2) is vastly more computable, more readable, and less bulky. In addition to its usage by Query Health, the NQF is taking steps to re-tool their HQMFr1 eMeasures into HQMFr2. They are balloting an HQMF Implementation Guide based on their Quality Data Model (QDM) [19], which defines a comprehensive set of health care data elements and associated states and attributes (eg, medications can have dose and frequency). This HQMF QDM Implementation Guide will enforce a standard methodology for eMeasure re-tooling. Allscripts is planning an eMeasures interpreter for MU2, well in advance of any federal incentive for handling HQMF [12]. Nonetheless, the MU3 RFC does indicate that mandated electronic processing of HQMF is on the horizon [7].

Throughout the remainder of this manuscript, HQMF will refer to the revised HQMFr2 developed for the ballot now in progress, which is used by the Query Health pilots. Areas under active discussion in the ballot are noted.

### Objective

Our goal is to integrate i2b2 with the Query Health HQMF architecture, to prepare for other HQMF use cases (such as MU2 and MU3), and to articulate the functional overlap between i2b2 and HQMF. Therefore, we analyze the structure of HQMF, and then we apply this understanding to HQMF computation on the i2b2 clinical analytical database platform. Specifically, we develop a translator between two query languages, HQMF and i2b2, so that the i2b2 platform can compute HQMF queries.

## Methods

### Understanding HQMF

HQMF is a language for defining quality measures expressed as logical combinations of clinical variables, intended to return an aggregate count or percentage. For example, HQMF can express, in a computable manner, the following query: "the number of diabetes patients who have had a hemoglobin A1C test greater than 9% in the past year".

HQMF is derived from the HL7 v3 Reference Implementation Model (RIM), which is also the basis for Clinical Document Architecture (CDA) documents. Therefore CDA bears some resemblance to HQMF. Articulating all of the details of the HQMF format is beyond the scope of this article. Our purpose here is to highlight the structure of HQMF necessary to develop an implementation. For complete documentation, refer to the balloted HQMFr2 [18], the balloted HQMF QDM

Implementation Guide [19], and the Query Health HQMF Implementation Guide [20].

We present HQMF through two examples, both based on the NQF0059 measure to detect poorly controlled diabetes [21]. In the first example, we outline the structure of an HQMF query for basic measurement. The second example highlights additional complexities needed to support the more challenging nuances of HQMF. (Note that both examples are illustrative but not complete HQMF; for compact presentation, some XML elements are not included. These include HL7 XML bulk, HQMF headers, and some elements described in the text that would not fit in the figure.)

### *Example 1: Basic HQMF*

The anatomy of a basic HQMF query is shown in Figure 1. This simplification of NQF0059 defines the following query: "Find all patients seen in the past year between 18-75 who have, during that year, had either a diabetes diagnosis or have taken a diabetes medication, have had an abnormally high HgbA1c test result, and have not been documented to have polycystic ovaries or steroid induced diabetes." Although this is a simplification, it is a meaningful quality measure. Figure 1 shows the three basic components of an HQMF document: Measure Period, Data Criteria, and Population Criteria.

**Figure 1.** The anatomy of a basic HQMF query (shown here: a portion of a simplified NQF0059 diabetes measure in HQMF).

**Measure Period**

```
<measurePeriod>
    <value>
        <width value="1" unit="a"/>
    </value>
</measurePeriod>
```

**Data Criteria**

```
<observationCriteria>
    <id extension="18-75_years_old"/>
    <code codeSystem="2.16.840.1.113883.6.96" code="424144002"/>
    <value>
        <low unit="a" value="18"/>
        <high unit="a" value="75"/>
    </value>
</observationCriteria>

<observationCriteria>
    <id extension="Diabetes_diagnosis"/>
    <value codeSystem="2.16.840.1.113883.6.103" code="250" />
</observationCriteria>

<observationCriteria>
    <id extension="HGB_A1C_GE_9_%" />
    <code codeSystem="2.16.840.1.113883.6.1" code="4548-4"/>
    <value>
        <low unit="%" value="9"/>
    </value>
</observationCriteria>

… (the three remaining data criteria are similar) …
```

**Population Criteria**

```
<patientPopulationCriteria>
    <allTrue>
        <observationReference>
            <id extension="18-75_years_old"/>
        </observationReference>
        <observationReference>
            <id extension="HGB_A1C_GE_9_%"/>
        </observationReference>

        <atLeastOneTrue>
            <observationReference>
                <id extension="Diabetes_diagnosis"/>
            </observationReference>
            <observationReference>
                <id extension="Diabetes_medication"/>
            </observationReference>
        </atLeastOneTrue>

        <allFalse>
            <observationReference>
                <id extension="Steroid_induced_diabetes"/>
            </observationReference>
            <observationReference>
                <id extension="Polycystic_ovaries"/>
            </observationReference>
        </allFalse>

    </allTrue>
</patientPopulationCriteria>
```

### Measure Period

Measure Period defines the time period the overall query covers, which in this case is 1 year (value *1*, unit *a*). Individual criteria can further restrict the time search.

### Data Criteria

The Data Criteria section defines the clinical variables in the query. Each variable is a *criteria* entry, the types of which correlate with the *act* classes in CDA. Three data criteria are shown in Figure 1. Here, all criteria are observation criteria, but criteria also exist for encounters, procedures, medications (supplied and administered), and general acts. Each criterion most commonly defines a code, value, and an id. *id* is used to reference the criterion in the population criteria section via the *extension* attribute. *code* defines what is being measured, and *value* defines its result. Note that what is a code and what is a value is not always intuitive. For example, the diagnosis of diabetes is a value (where the optional code refers to the "problem type", eg, diagnosis, complaint, etc.), but the HgbA1c test is a code and its result is a value.

*Code* can be either a coded value or value set. A coded value is a numeric code and a code system (eg, SNOMED, LOINC, ICD-9) represented by an HL7-registered Object Identifier (OID). A value set is a set of coded values referred to by a single OID. The coded values are OR'd together and are convenient shorthand when defining measures where an observation like "diabetes" could be recorded as multiple specific codes (eg, various ICD-9, ICD-10, or SNOMED codes). At the time of this writing, the National Library of Medicine has begun hosting a small number of value sets at their Value Set Authority Center (VSAC) repository [22,23].

*Value* can be any HL7 data type. When it is not a coded value, it is frequently one of: (1) a numeric value, when a specific value is desired, such as an HgbA1c of exactly 9%; or (2) an interval, seen in the age range and HgbA1c test in Figure 1.

### Population Criteria

The Population Criteria section defines the calculations on these data elements needed to compute the quality measure. Figure 1 defines one population through a series of nested lists. Each list contains references to data criteria (through the *extension* attribute) as well as other lists. Each list begins with one of six logical "combining operators", three of which are shown in the figure. The others are *atLeastOneFalse*, *onlyOneTrue*, and *onlyOneFalse*.

## *Example 2: Advanced HQMF*

Example 1 presented a pedagogical simplification of the NQF0059 measure. In this second example, we introduce additional features present in the actual measure. The updated query (with changes italicized) is: "Find all the patients who: *were seen at least twice within the two years before the end of the measure period*; are between 18 and 75; and within the measure period have had either a diabetes diagnosis or have taken a diabetes medication, had an abnormally high value for *their most recent* HgbA1c test result, and are not documented to have polycystic ovaries or steroid induced diabetes." Note that this still is not quite as complex as the actual measure, but it does use all of the HQMF features present in that measure. HQMF corresponding to the two modified data criteria and the new population criteria section can be seen in Figure 2. This uses the following additional HQMF features.

### Value Sets

The *code* elements in the data criteria now reference value set OIDs, rather than defining a coded value. These are published by the NQF, and some are available for download from the NLM's VSAC repository [23].

### Excerpting

Excerpting applies a filter to a data criterion and reports a summary value of the filtered results (eg, first, last, largest, smallest). This can been seen in the first data criterion in Figure 2. It is reminiscent of the HgbA1c criterion from Figure 1, but it now selects only the most recent HgbA1c test result with a value greater than 9%. This is done by wrapping the *value* element with an *excerpt* element with a subset code of "RECENT". This means, "find all HgbA1c test results within the measure period, filter out any results not greater than 9%, and report the most recent."

### Counting Repetitions

The second data criterion in Figure 2 defines ambulatory encounters using an NQF value set. The *repeatNumber* element specifies that at least two ambulatory encounters must have occurred.

#### *Temporal Relationships*

Also in the second data criterion in Figure 2, the *temporallyRelatedInformation* element specifies that these encounters must have occurred within 2 years (*pauseQuantity*) of the end (*typecode=SBE*—"starts before end") of the measure period (*observationReference*). Multiple temporal relationships can occur within a single criterion, and 17 types of relationships are defined. In addition to the measure period, temporal relationships can reference other data criteria.

#### *Multiple Population Criteria*

Figure 2 has four population criteria: an initial patient population (all people between 18 and 75), a denominator (those with diabetes), a numerator (those with abnormal HgbA1c values), and exceptions (those with polycystic ovaries). This multipopulation approach has two purposes. One, it allows measurement results to be reported as a percentage: numerator divided by denominator. Two, smaller population components are more modular. This has an organizational advantage for measure developers, but it also could speed computation. If multiple queries with the same denominator and exceptions are run with different numerators, the denominator can be computed just once.

## *Completing an HQMF Implementation*

These two examples cover all the major features in NQF0059. However, to support very complex measures, HQMF implementations should also include the following features.

**Figure 2.** The anatomy of a full-featured HQMF query (shown here: a portion of a simplified NQF0059 diabetes measure in HQMF, highlighting advanced computational features of HQMF).

**Data Criteria**

```
<observationCriteria>
    <id extension="HbA1Cgt9"/>
    <code valueSet="2.16.840.1.113883.3.464.1.72"/>
    <excerpt>
        <subsetCode code="RECENT"/>
        <observationCriteria>
            <value>
                <low value="9" unit="%"/>
            </value>
        </observationCriteria>
    </excerpt>
</observationCriteria>

<encounterCriteria>
    <id extension="AmbulatoryEncounter"/>
    <code valueSet="2.16.840.1.113883.3.464.1.1142"/>
    <repeatNumber>
        <low value="2"/>
    </repeatNumber>
    <temporallyRelatedInformation typeCode="SBE">
        <pauseQuantity value="-2" unit="a"/>
        <observationReference>
            <id extension="MeasurePeriod"/>
        </observationReference>
    </temporallyRelatedInformation>
</encounterCriteria>
```

**Population Criteria**

```
<patientPopulationCriteria>
  <allTrue>
    <observationReference>
      <id extension="ageBetween18and75"/>
    </observationReference>
  </allTrue>
</patientPopulationCriteria>

<denominatorCriteria>
  <allTrue>
    <observationReference>
      <id extension="HasDiabetes"/>
    </observationReference>
    <encounterReference>
      <id extension="AmbulatoryEncounter"/>
    </encounterReference>
  </allTrue>
</denominatorCriteria>

<numeratorCriteria>
  <allTrue>
    <observationReference>
      <id extension="HbA1Cgt9"/>
    </observationReference>
  </allTrue>
</numeratorCriteria>

<denominatorExceptionCriteria>
  <atLeastOneTrue>
    <observationReference>
      <id extension="HasPolycysticOvaries"/>
    </observationReference>
  </atLeastOneTrue>
</denominatorExceptionCriteria>
```

### Stratifiers

Normally, results are reported as one aggregate number, but stratifiers allow reporting to be broken down by other criteria (eg, zipcode or gender).

### Denominator Exclusions

An additional population group, this describes criteria for patients who should be excluded from the denominator only if they do not meet numerator criteria (for example, the measure could be modified to not penalize practices for uncontrolled diabetics who are high risk, eg, who have been to the emergency room at least five times in the past year).

### Calculations on Continuous Variables

This might include, for example, the average emergency department wait time.

### Modifiers

Data criteria can have other modifiers beyond those discussed here. Some of these are simply data elements, eg, the clinician interpretation code for a vital sign goes in an *interpretationCode* element. Others, such as details of medication route, admitting physician, or problem status require more complex XML structure using RIM data elements. Implementation Guides (IGs) will define how such elements should be expressed. The only available resources at the time of this writing are the Query Health IG draft and standards developer Keith Boone's blog

[20,24]. (The QDM IG leaves criteria-level templates to a future release.) Some modifier structure can be inferred from CDA IGs, though portability across implementations cannot be guaranteed until a complete set of templates in an HQMF IG are available.

#### Missing Information

Values can explicitly be "no information", allowing special behavior when information is missing.

#### Specific Occurrences

Under active discussion is a canonical representation for multiple observations that must be temporally related to a specific occurrence of another observation (eg, multiple observations referring to a single encounter).

### Understanding i2b2

i2b2 consists of a flexible relational data model and a somewhat more restricted set of Web services.

The data model stores observations as "facts" associated with a date, patient identifier, and encounter identifier. Each fact has a key that follows a hierarchical structure (eg, "Diagnosis\Diabetes Mellitus\Diabetes With Ketoacidosis") and a basecode that defines the source code and coding system (eg, "ICD9:250.6"). Optional modifier facts can supplement each primary fact (eg, "admit diagnosis"). Facts can have a value, stored in numeric or text formats. The set of possible facts (the ontology) are user-defined, though several standard ontologies exist.

The web services define an easy-to-understand query language expressed in an XML format. Queries are built in a web-based query tool and executed as highly optimized SQL statements by a "data repository" web service. A query consists of items dragged into a set of panels. The items in each panel are logically OR'd together, and all panels are logically AND'd. The NOT operation can optionally be applied to a panel. Each item can also have date constraints, though the only other temporal constraint currently supported is "all items in panel must be in the same encounter." Results can be stratified on any observation (eg, age, race, gender), and queries can be combined through query-within-a-query (eg, a previous query definition can be dragged into a panel), which allows deep nesting and more complex combinations of elements. Furthermore, ontology items can have arbitrary SQL statements embedded, for added flexibility.

This query XML is logically equivalent to a subset of HQMF, and it is therefore possible to translate between these two formats, provided that queries are restricted to this subset.

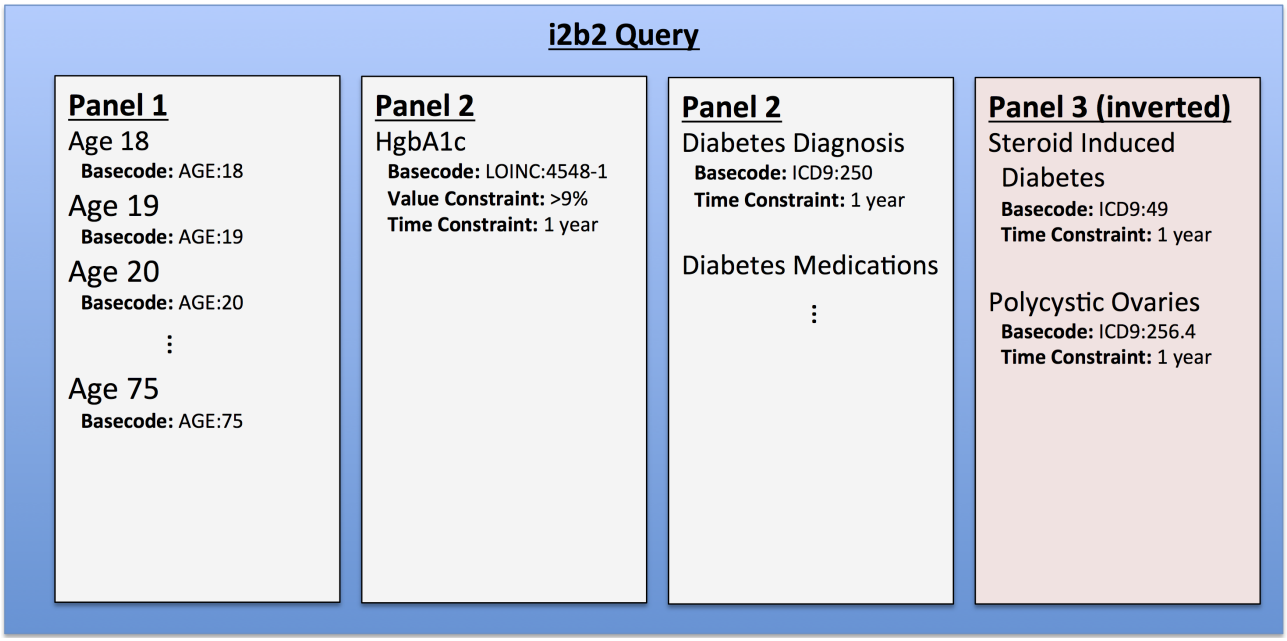
## Results

We first illustrate our translation method between i2b2-XML and HQMF by describing the conversion of Example 1 to an i2b2 query. Next, we analyze the HQMF subset supported by i2b2-XML, and we describe a web service we have implemented for i2b2 to both generate and consume HQMF conforming to this subset.

### Example 1 As i2b2

Figure 3 shows an i2b2 query version of Figure 1, displayed as i2b2 panels. The conversion of each element from HQMF to i2b2-XML is as follows.

**Figure 3.** The HQMF from Figure 1 converted to i2b2 query panels.




#### *Data Criteria Conversion*

First, each data criterion is converted to an i2b2 basecode. For age ranges (panel 1), age is expanded to the OR of every age in the range using the AGE basecode. (There is no standard terminology to encode age ranges; this is the only case of specialized logic for a particular data element.) For all other values, an OID-to-i2b2 lookup table determines the basecode, which is implementation-configurable.

Second, i2b2 basecodes are sent to the i2b2 ontology cell for conversion to associated i2b2 keys. An i2b2 key represents a

position in the ontology tree, so the ICD diagnosis codes in panels 2 and 3 include not only the listed basecode (eg, ICD9:250) but also all subcodes (eg, ICD9:250.xx). This nonstandard HQMF interpretation allowed us to specify an implicit diabetes value set with single coded value, prior to standardization of NQF value sets. Therefore, we feel this behavior (which will normally affect only ICD codes) is useful at the present time, and we will revisit it once the NLM VSAC is more complete.

Third, value constraints are added, such as the numeric interval ">9%" for HgbA1c.

Fourth, time constraints are added. When no specific time constraint is specified, a time constraint is added to match the measurement period.

### *Population Criteria Conversion*

For *allTrue*, every item or list following is placed in a separate panel. For *atLeastOneTrue*, every item following is placed in the same panel. For *allFalse*, every item following is placed in the same panel, which is set to the *exclude* type (ie, the NOT operation is performed on each item in the panel).

## HQMF i2b2 Translator

HQMF features currently supported by i2b2-XML are shown in Table 1. For the first version of the translator, we targeted a subset of these features that cover the functionality of the Shared Health Research Information Network (SHRINE), a distributed version of i2b2 in use around the country for performing distributed population queries [25].

We have implemented a bidirectional translator that can convert between HQMF-coded values and i2b2 keys (see Figure 4). For translation from HQMF, the basecode lookup method in the previous example is used. For translation to HQMF, a reverse lookup retrieves the basecode from the key. For consistency, all child nodes of the key are by default included (eg, \\I2B2\Demographics\Zipcodes\Alabama is augmented by an item for every zipcode in Alabama). The translator further supports result values, date constraints, repetition counting (at the panel level), and a single population with multiple *atLeastOneTrue* and *allFalse* groups inside an *allTrue* group.

Stratifiers, excerpts, and features requiring query-within-a-query are not currently supported. Stratifiers were a low priority for the Query Health pilots, so they have been tabled. Query-within-a-query features are waiting on the more portable implementation in i2b2-XML planned for i2b2's next release. Excerpts will be supported in the future through ontology definitions containing custom SQL code.

Modifiers and value sets are partially supported, but they also require an appropriate HQMF i2b2 ontology (containing approved value sets and supported modifiers). Dr. Michael Buck in the New York Department of Health is developing this ontology, which is an expression of CDA and QDM data elements in i2b2. We have tested his ontology's ability to support HQMF modifiers and value sets by successfully implementing support for health care providers and clinician interpretation codes and by testing conversion of his NQF0059-compatible diabetes value set. When completed, this ontology will significantly aid i2b2 in consuming complex CQMs.

To validate the translator's accuracy, we developed four query libraries in i2b2-XML, one for each supported ontology configuration of the translator (SHRINE, i2b2 default, Dr. Buck's terminology, and Beth Israel Deaconess Medical Center). Each library utilizes all features of the translator (see Table 1 and the discussion above). As we developed the translator, we (the authors or a member of the Query Health Reference Implementation team) periodically translated each library to HQMF and back to i2b2-XML. We visually inspected the HQMF and i2b2-XML to verify they were semantically equivalent to the original (because the translator inserts basecodes for child nodes, syntactic equivalence might not occur). We also had access to databases of (fake) test patients for all but the Beth Israel configuration, so we also ran both the original and post-translation i2b2-XML against that and verified the result was the same. Finally, we periodically translated and validated an HQMFr2 version of NQF0059 available in Query Health's repository. No other publicly available source of HQMFr2 presently exists (though this will rapidly change once HQMFr2 is through ballot).

Our translator supports a superset of SHRINE's features, and it is able to generate and process HQMF corresponding to the i2b2-XML feature set in Table 1 except as noted above. It can consume HQMF from other sources, as long as it is restricted to this feature set. The transformation process is summarized visually in Figure 5, with an illustration of i2b2 panels being built from the population criteria. When a criteria reference is encountered, an i2b2 item is inserted using information from the data criteria element. Supported HQMF elements include *value*, *code*, temporal constraints on the measure period, and some modifiers and value sets. The reverse translation follows the diagram in reverse (data criteria are generated from i2b2 items, and population criteria are built from the panel layout).

The translator is by default configured for diagnoses, labs, procedures, medication administration events, and the demographics in SHRINE (Age, Gender, Language, Marital Status, and Race and Ethnicity). Non-SHRINE ontologies are supported through a configuration file. The translator is open source and presently available from the Query Health repository [26]. It will likely be included in a future version of i2b2. The translator is implemented in Java and XSL as a Jersey servlet that runs within the i2b2 JBoss stack. A service to connect i2b2 to a Query Health-conformant, HQMF-based distributed query engine is being completed at the time of this writing and will also be available as an open-source project.

**Table 1.** HQMF features and their equivalence in the i2b2 XML query language (features not fully equivalent are italicized; note that some supported features are ontology-dependent).

| | HQMF | i2b2-XML |
|---|---|---|
| **Data Criteria** | | |
| | Coded values, with optional result value | Facts (looked up by basecode) |
| | Value sets, with optional result value | Facts (requires an ontology with value set OIDs as basecodes) |
| | Modifiers | Facts (requires an appropriate ontology) |
| | Excerpts | Facts (requires ontology definitions) |
| | Repetition counting | *Supported at the panel level.* |
| | Temporal relationships | *Date constraints (such as the measure period) only; others to be supported in summer 2013.* |
| **Population Criteria** | | |
| | Multiple populations | *Populations are constructed separately and can be combined with query-within-query.* |
| | Calculations on continuous variables | *Supported conceptually through patient-data objects and client-side analysis, but a plugin would be needed to perform HQMF server-side calculations.* |
| | allTrue | The ANDing of all panels. |
| | atLeastOneTrue | The ORing of elements in a panel. |
| | allFalse | An inverted panel |
| | atLeastOneFalse | A set of inverted panels, each consisting of a previous query. |
| | onlyOneTrue | Not supported |
| | onlyOneFalse | Not supported |
| | Nested “combining operators” | atLeastOneTrue and allFalse groups inside an allTrue |
| | Stratifiers | Supported (but requires server configuration). |

**Figure 4.** Translation from an HQMF coded value to i2b2 item key(s).

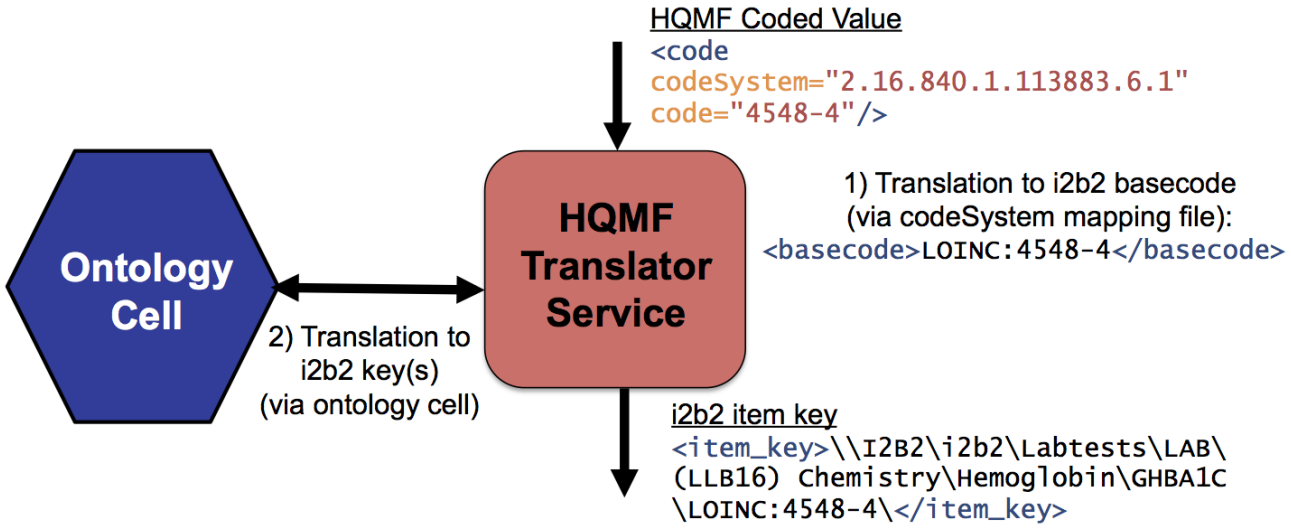
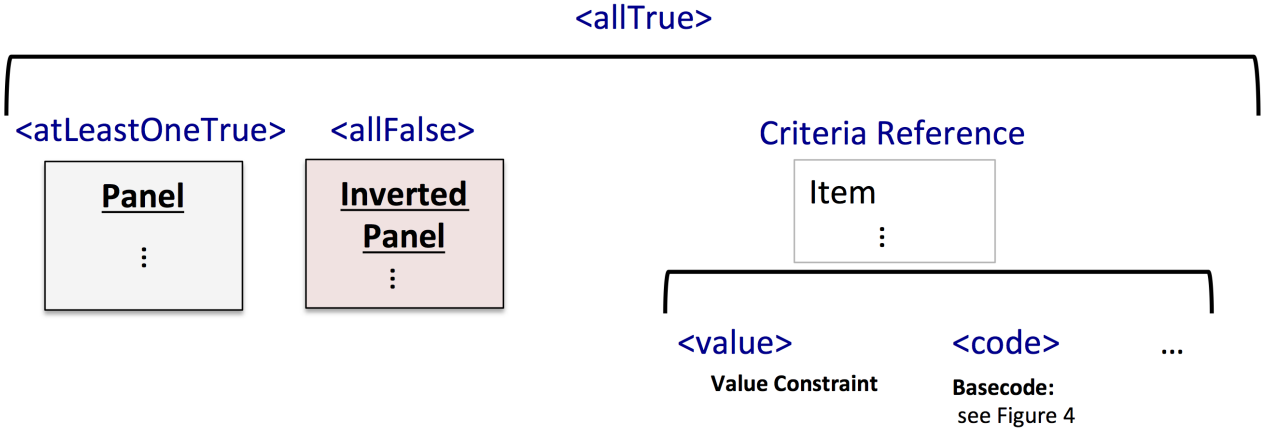


**Figure 5.** The transformation process from HQMF to i2b2-XML.




## Discussion

We have achieved significant interoperability between HQMF and i2b2 using only transformations of the two XML languages. Our web service successfully transforms queries bidirectionally between i2b2-XML and HQMF, supporting a superset of all queries possible in SHRINE. This HQMF translator has been integrated into the reference implementation of Query Health and is currently being used at two Query Health pilot sites. The first, through the New York State Department of Health, utilizes the bidirectional translator with Dr. Buck's ontology, allowing researchers to create custom HQMF queries and execute them at i2b2 sites throughout New York. The second pilot, at the Beth Israel Deaconess Medical Center, connects their i2b2 instance (using the SHRINE ontology) to the FDA Mini-Sentinel network and demonstrates the FDA's ability to send custom HQMF queries to i2b2 instances. i2b2 1.7, to be released in summer 2013, will extend the i2b2-XML language to support a larger subset of HQMF, including improved query-within-query and new temporal constraints. We hope this, along with Dr. Buck's ontology, will give us the tools needed to compute CQMs for MU2 and MU3 using the XML translation approach.

HQMF is an elegant but complex and challenging query language. It separates the definition of data elements from the logical operations used to combine them, and the format for clinical variables is similar to CDA, making basic understanding less difficult for HL7 developers. However, even the partial support we have achieved has been a massive effort. The work involved was more than 6 months of FTE time by an experienced software architect, and we had a preexisting sample transform from standards developer Keith Boone, support from Query Health team members, and the i2b2 analytical database system with a schema-defined XML language. Even with i2b2 1.7 and Dr. Buck's ontology, full translator support will involve much additional effort. Although future implementers will have both the advantage of our experiences and an open-source, HQMF-enabled version of i2b2, software development groups should carefully consider the required effort. Furthermore, the current HQMF specification allows some query constructions that are computationally challenging. Multiple time relationships on a single criterion, the possibility of nesting time relationships and excerpts, arbitrarily deep nesting of population criteria groups, and the fact that not all population criteria operators are equivalent to logical operators suggests to us that running very complex queries could tax even powerful servers. Also, HQMF allows unrealistic query constructions. In particular, supporting 16 temporal relationships between criteria does not reflect the limited temporal granularity we have seen in existing clinical data warehouses. Finally, HQMF has limited ability to specify behavior when data are missing or noisy. Revisions to implementation guides and the HQMF standard will be needed to address these issues. Some of these difficulties (eg, reducing complexity by limiting nesting in data criteria) are being discussed in the current ballot reconciliations.

### Conclusion

HQMF is a powerful language for developing and computing measures of population health. Despite complexity concerns, we believe this format can be supported, which will be important in stage two of meaningful use and possibly required in stage three. HQMF is also a key component in Query Health and will likely play important roles in other health informatics initiatives. We have created a fully bidirectional HQMF translator, which converts between HQMF and i2b2-XML formats, supporting a superset of SHRINE's features. This translator is freely available and has been integrated into the reference implementation of Query Health. Areas that we have been unable to implement will either be addressed in the next release of i2b2 or will be brought before appropriate decision-makers and standards developers. Expanding i2b2's HQMF support through this XML translation approach will allow i2b2 to fully support the requirements of CQM processing in MU. The work lays a foundation for dynamic, distributed queries across diverse clinical systems with disparate data models.

### Acknowledgments

Many thanks to the Query Health Reference Implementation team for making this work possible. In particular, we thank: Keith Boone at GE Healthcare, whose preliminary work on an i2b2 to HQMF transform paved the way for the translator described in

this paper; Marc Hadley at MITRE, whose deep analysis of HQMF's computability has led to changes in the balloted standard and a variety of helpful conversations; Michael Buck in the New York State Department of Health, whose work on a comprehensive i2b2 Query Health ontology was both synergistic and inspirational with regard to this work; and Nageshwara Bashyam, who has been active in making sure Query Health has had an important voice in HQMFr2, has created examples, and has worked on translators for other platforms.

This work was sponsored by the Office of the National Coordinator for Health Information Technology, Grant No. ONC 90TR001/01.

## Conflicts of Interest

None declared.

## References

1. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med 2010 Aug 5;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]
2. Office of the National Coordinator for Health IT. 2011 Sep. Federal health information technology strategic plan 2011-2015 URL: http://www.healthit.gov/sites/default/files/utility/final-federal-health-it-strategic-plan-0911.pdf [accessed 2013-01-23] [WebCite Cache ID 6DtXdhWYW]
3. Blumenthal D. Launching HITECH. N Engl J Med 2010 Feb 4;362(5):382-385. [doi: 10.1056/NEJMp0912825] [Medline: 20042745]
4. Marcotte L, Seidman J, Trudel K, Berwick DM, Blumenthal D, Mostashari F, et al. Achieving meaningful use of health information technology: a guide for physicians to the EHR incentive programs. Arch Intern Med 2012 May 14;172(9):731-736. [doi: 10.1001/archinternmed.2012.872] [Medline: 22782203]
5. 77 Federal Register 53967. 2012 Sep 04. Medicare and Medicaid electronic health record incentive program - stage 2 (CMS-0044-P) URL: https://www.federalregister.gov/articles/2012/09/04/2012-21050/medicare-and-medicaid-programs-electronic-health-record-incentive-program-stage-2 [accessed 2013-03-17] [WebCite Cache ID 6FF7RPADp]
6. 77 Federal Register 54163. 2012 Sep 04. Health information technology implementation specifications: standards and certification criteria for electronic health record technology 2014 edition URL: https://www.federalregister.gov/articles/2012/09/04/2012-20982/health-information-technology-standards-implementation-specifications-and-certification-criteria-for [accessed 2013-03-14] [WebCite Cache ID 6F9BzKsVD]
7. Health IT Policy Committee. 77 Federal Register 70444. 2012 Nov 26. Request for comment regarding the stage 3 definition of meaningful use of electronic health records URL: http://www.regulations.gov/#!documentDetail;D=HHS-OS-2012-0007-0001 [accessed 2013-03-19] [WebCite Cache ID 6FFdDIYTT]
8. Diamond CC, Mostashari F, Shirky C. Collecting and sharing data for population health: a new paradigm. Health Aff (Millwood) 2009 Mar;28(2):454-466 [FREE Full text] [doi: 10.1377/hlthaff.28.2.454] [Medline: 19276005]
9. Elmore R. Health IT Policy Committee and Health IT Standards Committee. 2012 Jun 07. Quality measures workgroup clinical quality public hearing URL: http://www.healthit.gov/sites/default/files/elmore_testimony_hitpc_hitsc_060712_0.pdf [accessed 2012-12-13] [WebCite Cache ID 6Ct72sBNn]
10. Elmore R. It Takes A Community. 2012 Jul 12. Toward a learning health system URL: http://blog.allscripts.com/2012/07/12/toward-a-learning-health-system/ [accessed 2012-12-13] [WebCite Cache ID 6Ct6xj4Gm]
11. Mostashari F. Twitter. 2012 Sep 19. (Farzad_ONC) @Fridsma: top 3 #SIframework (direct,lab,TOC) in stage 2: Will #queryhealth / Data segmentation make into stage3? URL: https://twitter.com/Farzad_ONC/statuses/248484298116722688 [accessed 2013-03-14] [WebCite Cache ID 6F9CIAskG]
12. Query Health Pilots Team. Standards and Interoperability Framework. 2012 Jun 22. Query Health pilots synopsis URL: http://wiki.siframework.org/file/view/Query+Health+Pilots+Synopsis_v0+1.docx [accessed 2012-12-13] [WebCite Cache ID 6Ct7T0JhE]
13. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med Inform Assoc 2010;17(2):124-130 [FREE Full text] [doi: 10.1136/jamia.2009.000893] [Medline: 20190053]
14. Zeng QT, Goryachev S, Weiss S, Sordo M, Murphy SN, Lazarus R. Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system. BMC Med Inform Decis Mak 2006;6:30 [FREE Full text] [doi: 10.1186/1472-6947-6-30] [Medline: 16872495]
15. Health Level Seven (HL7). 2010 Feb. Representation of the health quality measures format (eMeasure), release 1 URL: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=97 [accessed 2013-03-15] [WebCite Cache ID 6F9COuxne]
16. National Quality Forum (NQF). NQF Report to Congress. 2012 Mar. Changing healthcare by the numbers URL: http://www.qualityforum.org/News_And_Resources/Report_to_Congress/Report_to_Congress.aspx [accessed 2012-12-13] [WebCite Cache ID 6Ct6qAn7e]

17. Centers for Medicare and Medicaid Services. 2012 Dec 06. Electronic specifications for clinical quality measures URL: http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Electronic_Reporting_Spec.html [accessed 2012-12-13] [WebCite Cache ID 6CtD7eLHS]
18. Health Level Seven (HL7). 2012 Sep. Representation of the health quality measures format (eMeasure), release 2 URL: http://www.hl7.org/special/committees/projman/searchableprojectindex.cfm?action=edit&ProjectNumber=508 [accessed 2013-03-19] [WebCite Cache ID 6FFA7yfgq]
19. Health Level Seven (HL7). 2012 Sep. Quality data model based health quality measures format (eMeasure) implementation guide, release 1 (US Realm) - based on HL7 HQMF release 2.0 URL: http://wiki.siframework.org/file/detail/HQMFR2_IG_QDM_RI_SEP.docx [accessed 2013-03-15] [WebCite Cache ID 6F9CkL4uo]
20. Office of the National Coordinator for Health Information Technology. 2012 May. Query Format (HQMF) implementation guidance URL: http://wiki.siframework.org/file/view/HQMF+NEXT+Implementation+Guidance+v0.5.docx [accessed 2012-12-13] [WebCite Cache ID 6Ct7dwUZ1]
21. National Quality Forum (NQF). National Voluntary Consensus Standards for Ambulatory Care - Part I. 2009 Aug 10. Diabetes: hemaglobin A1c poor control URL: http://www.qualityforum.org/Publications/2008/03/National_Voluntary_Consensus_Standards_for_Ambulatory_Care%E2%80%93Part_1.aspx [accessed 2012-12-13] [WebCite Cache ID 6Ct8Hj7ee]
22. National Quality Forum (NQF). 2012 Jul 13. NQF Health Information Technology Advisory Committee (HITAC) meeting summary URL: http://www.qualityforum.org/About_NQF/HITAC/Meeting_Summary_07132012.aspx [accessed 2012-12-13] [WebCite Cache ID 6Ct8auOfX]
23. US National Library of Medicine. Value set authority center URL: https://vsac.nlm.nih.gov/ [accessed 2012-12-13] [WebCite Cache ID 6CtDcJtdZ]
24. Boone K. Healthcare Standards. Commentary and education on current and emerging healthcare standards: HQMF URL: http://motorcycleguy.blogspot.com/search/label/HQMF [accessed 2012-12-13] [WebCite Cache ID 6Ct9wRwSU]
25. Weber GM, Murphy SN, McMurry AJ, Macfadden D, Nigrin DJ, Churchill S, et al. The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories. J Am Med Inform Assoc 2009;16(5):624-630 [FREE Full text] [doi: 10.1197/jamia.M3191] [Medline: 19567788]
26. Standards and Interoperability Framework. Query Health reference implementation URL: http://code.google.com/p/query-health/source/browse [accessed 2012-12-13] [WebCite Cache ID 6CtA5QQBq]

## Abbreviations

**CDA:** HL7's Clinical Document Architecture
**CMS:** Centers for Medicare and Medicaid Services
**CQM:** clinical quality measure
**EHR:** electronic health record
**FTE:** full-time employee
**HgbA1c:** Hemoglobin A1c, a blood sugar test
**HL7:** Health Level Seven, a health care standards organization
**HQMF(r2):** Health Quality Measures Format (revision 2)
**i2b2:** Informatics for Integrating Biology and the Bedside
**ICD9:** International Classification of Diseases, 9th Edition
**IG:** implementation guide
**LOINC:** Logical Observation Identifiers Names and Codes
**MU:** Meaningful Use Incentive Program, where MU1, MU2, and MU3 refer to the specific stages of the program.
**NLM:** National Library of Medicine
**NQF:** National Quality Forum
**NQF0059:** NQF CQM 0059, defining poorly controlled diabetes
**ONC:** Office of the National Coordinator for Health Information Technology
**QDM:** NQF's Quality Data Model
**RFC:** request for comment
**RIM:** HL7's reference implementation model
**SHRINE:** Shared Health Research Information Network
**SNOMED:** Systematized Nomenclature of Medicine
**SQL:** Structured Query Language
**VSAC:** NLM's Value Set Authority Center
**XML:** extensible markup language

*Edited by G Eysenbach; submitted 19.12.12; peer-reviewed by R Elmore, R Gamache, J Halamka; comments to author 22.01.13; revised version received 28.01.13; accepted 22.02.13; published 19.04.13.*

*Please cite as:*
*Klann JG, Murphy SN*
*Computing Health Quality Measures Using Informatics for Integrating Biology and the Bedside*
*J Med Internet Res 2013;15(4):e75*
*URL: http://www.jmir.org/2013/4/e75/*
*doi:10.2196/jmir.2493*
*PMID:23603227*

Original Paper

# Using Social Networking to Understand Social Networks: Analysis of a Mobile Phone Closed User Group Used by a Ghanaian Health Team

Nadi Nina Kaonga[1,2*], BA, MHS; Alain Labrique[1*], MS, MHS, PhD; Patricia Mechael[2,3*], MHS, PhD; Eric Akosah[4], B.Sc; Seth Ohemeng-Dapaah[4], PhD; Joseph Sakyi Baah[4], MSc; Richmond Kodie[4*], BA; Andrew S Kanter[2], MPH, MD, FACMI; Orin Levine[1*], PhD

[1]Global Disease Epidemiology and Control, International Health Department, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[2]Earth Institute, Center on Globalization and Sustainable Development, Columbia University, New York, NY, United States

[3]mHealth Alliance, United Nations Foundation, Washington, DC, United States

[4]Bonsaaso, Ghana, Millennium Villages Project, Manso-Nkwanta, Ghana

[*]these authors contributed equally

**Corresponding Author:**
Nadi Nina Kaonga, BA, MHS
Earth Institute
Center on Globalization and Sustainable Development
Columbia University
Suite 401
475 Riverside Drive
New York, NY,
United States
Phone: 1 6039694883
Email: nkaonga@gmail.com

## Abstract

**Background:** The network structure of an organization influences how well or poorly an organization communicates and manages its resources. In the Millennium Villages Project site in Bonsaaso, Ghana, a mobile phone closed user group has been introduced for use by the Bonsaaso Millennium Villages Project Health Team and other key individuals. No assessment on the benefits or barriers of the use of the closed user group had been carried out.

**Objective:** The purpose of this research was to make the case for the use of social network analysis methods to be applied in health systems research—specifically related to mobile health.

**Methods:** This study used mobile phone voice records of, conducted interviews with, and reviewed call journals kept by a mobile phone closed user group consisting of the Bonsaaso Millennium Villages Project Health Team. Social network analysis methodology complemented by a qualitative component was used. Monthly voice data of the closed user group from Airtel Bharti Ghana were analyzed using UCINET and visual depictions of the network were created using NetDraw. Interviews and call journals kept by informants were analyzed using NVivo.

**Results:** The methodology was successful in helping identify effective organizational structure. Members of the Health Management Team were the more central players in the network, rather than the Community Health Nurses (who might have been expected to be central).

**Conclusions:** Social network analysis methodology can be used to determine the most productive structure for an organization or team, identify gaps in communication, identify key actors with greatest influence, and more. In conclusion, this methodology can be a useful analytical tool, especially in the context of mobile health, health services, and operational and managerial research.

***(J Med Internet Res 2013;15(4):e74)*** doi:10.2196/jmir.2332

**KEYWORDS**

mobile health; electronic health; telehealth; sociology; social network analysis; rural health; global health; evaluation research; Ghana

## Introduction

The Millennium Villages Project (MVP) is a collaborative initiative that seeks to support and empower rural African communities to achieve the Millennium Development Goals by 2015. Initiatives are centered on five sectors: agriculture, health, education, enterprise, and infrastructure. Within the health sector, science and technologies are being integrated into health systems and are evaluated in close partnership with the communities and other key stakeholders (ie, Ministries of Health) in each of the 14 MVP sites.

In the MVP site in Bonsaaso, Ghana, a mobile phone closed user group (CUG) has been introduced for use by the Bonsaaso MVP Health Team and other key individuals representing the Ghana Ministry of Health and Ghana Health Service at a local level. Cited benefits of mobile phone CUG include the following [1-4]:

- Improved management and savings: the ability for the organization to more effectively manage communication costs among staff (and potential savings).
- Limitless network size: flexibility in the number of members that can be added to the group (the closed network can scale up or down).
- Increased intra-network access: members of the group can call one another, at any time, in an unlimited capacity.

In part, it was these benefits that prompted the establishment of a mobile phone CUG in each of the MVP sites for the health teams. The CUG has been in place in Bonsaaso since 2009. No assessment of the CUG has been carried out. Consequently, it was not known if and how beneficial the closed user group is to the health staff and their clients. Based on the organizational structure of the Bonsaaso MVP Health Team, it was inferred that aspects of the traditional social network would be mirrored in the social network resulting from the introduction of the mobile phone CUG. Therefore, it was hypothesized that Community Health Nurses would be the most central nodes (or are central actors) in the closed user group network as they serve as intermediaries in the command structure. To determine the validity of this hypothesis, social network analysis methods were used to explore the "social networks" within the mobile phone CUG. By using social network analysis methods, it was believed that the evaluation of the relational ties would allow for more in-depth analysis of the social context in which the closed user group was introduced, potentially changed, and now operates.

For the purposes of this paper, mobile health (mHealth) has been defined in accordance with the definition found in the World Health Organization's Global Observatory for eHealth 2011 Survey as "medical and public health practice supported by mobile devices, such as mobile phones…" [5]. Experts in mHealth have noted that there is not only a need for "high-quality research" in mHealth, but traditional public health methods, such as randomized controlled trials, may not be possible for all mHealth project evaluations [6]. This applies to the context of the MVP closed user groups where randomized controlled trials would require extensive resources and a highly complex study design and analysis.

Douglas Luke underlined the need to move beyond traditional public health methods and look at methods that would allow for more contextual information in community-based research [7]. Social network analysis is a new tool for public health that can be used to explore and answer questions that were not answerable by previous methods [7,8]. For example, this tool can be used to provide information on social and environmental contexts, better explain or provide structure to abstract concepts through mapping and visualization, and present new or previously known information in unique ways for different and additional insights [9-11].

Social network analysis can be used as a managerial tool as well [12-16]. Those managing organizations can maximize their network and employee performance by understanding the informal network. They can gain insight on advice, trust, and communication, in addition to identifying the influential players, gaps in the information flow, and inefficient use of resources [13,17]. This information can then be used to develop appropriate interventions [16].

Important attributes of the [informal] networks that should be assessed to better understand the network include centrality, degree, and betweenness [14,18]. Balkundi and Harrison found that the more dense a network, the better the performance and viability of the team [19]. Granovetter and Burt, separately, suggest that weak ties should also be assessed. Weak ties provide a larger picture of the network and may also explain the formation of subgroups or cliques within a network [20]. They may also provide insight into who may have social capital within a network (eg, those who broker information across the weak ties and, in effect, prevent disconnection in the network) and potentially help explain performance in the network [21,22]. All of these measures were taken into consideration for the assessment of the Bonsaaso MVP CUG.

## Methods

This research was part of a larger evaluation study, the Millennium Villages Project and OASIS II Research on MVG-Net, approved by Columbia University (IRB-AAAF1647).

Five sectors are represented at the Bonsaaso MVP site. The teams (or sectors) are agriculture, infrastructure, enterprise, education, and health. Coordinators lead each of the five sectors and report to the Team Leader, who also functions as the Science Coordinator. The health team's traditional organogram depicts the non-CUG flow of information, highlighting the siloed and rigorous chain of command in the structure (see Figure 1). The Health Coordinator leads the Health Team. Two Health Facilitators are designated to manage the community-based Health Team. The community-based Health Team includes

midwives, Community Health Nurses, and Community Health Extension Workers. One facilitator oversees Community Health Nurses and Community Health Nurses, in turn, oversee Community Health Extension Workers. A separate Health Facilitator oversees midwives. Midwives run the seven clinics in Bonsaaso and work with both the Community Health Nurses and clinic-based Community Health Extension Workers, who are distributed across the 30 communities, or villages. Based on the traditional hierarchy, it appears that Community Health Nurses relay information not only between Community Health Extension Workers and midwives, but also between Community Health Extension Workers and Health Team management members.

Figure 1 shows a traditional organogram of the MVP Bonsaaso Health Team. This organogram is not inclusive of the two ambulance drivers. While they are not included in the organogram, they are considered to be members of the Health Team. They typically report to the Health Coordinator but work closely with the midwives.

Officially, the mobile phone CUG in MVP consists of 79 members. This includes the midwives, Community Health Nurses, Community Health Extension Workers, Health Facilitators, Health Coordinator, Team Leader, physicians at the local referral hospital, ambulance drivers, and local government representatives.

As this network involved a single set of actors and their relations, subsequent analyses were based on one-mode networks. The type of relational tie being assessed was nondirectional communication, or exchange of information, among the actors in the network, as the existence of communication (or a tie) was the most important factor.

The datasets for the social network analyses were based on monthly voice data from Airtel Bharti Ghana. The voice data were formatted in an excel matrix that had the numbers of those in the CUG as the top row and left-most column. Initially, the relational data were provided as the total duration of calls made between two actors over the time period of interest in minutes. This information was converted to binary form as the interest was primarily in the establishment of communication among the actors in the network within a 1-month time period, rather than the duration of the communication. If no communication took place between 2 actors, this was represented in the matrix as “0”. If communication did take place between 2 actors, this was represented as “1” in the dataset. The available datasets represent each of the months from March 2011 through September 2011. A facsimile of the traditional network was also used to illustrate the foundation of the hypothesis. The data were not initially in symmetric form, therefore, the primary researcher reformatted the datasets appropriately for modification into UCINET-ready symmetric files. The researcher then made copies of the spreadsheet and converted the call information into binary form (0 = no tie or call, 1 = tie or call) and formatted appropriately for modification into UCINET-ready excel files. The files were then uploaded into UCINET and converted into UCINET files. A UCINET data file on the traditional social network structure was also generated. As noted previously, important attributes of the network that should be assessed to better understand the network include centrality, degree, and betweenness [14,18]. Therefore, analyses primarily focused on such attributes.

Using UCINET, each of the following analyses were run: (1) ego-network density to look at the size of the ego-network, ties (or degree), diameter, two-step reach, reach efficiency, brokerage and betweenness, (2) network degree centrality: Freeman’s approach and Bonacich’s approach, and (3) cliques to determine subgroups.

To obtain a visual depiction of the network, sociograms were created using NetDraw. The UCINET data files were individually uploaded into NetDraw. To differentiate the actors in the mobile phone CUG by their role on the MVP Health Team, colors were assigned to the different cadres of health workers (see Figure 2). As networks can change over time, these network maps allowed for quick review of whether and how the network may have changed with time. From these sociograms, connectedness of the network was also noted.

Network data from interviews with 23 key informants, from the mobile phone CUG, were used as contextual and supplementary information to the analyses of the entire network. The interviews primarily focused on mobile phone use within and outside of the mobile phone CUG, in-person communication, and advice channels. Call journals that were kept by the same key informants as the interviews for at least a 2-week period were used in a similar manner to the interviews. The journals were used to document calls made or received, purpose of call, and whether those calls were with members of the CUG. Transcripts of the interview were coded and analyzed using NVivo9. The call journals were transcribed into soft copies and analyzed using Excel calculation features and NVivo9 (for the purpose of call section).

For the ego-network analysis, the 10 largest ego-network sizes and the accompanying actors for each of the months were noted in an Excel sheet and then compared across all of the months for which there were data. Ten was chosen as the cut-off of the largest ego-network sizes across all of the months as the ego-network sizes dropped drastically in magnitude, to as low as zero (or no connection), typically after the first dozen largest ego-network sizes were identified. Furthermore, this allowed for consistency across the entire time period. For March, April, July, and August 2011, there were 11 actors that were identified with the largest ego-network sizes. In these instances, there were at least 2 actors in the top 10 who had the same value for their ego-network size. Overall, the actors who had among the 10 largest ego-network sizes were designated as central actors. Interview and call journal information provided additional contextual information on these central actors, where possible.

A similar approach was taken for the socionetwork-level (or complete network-level) centrality analyses as similar trends in the data were observed. Both Bonacich and Freeman tests were run on the network data and the top 10 values were compared across each month. Bonacich represents influence and Freeman represents power. Analyses of cliques were conducted, as well. The clique analysis was set to look at subgroups with 3 or more actors, as the smallest health team

unit includes one Community Health Nurse and two Community Health Extension Workers.

**Figure 1.** Traditional organogram of the MVP Bonsaaso Health Team.

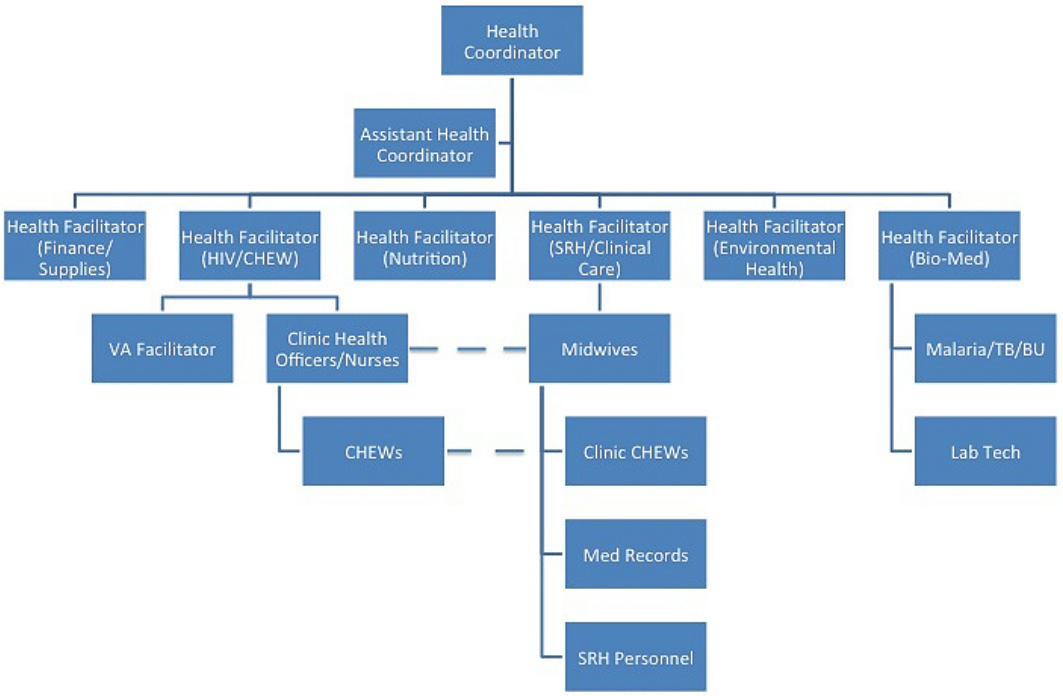
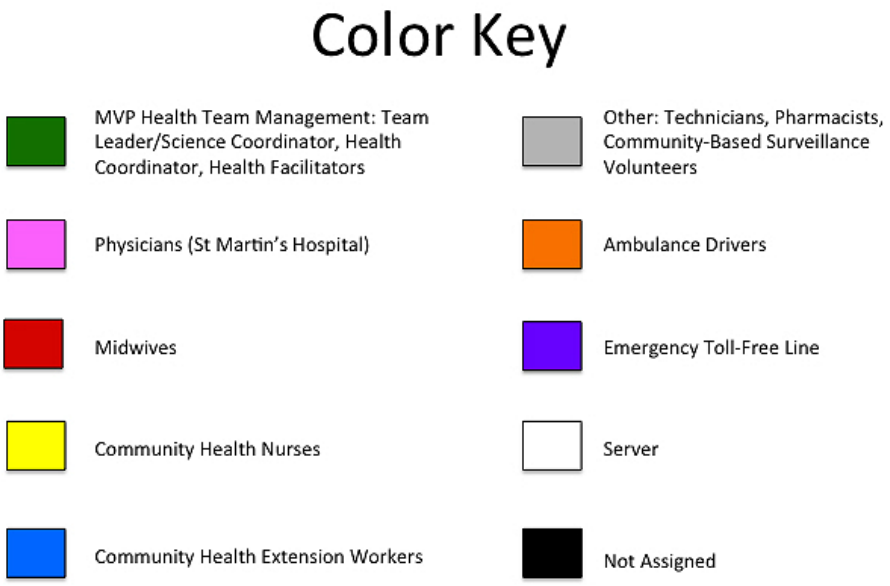



**Figure 2.** Color key for sociograms made in NetDraw on the traditional organizational structure and monthly Airtel closed user group call data.



## Results

The results are discussed below. The monthly results are then followed by a comparison and overview of all the time periods, inclusive of the interview and call journal information.

### Facsimile of Traditional Network

A facsimile of the traditional network was developed and analyzed. There were no disconnected actors in this network. According to the results of the analyses, Health Facilitators, midwives, and Community Health Nurses have the largest ego-networks. Midwives and Community Health Nurses have the highest values for the normalized broker measure. Based on the centrality measures, the midwives and Community Health Nurses have the most power (largest Freeman Betweenness Centrality measures,) and Community Health Nurses have intermediate influence in the network (Health Facilitators and midwives had the greatest influence/largest Bonacich values). A NetDraw image of the facsimile of the traditional network provides an illustration of the above results (see Figure 3).

### Mobile Phone Closed User Group Network

#### *On Relational Ties and Disconnections*

While the number of members of the mobile phone CUG is noted to be 79, the number does vary over time due to new hires and individuals leaving. In March 2011, the number of members in the CUG was 78. Over this month, 525 relational ties were made out of a possible 3081 ties. For the months of April 2011 to September 2011, the mobile phone CUG had 79 members. In April, the number of relational ties increased to 770. The number of relational ties increased in May to 856 and then decreased to 814 in June and July. The number of ties increased substantially to 1106 in August, but then decreased to 908 ties in September.

Over the months of March 2011 to September 2011, the number of actors having no relational ties decreased over time. In March, there were 23 actors disconnected from the network and in September, there were only 10 disconnected actors in the network (see Figure 4).

#### *Identifying the Central Actors*

There were several key actors that were consistently influential or central across all of the 7 months, based on their ego-network measures. All four were members of the Health Team Management. One individual, a Community Health Nurse, was a key central actor in all but one month. For Bonacich Centrality, 2 Community Health Extension Workers were highly prominent in 5 of the 7 months. A midwife, Community Health Nurse, and Community Health Extension Worker were identified as central actors in 4 of the 7 months. With the Freeman Betweenness Centrality, there were 5 individuals that were identified as central actors for 6 of the 7 months. They were 4 Health Team Management members and 1 ambulance driver. When taking the ego-network size, Freeman Betweenness Centrality and Bonacich Centrality measures into consideration, 2 Health Team Management members were identified as the most central actors (high values across all measures).

### Ego-Networks

Ego-network size, density, diameter, reach efficiency, and normalized brokerage values of ego-networks were calculated to help identify the most central actors in the mobile phone CUG network (see Table 1). Across the months, the largest ego-network size ranged from 24 to 40 with a median of 27. Density ranged from 43.5 to 52.8, indicating that, overall, the network is loosely connected and this was consistent over time. The largest diameter of an ego-network was in March 2011 (diameter size of 19.6). For the remaining months, this reduced to 4 or 5, with most months at 4. Across all 7 months, most of the actors had little (<50) to no reach efficiency. The highest normalized brokerage values ranged from 70% to 78%. Over the months, the general trend for the number of subgroups in the network consisting of 3 or more actors increased. The number of subgroups ranged from 92 in March to 423 in September. Among the health worker types, midwives tended to be clustered together in March, April, May, June, August, and September 2011.

**Figure 3.** Sociogram of the traditional network where the relational tie is communication.

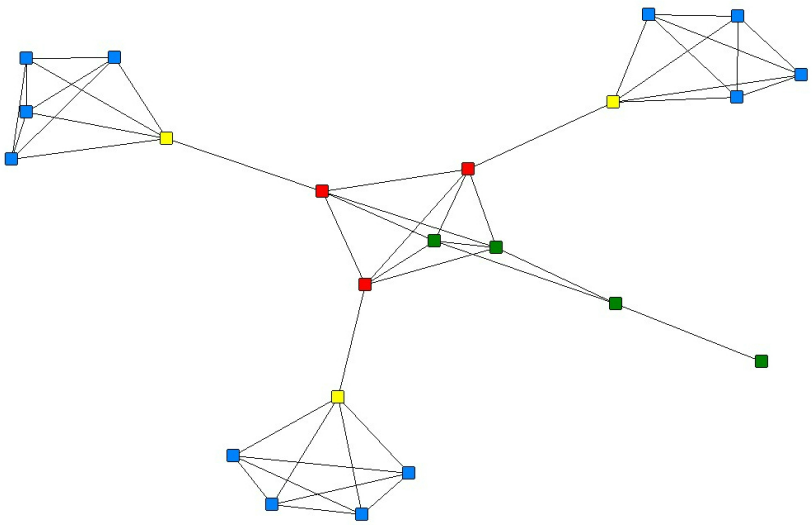


**Figure 4.** Depiction of sociograms from March 2011 through September 2011 juxtaposed with one another and the sociogram of the traditional network.

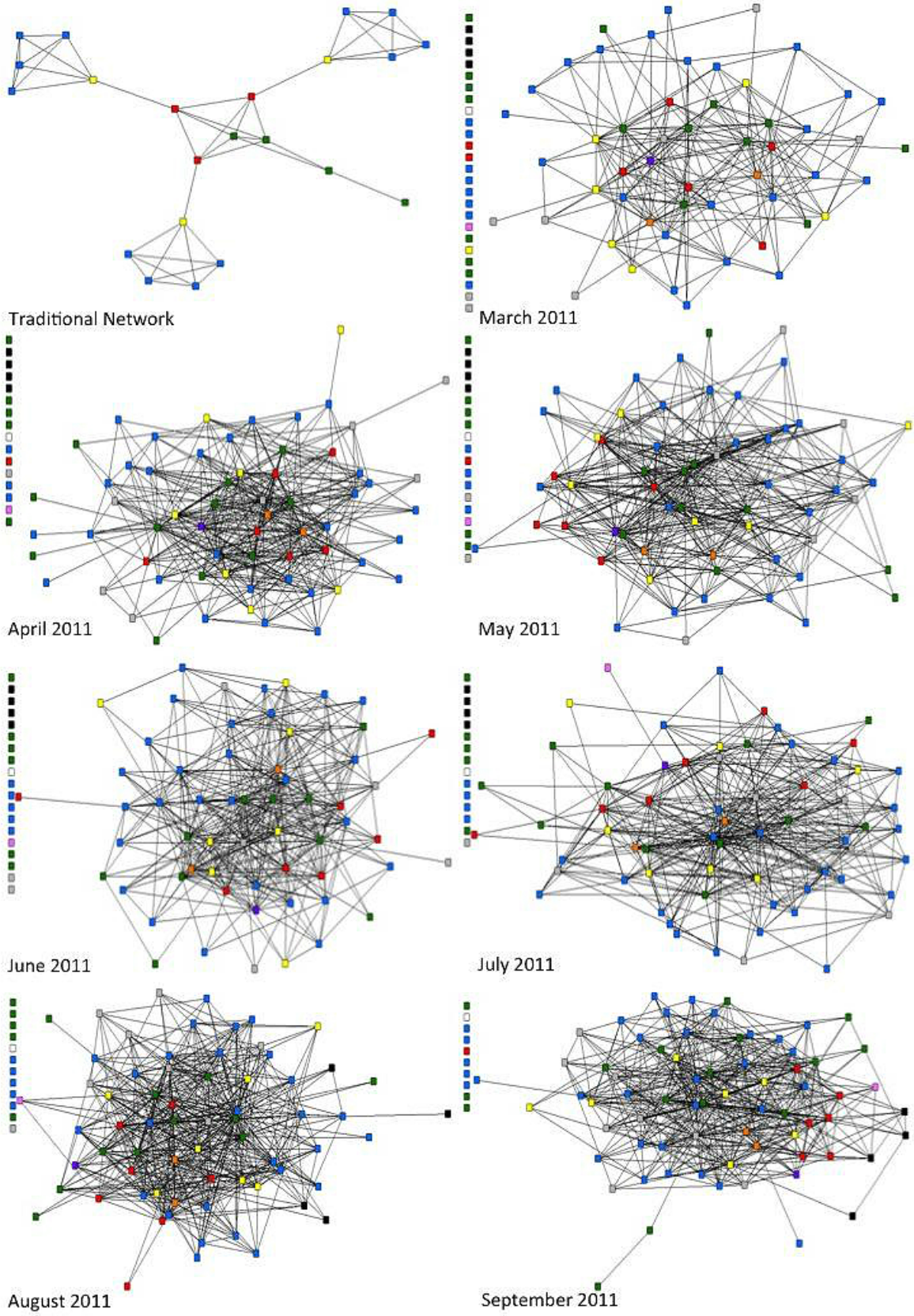


**Table 1.** Summary of ego-network measures of the Airtel mobile phone closed user group by month from March 2011 to September 2011 (source: Airtel closed user group call data).

| | March | April | May | June | July | August | September |
|---|---|---|---|---|---|---|---|
| Largest ENS[a] | 24 | 42 | 41 | 41 | 38 | 45 | 40 |
| Range of top 10 largest ENS | 19–24 | 22–42 | 25–41 | 24–41 | 19–38 | 28–45 | 23–40 |
| Largest density | 43.5 | 52.8 | 48.7 | 47.7 | 46.8 | 52.7 | 43.9 |
| Density range for 10 largest ENS | 30–43.5 | 21.6–52.8 | 29.3–48.7 | 29–47.7 | 22.3–46.8 | 30.6–52.7 | 24.6–43.9 |
| Largest diameter | 19.6 | 4 | 4 | 5 | 4 | 4 | 4 |
| # Actors with no RE[b] | 23 | 16 | 19 | 18 | 15 | 12 | 10 |
| # Actors with RE > 50 | 12 | 13 | 8 | 8 | 8 | 7 | 9 |
| Highest NBV[c] | 70% | 78% | 71% | 71% | 78% | 71% | 75% |
| NBV range for 10 largest ENS | 59%–70% | 54%–78% | 51%–71% | 52%–71% | 53%–78% | 47%–71% | 56%–75% |
| # Subgroups w/ > 3 actors | 92 | 225 | 313 | 273 | 284 | 380 | 423 |

[a]ENS–Ego-network size.

[b]RE–Reach efficiency.

[c]NBV–Normalized brokerage values.

## Centrality

The highest normalized Bonacich Centrality values increased from March through August and then decreased in September. The highest Bonacich Centrality values across the 7 months of network data ranged from 30.7 to 54.7. There was some overlap in the actors identified as central by their ego-network size and Bonacich Centrality measure. In March, there were 8 actors identified as central to the network based on their ego-network size and Bonacich Centrality. In June, July, and August, the number dropped to 3 actors.

There was more overlap of ego-network size and Freeman Betweenness Centrality than ego-network size and Bonacich Centrality. The range of the overlap across the months was small. The values ranged from 6 to 9 actors. Across all the months, the Freeman Betweenness Centrality was varied. This is evidenced by the range, mean, and standard deviation of the normalized values (see Table 2). Overall, the network centralization index was low, ranging from 5.3% in March to 11.01% in April, with a global median of 7.3% and global mean of 7.6%.

**Table 2.** Summary of Bonacich and Freeman centrality measures of the Airtel mobile phone closed user group by month from March 2011 to September 2011 (source: Airtel closed user group call data).

| | | March | April | May | June | July | August | September |
|---|---|---|---|---|---|---|---|---|
| **Freeman Betweenness Centrality (FBC)** | | | | | | | | |
| | # Actors with highest FBC and ENS | 9 | 8 | 7 | 6 | 8 | 7 | 8 |
| | Range of normalized values | 0–5.9 | 0–11.6 | 0–7.8 | 0–9.6 | 0–6.6 | 0–6.8 | 0–9.6 |
| | Mean | 0.68 | 0.7 | 0.65 | 0.67 | 0.82 | 0.8 | 0.95 |
| | Standard deviation | 1.3 | 1.7 | 1.2 | 1.3 | 1.4 | 1.4 | 1.6 |
| | NCI[a] | 5.3% | 11.01% | 7.3% | 9.1% | 5.8% | 6.1% | 8.8% |
| **Bonacich Centrality (BC)** | | | | | | | | |
| | # Actors with highest BC and ENS | 8 | 6 | 5 | 3 | 3 | 3 | 4 |
| | Highest BC value | 37 | 41 | 36.5 | 54.2 | 52 | 54.7 | 30.7 |

[a] network centralization index

## Qualitative Findings

The qualitative findings corresponded with the social network analysis results. Individuals within the mobile phone CUG called other members of the CUG, whether they were within or outside of their catchment area. This was unlike the in-person communication and pre–mobile phone CUG communication that tended to be limited to intracatchment communication. In addition, the informants regularly established communication with the ambulance drivers in cases of emergencies.

## Discussion

We identified the central players in the mobile phone CUG network using both ego-network and sociocentric (or complete network) approaches. The ego-network size was the first indicator to be analyzed, as the size of the individual network is indicative of being a central figure in the network. Individuals with larger ego-networks may have greater influence in the network as compared to others with smaller ego-networks.

Based on the traditional network structure, Community Health Nurses form a bridge between the Community Health Extension Workers and the Health Team management (directly linking to the Health Facilitators). Community Health Nurses also connect the Community Health Extension Workers with midwives. Therefore, it was surmised that Community Health Nurses would be central actors in the network as they should have a large ego-network size and serve as intermediaries (high broker values). However, when analyzing the actual mobile phone CUG network data, this was found to not be the case.

Across all of the months, the ego-network diameters, even at their highest levels, were low. This indicates that the ego-networks were not very extensive. This may be due to most individuals in the network having set communication patterns and these communications taking place in a pre-defined subgroup (such as a catchment area). Similarly, the reach efficiencies of the ego-networks were low to moderate. Those with smaller ego-network size tended to have higher reach efficiencies. This could be due to the fact that those with smaller ego-networks made relational ties with those with larger ego-networks, while those with larger ego-network sizes had relational ties to those with small ego-networks (and thus, the extent of the network beyond the alters was limited). Accordingly, those with larger ego-networks tended to play the broker role relatively frequently. In the overall networks, those who did fulfill the role of a broker filled that role at a moderate to high frequency.

The socionetwork indicators allowed for additional information on key actors in the network. Since the Bonacich Centrality calculation was based on a positive beta, the resulting data were representative of the most central actors who hold power due to being connected to others in the network who are well connected. From the results, it appears that power (based on being connected to those who are well connected) is limited. The power lies with few individuals. This finding may help explain why there was less overlap between the ego-network size and Bonacich Centrality than the Freeman Betweenness Centrality. The Bonacich Centrality findings of having a network with limited power correlate with the ego-network size results. If the network were to have more power, those with larger ego-network sizes would be connected to others with larger ego-network sizes.

Looking at betweenness via the Freeman Betweenness Centrality measure, across all of the months, there was considerable variation in this measure. With that noted, it was not surprising to find that ambulance drivers were among the central actors in the network and had high levels of betweenness activity. Ambulance drivers receive calls from any of the health care workers and then may have to relay that information to others in the network or beyond. However, the overall network centralization index values for the Freeman Betweenness Centrality measures were low, meaning that connections in the network can be made without intermediaries. Despite the traditional network structure, the mobile phone CUG structure has disrupted this, allowing individuals who would not normally communicate with one another to make such connections directly, rather than going through other actors.

Subgroups, or cliques, exist in the network, as evidenced by the results from the clique analysis in UCINET. Therefore, the mobile phone CUG has a community structure. This community structure may partially explain the fact that the networks were found to be moderately connected. Some of the subgroups may be more tightly connected within their respective group and weakly connected with other subgroups. As most of the Health Team is divided into catchment areas, there may be natural divisions. There may also be natural divisions among the health worker types (eg, midwives communicate closely among one another). The mobile phones would also allow for additional nontraditional cliques to form. Further analyses on cliques and an additional measure for power, structural holes, would need to be carried out to better determine the community structure [23,24].

Limitations of these analyses include focusing primarily on the largest values for the measures. Also, 10 was an arbitrary cut-off number but allowed for interesting comparisons among the various measures and more manageable management of the results. In terms of context, there were individuals experiencing phone issues at this time period due to blocked SIMs over the time period of interest. The blocked SIM issue was resolved after September 2011. Therefore, those who were disconnected from the network may, in reality, may not truly be disconnected. Additionally, there may have been key players experiencing phone issues throughout the duration of the time period who were not labeled as such due to their limited phone connectivity.

### Conclusion

Based on the traditional network structure, it would appear that the Community Health Nurses would be highly prominent, or central, in the network. However, this was not the case, meaning that the evidence does not support Community Health Nurses being the most central actors in the network. Rather than Community Health Nurses, it was the members of the Health Team Management that were the more central players in the network. While this finding appears to support evidence that the traditional chain of command and communication flows have been “disrupted” by the mobile phone CUG, the Health Team management structure may, itself, be a confounder to the relationship between the mobile phone CUG and communication flows.

Nevertheless, it is evident that social network analysis methods can be a useful analytical tool, especially in the context of mHealth, health services, and operational/managerial research. The results on the voice traffic in the mobile phone CUG and the resulting social network structure can be used by the MVP Ghana Health Team to reflect on their current work-related mobile phone use and identify opportunities and areas to better

manage team communication. For example, with the Health Team Management being such central actors within the network, they may be receiving multiple calls on the same topic or calls that could be better addressed by another actor in the network. Therefore, policies could be put into effect that could divert such calls to the appropriate channels and free up this time for the Health Team Management to attend to other tasks. The Health Team could also conduct similar social network analyses at later time periods to follow-up on any changes that they may put into place regarding the use of the mobile phone CUG.

Therefore, the social network analysis methodology can be used to determine the most conducive/productive structure for an organization/team, identify gaps in communication, identify key or central actors with greatest influence, and more. In the future, health outcomes research should also attempt to incorporate social network analysis methodology, as organizational structure of health service teams, organizations, and even communities may have an impact on the health outcomes of the communities being served.

## Acknowledgments

The authors would like to acknowledge the Novartis Foundation for Sustainable Development for their support of the Ghana Telemedicine Project and IDRC for their support to the Millennium Villages Project through the OASIS II Research Project, through which this project was made possible. Additional support was provided to NNK through the Global mHealth Initiative at Johns Hopkins University. The authors would like to extend their gratitude to colleagues at Airtel Bharti in Ghana, the Earth Institute at Columbia University, the Ghana Millennium Villages Project and the Johns Hopkins School of Public Health, without whom, this project would not have come to fruition.

## Conflicts of Interest

None declared.

## References

1. Airtel. Airtel Corporate User Group (CUG) URL: http://www.ng.airtel.com/Business/AirtelCorporatePackage/AirtelCUG.aspx [WebCite Cache ID 6CLsUjUw0]
2. Digicel. FAQs: Closed User Group URL: http://www.digicelbonaire.com/en/help_faqs/products/faqs-closed-user-group [accessed 2012-11-21] [WebCite Cache ID 6CLsFActH]
3. Roshan. Closed User Group URL: http://www.roshan.af/Roshan/Business/Solutions/Talk/Closed_User_Group.aspx [accessed 2012-11-21] [WebCite Cache ID 6CLsHdI1s]
4. Setar. Closed User Group Calling URL: http://setar.aw/mobile/closed-group-calling.html [accessed 2012-11-21] [WebCite Cache ID 6CLsIt3Ff]
5. World Health Organization. Global Observatory for eHealth Series;3. 2011. mHealth: New horizons for health through mobile technologies URL: http://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2013-02-04] [WebCite Cache ID 6EBskXEaM]
6. Whittaker R. Issues in mHealth: findings from key informant interviews. J Med Internet Res 2012;14(5):e129 [FREE Full text] [doi: 10.2196/jmir.1989] [Medline: 23032424]
7. Luke DA, Harris JK. Network analysis in public health: history, methods, and applications. Annu Rev Public Health 2007;28:69-93. [doi: 10.1146/annurev.publhealth.28.021406.144132] [Medline: 17222078]
8. Luke DA. Getting the big picture in community science: methods that capture context. Am J Community Psychol 2005 Jun;35(3-4):185-200. [Medline: 15909794]
9. Leischow SJ, Milstein B. Systems thinking and modeling for public health practice. Am J Public Health 2006 Mar;96(3):403-405. [doi: 10.2105/AJPH.2005.082842] [Medline: 16449572]
10. Borgatti SP, Mehra A, Brass DJ, Labianca G. Network analysis in the social sciences. Science 2009 Feb 13;323(5916):892-895. [doi: 10.1126/science.1165821] [Medline: 19213908]
11. Fisher D. Using Egocentric Networks to Understand Communication. IEEE Internet Comput 2005 Sep 2005;9(5):20-28. [doi: 10.1109/MIC.2005.114]
12. Balkundi P, Kilduff M. The ties that lead: A social network approach to leadership. The Leadership Quarterly 2006 Aug 2006;17(4):419-439. [doi: 10.1016/j.leaqua.2006.01.001]
13. Chauvet V, Chollet B, Soda G, Huault I. The contribution of network research to managerial culture and practice. European Management Journal 2011 Oct 2011;29(5):321-334. [doi: 10.1016/j.emj.2011.06.005]
14. Johanson J. Formal structure and intra-organisational networks. An analysis in a combined social and health organisation in Finland. Scandinavian Journal of Management 2000 Sep 2000;16(3):249-267. [doi: 10.1016/S0956-5221(99)00027-5]
15. Krackhardt D, Hanson JR. Informal networks: the company behind the chart. Harv Bus Rev 1993;71(4):104-111. [Medline: 10127036]
16. Woodard KL, Doreian P. Utilizing and understanding community service provision networks: a report of three case studies having 583 participants. Journal of Social Service Research 1994 Mar;18(3-4):1-41. [doi: 10.1300/J079v18n03_01]

17. Stewart SA, Abidi SS. Applying social network analysis to understand the knowledge sharing behaviour of practitioners in a clinical online discussion forum. J Med Internet Res 2012;14(6):e170 [FREE Full text] [doi: 10.2196/jmir.1982] [Medline: 23211783]
18. Freeman LC. Centrality in Social Networks: Conceptual Clarification. Soc Networks 1978;1:215-239 [FREE Full text]
19. Balkundi P, Harrison DA. Ties, leaders, and time in teams: strong inference about network structure's effects on team viability and performance. Academy of Management Journal 2006 Feb;49(1):49-68. [doi: 10.5465/AMJ.2006.20785500]
20. Granovetter MS. The Strength of Weak Ties. Am J Sociol 1973;78(6):1360-1380 [FREE Full text]
21. Balkundi P, Kilduff M, Barsness ZI, Michael JH. Demographic antecedents and performance consequences of structural holes in work teams. J. Organiz. Behav 2007 Feb 2007;28(2):241-260. [doi: 10.1002/job.428]
22. Burt RS. The network structure of social capital. Research in Organizational Behavior 2000 Jan 2000;22:345-423. [doi: 10.1016/S0191-3085(00)22009-1]
23. Fortunato S. Community detection in graphs. Physics Reports 2010 Feb;486(3-5):75-174. [doi: 10.1016/j.physrep.2009.11.002]
24. Gustafsson M, Hörnquist M, Lombardi A. Comparison and validation of community structures in complex networks. Physica A: Statistical Mechanics and its Applications 2006 Jul 2006;367(15 July 2006):559-576. [doi: 10.1016/j.physa.2005.12.017]

## Abbreviations

**BC:** Bonacich centrality
**CUG:** closed user group
**ENS:** ego-network size
**FBC:** Freeman betweenness centrality
**mHealth:** mobile health
**MVP:** Millennium Villages Project
**NBV:** normalized brokerage values
**NCI:** network centralization index
**RE:** reach efficiency

*Edited by G Eysenbach; submitted 02.09.12; peer-reviewed by H Cole-Lewis, A Fuad; comments to author 08.11.12; revised version received 11.12.12; accepted 06.02.13; published 03.04.13.*

*Please cite as:*
*Kaonga NN, Labrique A, Mechael P, Akosah E, Ohemeng-Dapaah S, Sakyi Baah J, Kodie R, Kanter AS, Levine O*
*Using Social Networking to Understand Social Networks: Analysis of a Mobile Phone Closed User Group Used by a Ghanaian Health Team*
*J Med Internet Res 2013;15(4):e74*
*URL: http://www.jmir.org/2013/4/e74/*
*doi:10.2196/jmir.2332*
*PMID:23552721*

Original Paper

# The Personal Health Record Paradox: Health Care Professionals' Perspectives and the Information Ecology of Personal Health Record Systems in Organizational and Clinical Settings

Kim M Nazi[1], BS, MA, PhD

Veterans and Consumers Health Informatics Office, Veterans Health Administration, Department of Veterans Affairs, Albany, NY, United States

**Corresponding Author:**
Kim M Nazi, BS, MA, PhD
Veterans and Consumers Health Informatics Office
Veterans Health Administration
Department of Veterans Affairs
c/o Stratton VA Medical Center
113 Holland Avenue
Albany, NY, 12208
United States
Phone: 1 449 0621
Fax: 1 449 0200
Email: kim.nazi@va.gov

## Abstract

**Background:** Despite significant consumer interest and anticipated benefits, overall adoption of personal health records (PHRs) remains relatively low. Understanding the consumer perspective is necessary, but insufficient by itself. Consumer PHR use also has broad implications for health care professionals and organizational delivery systems; however, these have received less attention. An exclusive focus on the PHR as a tool for consumer empowerment does not adequately take into account the social and organizational context of health care delivery, and the reciprocal nature of patient engagement.

**Objective:** The purpose of this study was to examine the experiences of physicians, nurses, and pharmacists at the Department of Veterans Affairs (VA) using an organizationally sponsored PHR to develop insights into the interaction of technology and processes of health care delivery. The conceptual framework for the study draws on an information ecology perspective, which recognizes that a vibrant dynamic exists among technologies, people, practices, and values, accounting for both the values and norms of the participants and the practices of the local setting. The study explores the experiences and perspectives of VA health care professionals related to patient use of the My HealtheVet PHR portal and secure messaging systems.

**Methods:** In-depth interviews were conducted with 30 VA health care professionals engaged in providing direct patient care who self-reported that they had experiences with at least 1 of 4 PHR features. Interviews were transcribed, coded, and analyzed to identify inductive themes. Organizational documents and artifacts were reviewed and analyzed to trace the trajectory of secure messaging implementation as part of the VA Patient Aligned Care Team (PACT) model.

**Results:** Study findings revealed a variety of factors that have facilitated or inhibited PHR adoption, use, and endorsement of patient use by health care professionals. Health care professionals' accounts and analysis of organizational documents revealed a multidimensional dynamic between the trajectory of secure messaging implementation and its impact on organizational actors and their use of technology, influencing workflow, practices, and the flow of information. In effect, secure messaging was the missing element of complex information ecology and its implementation acted as a catalyst for change. Secure messaging was found to have important consequences for access, communication, patient self-report, and patient/provider relationships.

**Conclusions:** Study findings have direct implications for the development and implementation of PHR systems to ensure adequate training and support for health care professionals, alignment with clinical workflow, and features that enable information sharing and communication. Study findings highlight the importance of clinician endorsement and engagement, and the need to further examine both intended and unintended consequences of use. This research provides an integral step toward better understanding the social and organizational context and impact of PHR and secure messaging use in clinical practice settings.

***(J Med Internet Res 2013;15(4):e70)*** doi:10.2196/jmir.2443

**KEYWORDS**

personal health record (PHR); health professionals; attitudes; eHealth; electronic health records (EHRs); secure messaging; qualitative analysis

## Introduction

Personal health records (PHRs) are designed as tools to engage patients in their health care and to enable them to manage their personal health information [1-6]. Significant investments have been made by organizations to offer PHRs based on the desire to enhance patient-centered care [7-11], and the perceived potential for health care system improvement [12,13]. Historically, most PHR implementation efforts have focused on broad conceptions of consumer empowerment in which advocates emphasize the potential for PHRs to (1) increase consumer access to and control over health information, and (2) enable active patient participation in health care decision making and health management [14,15]. Despite high consumer interest in PHRs [16-21] and growing availability [22-25], a paradox exists in that adoption remains relatively low overall [26-28]. A national consumer survey conducted by the Markle Foundation revealed that only 10% of American adults currently use a PHR [29]. One notable exception to the low rate of PHR adoption is Kaiser Permanente: at the end of 2012, 4 million of its 9 million members had registered to use its patient portal, My Health Manager [30]. Among veterans, 71% utilize the Internet, and approximately one-fifth report using the US Department of Veterans Affairs (VA) PHR, My HealtheVet [31].

Preliminary findings in the literature suggest that provider endorsement may be an important factor in a patient's choice to adopt a PHR, and that continued clinician engagement with patient PHR use may be required to achieve and sustain anticipated positive outcomes [32,33]. A national consumer study of public use and attitudes toward PHRs concluded that doctors may hold the key to increasing adoption [28]. Although there has been a prominent focus on PHRs as tools to support consumers, the value that consumers derive from the use of a PHR will likely be directly affected by the attitudes and actions of health care providers and team members within the context of the clinical setting. What may be missing from the current consumer empowerment paradigm is a deeper understanding of how patient PHR use unfolds within the context of the health care interaction and how it influences the provision of services by health care professionals in organizational settings.

Although PHRs are designed as consumer-oriented tools, understanding the consumer perspective is necessary, but insufficient by itself. Consumer PHR use also has broad implications for health care providers and delivery systems; however, these have received less attention. Many authors agree that the social and organizational changes implicated in patient use of PHRs will require a significant culture change for medicine [34-38]. Others raise more practical issues for physician practices, such as ensuring appropriate safeguards for release of information to patients [39-42], and determining how best to manage potentially large volumes of self-reported information within the limited time allocated to the clinical visit [43].

Greenhalgh and Swinglehurst [44] present a strong case for studying technology use as a social practice calling for ethnographic approaches that recognize technologies both shape, and are shaped by, human action. Although a variety of anticipated benefits are attributed to the use of PHRs, there is a critical need to examine use from a social and organizational perspective. Given the persistent paradox between reported patient interest and anticipation of benefits with relatively low adoption and little evidence about impact, PHR use must be examined as a component of health care work, influenced by and influencing organizational actors and their work within the health care ecosystem. Without this understanding, the rhetoric of consumer empowerment may have minimal effects in everyday settings in which patients and health care professionals interact.

### Objective

The purpose of this study was to examine the experience of VA health care professionals whose patients use an organizationally sponsored PHR system (My HealtheVet) to develop insights into the interaction of technology and the processes of health care delivery. The study aimed to explore the experiences and perspectives of health care professionals related to patient use of 4 PHR features that could have important ramifications for health care professionals: patient health education resources, tools to support medication reconciliation, tools to enable patient tracking and self-reporting of data, and electronic communication via secure messaging.

### Setting

The Veterans Health Administration (VHA) of the VA is the largest integrated health care system in the United States with over 1600 sites of care. VA health care facilities provide a broad spectrum of medical, surgical, and rehabilitative care to approximately 5.5 million patients annually. The VHA was an early pioneer in utilizing an enterprise-wide electronic health record (EHR) system [45] and piloting a tethered PHR prototype [46]. The national My HealtheVet PHR portal [47] was released in November 2003.

The My HealtheVet portal enables veterans to create and maintain a Web-based PHR that provides access to patient health education information and resources, a comprehensive personal health journal, and electronic services, such as prescription refills and secure messaging [48-50]. All site visitors can access health education resources, and veterans who self-register for an account can create a customized PHR and request VA prescription refills. For veterans who are VA patients, a 1-time process of authentication enables access to selected data from the VA EHR, such as laboratory test results and medication history. The VA Blue Button was added to the portal in August 2010, enabling veterans to generate and download an electronic file that contains their personal health information [51].

My HealtheVet secure messaging enables authenticated users who are VA patients to interact with their health care providers

and VA staff electronically to exchange nonurgent health-related information, to request an appointment or prescription renewal, or ask health-related questions. A triage process similar to the telephone triage process enables a member of the health care team to read and respond to incoming messages or assign action to another member of the team. Users can set their preferences to receive a notification via email that a new secure message is available. Messages that are not marked as completed in the system are escalated, generating an alert to health care team members. Members of the health care team can elect to save selected parts of the interaction as a progress note in the VA EHR.

Secure messaging was initially available for a limited number of participants to support alignment with clinical practice workflows and enable participant feedback to guide refinement of the application. With the VA transformation to the Patient Aligned Care Team (PACT) model [52], based on the medical home model, incremental expansion of secure messaging made the service available in primary care at all VA facilities in early 2011. Expansion to specialty and surgical care settings and nonclinical areas continues.

As of September 2012, the My HealtheVet PHR portal had served more than 1.9 million registrants, with more than 896,000 authenticated VA patients. The site had logged more than 76 million visits, and veterans had requested more than 33.5 million prescription refills. The VA Blue Button had been used by more than 636,000 unique users with more than 2.4 million file downloads. More than 479,000 VA patients had also opted-in to use secure messaging. Most veterans currently visit the site to use the pharmacy-related features and satisfaction with the site is high [9].

## Methods

Given the paucity of research about how patient use of PHRs is experienced by health care professionals in the social and organizational contexts in which they are situated, a qualitative methodology was deemed most appropriate because the goal of this study was to gain an understanding of actors' experiences and perspectives [53-56]. The study consisted of conducting in-depth interviews with VA health care professionals to better understand their experiences and perspectives related to use of the organizationally sponsored PHR, My HealtheVet. Although the initial aim of the study was to focus on 4 specific PHR features that could have important ramifications for health care professionals (patient health education resources, tools to support medication reconciliation, tools to enable patient tracking and self-reporting of data, and electronic communication via secure messaging), secure messaging emerged as the prominent focus for the study because it was used most often by study participants. Given this shift in focus, an analysis of the organizational implementation of secure messaging and the parallel development and implementation of PACT was also undertaken to gain a deeper understanding of the adoption and use of this technology in an evolving situated context. The study was reviewed and approved by both the Washington DC VA Medical Center and the State University of New York at Albany Institutional Review Boards (IRBs).

### Sampling and Recruitment

The sampling strategy used for this study was purposeful and theoretical because of the variable level of adoption and uncertainty about the degree to which participants had experiences with patient PHR use [53]. The sample consisted of a stratified purposive sample of VHA health care professionals in 3 groups: health care providers (including physicians, physician assistants, nurse practitioners, and advance practice nurses), nurses, and pharmacists. Criteria for participation in the study included that the participant: (1) was a VHA staff member (health care provider, nurse, or pharmacist) involved in direct patient care, (2) reported having experience with patient use of 1 or more of 4 My HealtheVet features, and (3) was willing and available to participate in a 45 to 60 minute in-depth interview.

Participants were recruited directly at 3 VHA discipline-specific organizational meetings in which the target audience were health care professionals. Interested individuals who could self-identify as meeting the sampling criteria were invited to contact the research investigator. A second recruitment strategy consisted of snowball sampling using My HealtheVet coordinators as informants to identify potential study participants [57]. The My HealtheVet coordinator is an organizational role at each VA Medical Center and Veteran Integrated Service Network (VISN) tasked with leading local efforts to implement My HealtheVet. Similarly, additional referrals were received during the actual in-depth interviews. Theoretical saturation was reached with the completion of 30 interviews.

### Data Collection

Recruited participants consented in writing and were then asked to complete a short questionnaire that included general demographic questions and questions about their experience with patient use of the 4 My HealtheVet features of interest to validate study criteria and to enable a purposeful sampling of the participant pool. An interview guide was used for each interview, and the data collected through the short background questionnaire used to refine the focus of the individual interviews.

An initial round of 3 interviews (1 health care provider, 1 nurse, and 1 pharmacist) was conducted in January and February 2011 to pilot-test the interview guide and allow for refinements. Data collection was staged over time (February through July 2011) to enable iterative coding and analysis. Because study participants came from various locations across the country, interviews were conducted by telephone at a time that was convenient for the health care professional given their clinical schedule. All interviews were recorded using a digital audio recorder and were transcribed verbatim by the research investigator.

### Data Analysis

Responses to the short background questionnaires were recorded in a spreadsheet to facilitate analysis of the sample and continued purposive sampling. Review and analysis of the interview data was ongoing throughout the study in conjunction with continued data collection until theoretical saturation was reached [58]. The analysis process involved an ongoing review

of all data and project documents to identify common themes utilizing modern techniques of qualitative analysis. Coding and analysis of the interview transcripts employed an inductive approach using the perceptions and reported experiences of participants as the basis for constructing and organizing the codes and categories [53]. Atlas Ti version 6.2 qualitative analysis software (Scientific Software Development, Berlin, Germany) was used to organize the interview transcripts, facilitate data coding and sorting, and document memos.

Interview transcripts were reviewed systematically using open coding [53,58-60] to develop an initial coding scheme. After an initial coding structure was developed, the data were reviewed iteratively as the study progressed to review and refine codes, and to identify additional codes. Analysis continued with axial coding to identify relationships between code categories. The constant comparison method [61] was used to continuously refine the codes and coding categories. After developing preliminary interpretations, all data were reviewed for possible alternative interpretations and rival conclusions. Field notes were prepared immediately following each interview and investigator memos were written to further document findings as they emerged from the data and analysis. Analytic memos captured emerging insights and connections between codes and themes, and integrative memos were used to develop theoretical connections between the coded data excerpts.

Based on the study findings generated by the analysis of the interview data, a review and analysis of organizational documents was also undertaken in October 2011 to reconstruct the history of secure messaging implementation and relevant organizational changes that happened in parallel. Documents that were included in this analysis consisted of workgroup meeting minutes, status reports, memorandums, statistical reports, and project implementation reports. Further analysis of the study participant characteristics based on their responses to the short background questionnaire coupled with data from their interview transcripts then enabled a deeper understanding of each participant and the context of their perspective in terms of their experience with respect to this trajectory.

Several analytic techniques were used to improve the validity and reliability of the study [62]. Study participants were assured of the anonymity of their comments. Other VA researchers with experience in qualitative research were consulted at key milestones in the project to review sample codes, memos, and themes. As the study progressed, participants in later interviews were asked at the end of their interview to comment on emerging themes from earlier participants. A summary of the implementation milestones and timeline was reviewed and validated by implementation leaders. Relevant reflections generated through reflexive processing were documented in a project journal and reviewed as part of the analytic process. Member checking was also performed by inviting all study participants to review a summary of findings to check the authenticity of the investigator's interpretations [63].

## Results

A total of 30 VA health care professionals participated in the study (10 health care providers, 10 nurses, and 10 pharmacists). As shown in Table 1, study participants reported working predominantly in primary care settings and spending the majority of their work time (62% on average) on direct patient care activities. Four health care professionals reported working in specialty care (eg, audiology).

### Comparison of Health Care Professionals' Perspectives

The design of this study was intended to allow for a comparison of perspectives for 3 different types of health care professional roles: health care providers, nurses, and pharmacists. Themes were consistent across the 3 professions. This consistency may reflect the unifying nature of an integrated delivery system, especially because the system as a whole has recently undergone a systemic transformation to the PACT model. Where variations did exist, they represented differences in areas of focus reflective of varying roles rather than disagreements. For example, although health care providers vocalized concerns about the lack of workflow fit for tools to support patient self-reported data, nurses focused more on patient motivation, and pharmacists focused on practical workload implications. All 3 groups consistently emphasized the barriers associated with lack of access to patient self-reported data, and expressed concerns about the potential for mismatched patient expectations. For secure messaging, health care providers emphasized improving the quality of the clinical visit, whereas nurses uniquely emphasized the importance of providing patients with information that they could refer back to later. Pharmacists emphasized the potential workload burden for secure messaging because many messages request prescription refills, and these are triaged and assigned to pharmacists for completion. These nuances reflect the varying work tasks that health care professionals perform. All 3 groups consistently emphasized the positive consequences of secure messaging.

**Table 1.** Study participant gender and self-reported type of work and activity.

| Characteristics | Providers (n=10) | Nurses (n=10) | Pharmacists (n=10) | All groups (N=30) |
|---|---|---|---|---|
| **Gender, n** | | | | |
| Male | 6 | 0 | 4 | 10 |
| Female | 4 | 10 | 6 | 20 |
| **Type of work setting, n** | | | | |
| Primary care | 7 | 8 | 9 | 24 |
| Specialty care | 2 | 1 | 1 | 4 |
| Both/other | 1 | 1 | 0 | 2 |
| **Type of work activities, %** | | | | |
| Direct patient care | 55 | 53 | 78 | 62 |
| Administrative work | 28 | 43 | 14 | 28 |
| Other | 17 | 4 | 8 | 10 |

## Key Findings

Prominent themes emerging from the study were organized into 5 key findings as shown in Textbox 1.

### *Underutilization of My HealtheVet Personal Health Record Features*

In general, health care professionals reported limited experiences with patient use (and their own use) of My HealtheVet health education resources, tools to support medication reconciliation, and tools to support patient self-reported data (with some exceptions), often using alternative tools and resources instead. Health care professionals identified several barriers to use of the My HealtheVet PHR features, and commented that these barriers have also limited their endorsement of patient use.

### *Factors Inhibiting My HealtheVet Personal Health Record Adoption and Use*

Health care professional's accounts provide evidence that the My HealtheVet portal has been conceptualized by many as a tool for patients and separate from the clinical encounter, with some notable exceptions. Several study participants indicated that they had not experienced patient use of the tools, nor advised patients to use the tools, viewing My HealtheVet as a self-service portal for patients.

Many health care professionals reported general awareness of My HealtheVet but limited familiarity with its features, with the exception of secure messaging. Health care professionals note that this lack of knowledge limits their ability to endorse patient use, or to integrate use of My HealtheVet features within the clinical practice setting. They emphasize that increasing staff knowledge about the various features would enable staff to better utilize the available tools and resources, and to encourage patient use. Many commented that time constraints hamper their ability to become more familiar with these resources, and also to educate patients about them.

To educate both staff and patients, health care professionals emphasize that demonstrating functionality is important, advocating for approaches that enable hands-on experiences. Several commented that providing staff with opportunities to learn more about My HealtheVet will require that time be allocated to these activities with sufficient coverage of their patient care responsibilities. Availability of patient-accessible computers in the clinic setting was identified as an important structural need that would enable patients to learn more about the tools and to make use of them in concert with their clinical visit.

Health care professionals often reported using alternative tools and resources. For example, although My HealtheVet provides a significant library of health education resources, health care professionals already use alternative resources, such as subscription-based software that is linked from within the primary clinical workflow system or resources retrieved from the Internet, with little incentive to change. Health care professionals reported that they increasingly use Internet resources easily found by search engines, and speculated that patients do as well. As one health care provider said: "Why not just Google?"

Several health care professionals commented about the need to enable a delegation feature that would provide veterans with the ability to share their personal health information. They noted that given the current My HealtheVet system capabilities, patient-accessible computers within the clinic setting are needed so that patients can access these tools in conjunction with their visit. Health care providers could then view and discuss patient-tracked data, nurses could demonstrate relevant patient health education materials, and pharmacists could review patient self-entered medication lists with the patients to update the medication list within the VA EHR. Some health care professionals noted that even with a delegation feature in the portal, logging into a system outside of the primary clinical workflow system would be a barrier for them.

Likewise, health care professionals indicated that the inability for patients to share information with their health care team through the My HealtheVet system limits the value of tools to support patient self-entered data. Although health care professionals report that tools to support patient-tracked health

metrics, such as blood glucose readings, can be helpful, the lack of timely report and communication has made these tools less useful because delegation is not yet available. Health care professionals also expressed concerns that patients often believed that the health care team could view data that they had entered, further influencing health care professionals' endorsement of use. Several noted that this lack of integration with the primary clinical information system deters use because self-entered data are inaccessible to the health care team.

Although My HealtheVet also provides tools to support medication reconciliation, health care professionals consistently reported that, regardless of whether patients supply their own medication lists, the process of medication reconciliation always comes down to communication. They consistently described medication reconciliation as a standardized process that inherently involves a dialog between the patient and members of the health care team, and often also involves the patient's family member(s) or caregiver(s). One health care provider summarized this by saying that medication reconciliation is "a very complex animal" the goal of which is to "really figure out what the patient is taking." Because the primary focus of the process is to update the medication list in the medical record, health care professionals begin and end the process with the medication list on record within the VA EHR, reviewing this list with the patient and providing the patient with a copy of the updated list. Although patient input is seen as essential to this process, health care professionals report that patient-supplied lists are often suspect, either because they are not up-to-date when brought to the periodic clinical visit or because they may contain other inaccuracies.

### *Implementation, Use, and Endorsement of Secure Messaging*

In contrast to their experiences with other My HealtheVet features, health care professionals reported successfully using secure messaging and routinely endorsing patient use. Although several of the study participants were early adopters of secure messaging, others began use of the system as part of the organization-wide implementation of the PACT model. Analysis of the trajectory of implementation revealed that secure messaging began as an innovation project, spread to other sites via early adopters, and was ultimately assimilated and routinized throughout the system. A history of implementation milestones is shown in Table 2. Health care professionals' accounts revealed several factors that have facilitated the adoption and use of secure messaging.

### *Factors Facilitating Secure Messaging Adoption and Use*

Health care professionals' accounts revealed that in contrast to other My HealtheVet features, secure messaging has been perceived by health care professionals as a tool that has significant value for both themselves and their patients, and as having attributes that encourage adoption, including relative advantage over existing alternatives, compatibility with existing clinical systems, and fit with existing workflow. Secure messaging was implemented in a way that allowed for organic growth in use of the system, with opportunities to try out the system with a small number of patients, observe the success of others, and interact with other users. As implementation progressed, it led to the emergence of structures that further facilitated adoption, including the development of performance measures, decentralization of patient authentication processes within local clinic settings, and the emergence of clinical reminders within the VA EHR to prompt endorsement at patient visits.

In contrast to the My HealtheVet PHR, secure messaging implementation was also accompanied by training and education programs for health care professionals. Part of this training was devoted to alignment of system use with the clinical workflow. With the organizational implementation of the PACT model, this training was integrated into the PACT curriculum, which was then systematically offered to VA health care team members across the country.

Health care professionals report that the secure messaging system has been fairly well aligned with clinical workflow and implementation teams have invested time in structuring their triage teams and associated tasks to optimize this alignment. Although they suggest ways in which the system could be further integrated into the primary clinical information system, the alignments that do exist (such as the automatic notification of new and escalated messages and the ability to save secure messaging interactions in the VA EHR) facilitate use of the system. Health care professional's accounts provide evidence that secure messaging has also addressed some of the barriers that previously constrained use of My HealtheVet PHR features, for example, by enabling more timely self-reports from patients between clinical visits, supporting communication and feedback, and facilitating documentation updates in the VA EHR.

### *Reported Consequences of Secure Messaging Use*

Health care professionals report several consequences as a result of secure messaging use. These are summarized in Table 3. Health care professionals perceive that secure messaging improves patient access to the health care team and health care services, and makes it easier for health care team members to respond directly to patients. Secure messaging enables better connectivity between patients and members of their health care team and avoids some of the challenges encountered with telephone calls, such as phone tag. Patient perceptions about access are positively influenced, increasing patient's confidence that they can easily reach their health care provider when needed.

Health care professionals perceive that secure messaging changes both communication and the patterns of communication. Communication is more direct and focused, in contrast to telephone communication. Health care professionals report differences in the patterns of communication with secure messaging as a result of asynchronicity, including more frequent communication with patients between periodic in-person visits, and a lowering of the threshold at which patients will initiate communication with their health care team.

Asynchronicity is perceived as beneficial because it enables patients and health care professionals to send and respond to messages when it is convenient for them. This enables patients to communicate in their own time and be thoughtful about their

needs. For staff, asynchronicity enables them to respond to patient requests when they have time within their workflow and to give patient requests more focused attention. Health care professionals report that being able to save the interaction as a progress note in the VA EHR ensures needed documentation and also allows them to capture the patient's description in their own words.

With secure messaging, health care professionals report that more frequent communication enables them to keep track of what is going on with their patients between face-to-face visits, and they know their patients better as a result. Exchange of information before the clinical visit also enhances the quality of the visit. As 1 provider described, accomplishing administrative work in advance of the visit enables the health care provider to focus on the patient's agenda at the visit:

> *It gives you a conversation that you might not have otherwise had, except that you see them once every 7 to 8 months or 9 months or a year, you now have this interjected conversation piece that's going on that allows you to find out what their value system is, what their reasons are, what the barriers are, how is it that they're able to be successful with this piece or that piece. And then you can launch a change talk about other issues that may be the underlying root cause of why they've never been successful in the first place...So it's a different kind of information gathering journey I would say...I would say that it supports or strengthens the relationship: patient to provider. In the most simple sentence that I could provide I would say that it strengthens the relationship. It certainly builds trust...I would say that it affects them all to date in a very positive way...the face-to-face visits seem to have a better flow. I have the patient set an agenda when I first walk into the room rather than 'What are you here for?' I say 'What would YOU like to accomplish in this visit?' And if they've been secure messaging me, then we've taken care of a lot of their list that they want to take forward to the provider, and I usually know...and we launch from there in the direction that the patient really wants to travel.*

Another phenomenon that health care professionals describe is the impact of secure messaging on the threshold of communication for patients, facilitating improved communication with patients. Health care professionals' accounts provide evidence that the interrelated effects of these changes leads to improved relationships between patients and members of the health care team. Secure messaging is perceived to increase patient engagement, trust, and satisfaction:

> *I think people get to communicate without the intensity of a visit. They get to do it in their own time. They get to be more thoughtful...it's just a slam dunk...for them!...I think there's 2 things that have changed...1 is the care coordination. I had an email: 'I went to the Emergency Room. I had chest pains.' Well, you know, boy I'm going to make sure that person has an appointment...I'm going to ask the clinic facilitator to get the records...and so...so...the patient has made me a better doctor...because, what if they didn't let me know?...Well in addition to what I said about improving access, improving coordination...increasing my knowledge of the patients...between visits...and learning more about your patients...But that whole paradox of...even though it's a computer communication, I actually know the patients better...would they have called me and told me that? On the phone? I'm not sure! I think...I think patients are behaving differently because we've lowered the threshold to share the information. And so...and because of that, I get to be more diligent.*

Health care professionals also expressed concerns about workload, especially as use of the system increases. They consistently noted that workload so far has been manageable because use of the system has grown organically. They caution that as use of the system increases it will be important to continue to enhance the system, especially for integration with the primary clinical workflow system.

**Textbox 1.** Key findings and related themes.

1. My HealtheVet PHR features have been underutilized, with limited patient endorsement.

- Health care professionals report limited experiences with patient use (and their own use) of patient health education resources, tools to support medication reconciliation, and tools to enable patient tracking and self-reporting of data (with some notable exceptions)
- Endorsement of patient use has been limited

2. Several factors have inhibited the My HealtheVet PHR adoption, use, and endorsement of patient use.

- Lack of knowledge
- Lack of perceived relevance
- Perceived lack of relative advantage
- Time constraints
- Lack of alignment with workflow (eg, lack of integration with the primary clinical information system)
- Lack of alignment with structures (eg, lack of patient-accessible computers in the clinic setting)
- Lack of alignment with processes (eg, barriers to information flow)

3. In contrast, secure messaging has been more readily implemented, used, and endorsed by health care professionals.

- Health care professionals report successfully using secure messaging, and endorsing patient use
- Analysis of the trajectory of secure messaging implementation reflects spread and significant growth in use

4. Several factors have facilitated secure messaging adoption, use, and endorsement of patient use by health care professionals.

- Perceived relevance
- Perceived relative advantage
- Education and training opportunities
- Integration with the existing technology used to accomplish work tasks
- Alignment with workflow within the clinical setting
- Incentives that affect intended users (eg, performance measures)
- Access to information entered by patients
- Asynchronous, bidirectional communication for collaborative work

5. Secure messaging has had dramatic consequences for communication, patterns of communication, and patient/provider relationships.

- Improves access and patient perceptions of access
- Avoidance of telephone tag
- Communication is more direct and focused
- Improves convenience and efficiency
- More frequent communication between periodic in-person visits
- Lowers the threshold at which patients will initiate communication
- Improved patient engagement, satisfaction, and trust
- Enhances patient/provider relationships
- Concerns about workload implications with increased use

**Table 2.** Milestones in the history of secure messaging implementation.

| Date | Milestone |
|---|---|
| MAR 2004 | Workgroup established to develop strategy for secure messaging. |
| MAY 2006 | Workgroup initiates design and development of the secure messaging. |
| NOV 2007 | Secure messaging deployed at 3 early adopter sites for pilot testing. |
| DEC 2007 | Clinical workflow and triage process documents developed and distributed. |
| JAN 2008 | Three additional sites added to initial 3 early adopter sites. |
| JUN 2008 | Secure messaging application undergoes formal functionality testing. |
| SEP 2008 | National release of secure messaging application within the My HealtheVet portal. Secure messaging tab appears for authenticated VA patients. |
| OCT 2008 | Workload code approved and activated to capture workload credit. Encounter form developed to capture secure messaging progress note in the VA EHR. |
| DEC 2008 | Secure messaging in limited use at 12 facilities in 8 VA health care systems. Every network (VISN) is required to establish a local implementation team. |
| JAN 2009 | Clinical adoption toolkit released to field to support local implementation. |
| FEB 2009 | VA National Universal Task Force releases report recommending transformation initiatives including new models of care. |
| APR 2010 | VA initiates 3-year plan to implement Patient Aligned Care Teams (PACT) in more than 900 VA primary care clinics. More than 700 VA patients opted-in and actively using secure messaging with 136 triage groups. |
| JUL 2010 | My HealtheVet coordinator positions formalized with initiation of recruitment. |
| AUG 2010 | Secure messaging becomes part of the formal Operating Plan for New Models of Care (PACT). |
| SEP 2010 | VA National Leadership Board formalizes performance targets: use of secure messaging within primary care at a minimum of 1 medical center per VISN within 30 days, availability of secure messaging within primary care at all medical centers within 1 year (September 2011), 100% penetration of secure messaging in all primary care clinics by September 2012. |
| OCT 2010 | Annual national performance measures for fiscal year 2011 include 3 secure messaging–related goals (increase authentication, increase patients opted-in for secure messaging, increase number of sites offering secure messaging). Secure messaging enhancements released. |
| MAY 2011 | Secure messaging offered within primary care at all VA medical centers, meeting national target in advance of September 2011 deadline. |
| OCT 2011 | Annual national performance measures for fiscal year 2012 include 100% secure messaging penetration in primary care by March 2012, implementation within specialty and surgical care by September 2012, and aggressive targets for in-person authentication. |
| NOV 2011 | More than 60 facilities reach FY12 milestone goal of 100% secure messaging penetration rate in primary care in advance of September 2012 deadline. One VISN has 100% secure messaging penetration in primary care for all facilities in the VISN. More than 58,019 patients actively using secure messaging with 6613 triage groups. |

**Table 3.** Consequences of secure messaging.

| Theme | Description |
|---|---|
| Improving access and patient perceptions about access | Health care professionals report that secure messaging improves patient access and influences patient perceptions about access by enabling better connectivity with the health care team and avoiding some of the difficulties encountered with telephone calls. |
| More direct communication | Health care professionals report that secure messaging has enabled more direct communication by enabling patients to send questions directly to their health care team and allowing health care team members to respond directly to patient inquiries. |
| Changing communication patterns/asynchronicity | Health care professionals report that for many kinds of needs an asynchronous Secure message is a more effective way to support patient communication with the health care team. |
| Changing communication patterns/lowering the threshold | Health care professionals perceive that secure messaging lowers the threshold at which patients will initiate communication with their health care team. |
| Changing communication patterns/enhancing relationships | Health care professionals report that secure messaging has had a positive impact on patient/provider relationships. Health care professionals attribute this to the patient's perception of greater and more direct access to their health care team, the patient's perception of better responsiveness of the health care team to their needs leading to greater respect, trust, comfort, and appreciation, and increased frequency of communication. |
| Concerns about workload | Health care professional express some concerns about workload implications as use of the secure messaging system increases. |

## Discussion

Study findings revealed that 3 My HealtheVet features (patient health education resources, tools to support medication reconciliation and tools to enable patient tracking, and self-reporting of data) have been generally underutilized, whereas secure messaging has been successfully implemented and used by health care professionals. Findings revealed several factors that have facilitated or inhibited the adoption, use, and endorsement of patient use by health care professionals. Health care professional's accounts and analysis of organizational documents revealed a multidimensional dynamic between the trajectory of secure messaging and PACT model implementation and its impact on organizational actors and their use and endorsement of My HealtheVet. This dynamic has influenced workflow, work practices, communication, and the flow of information between patients and members of their health care team. In effect, secure messaging was the missing element of a complex information ecology and its implementation acted as a catalyst for change. Figure 1 illustrates the accelerated rate of growth in new My HealtheVet account registrations as secure messaging became more fully implemented. Secure messaging was also found to have dramatic consequences for communication, patterns of communication, and patient/provider relationships.

### Key Factors in the Implementation, Adoption, and Use of Technology

A comparison of the underutilized My HealtheVet features with use of secure messaging revealed 8 key factors that are important for the implementation, adoption, and use of a new technology in organizational settings (see Table 4).

### Perceived Relevance

Like other PHRs, My HealtheVet has generally been conceptualized as a set of tools for patients to utilize and as less relevant for health care professionals. Historically, promotional efforts have focused explicitly on patient use of the system with less attention to the potential relevance of these tools in the work of health care professionals. In contrast, secure messaging has been perceived by health care professionals as a tool that has significant relevance to their work. Promotional efforts have not only targeted patients, but have also focused on health care professionals to facilitate their adoption and use of the technology. As the organization has assimilated the PACT model, the relevance of secure messaging has been continually reinforced for professionals as an effective way to accomplish patient-centered care.

### Perceived Value

Study findings call attention to the question of value. Participants in this study recommended focusing on unique services that the My HealtheVet PHR offers, with great attention to secure messaging. For example, although My HealtheVet provides a significant library of patient health education resources, health care professionals already use alternative resources, with little incentive to change their practices. In contrast, secure messaging is perceived to offer specific advantages over existing alternatives. Many work tasks that health care professionals are responsible for require them to communicate with patients. Health care professionals report that secure messaging is more convenient than contacting patients by telephone, increases efficiency by avoiding telephone tag, and improves communication with patients by enabling increased communication between face-to-face visits. Each of these characteristics of secure messaging contributes to health care professional's perceptions about its value, both for themselves and for their patients.

**Figure 1.** New My HealtheVet account registrations by fiscal year.

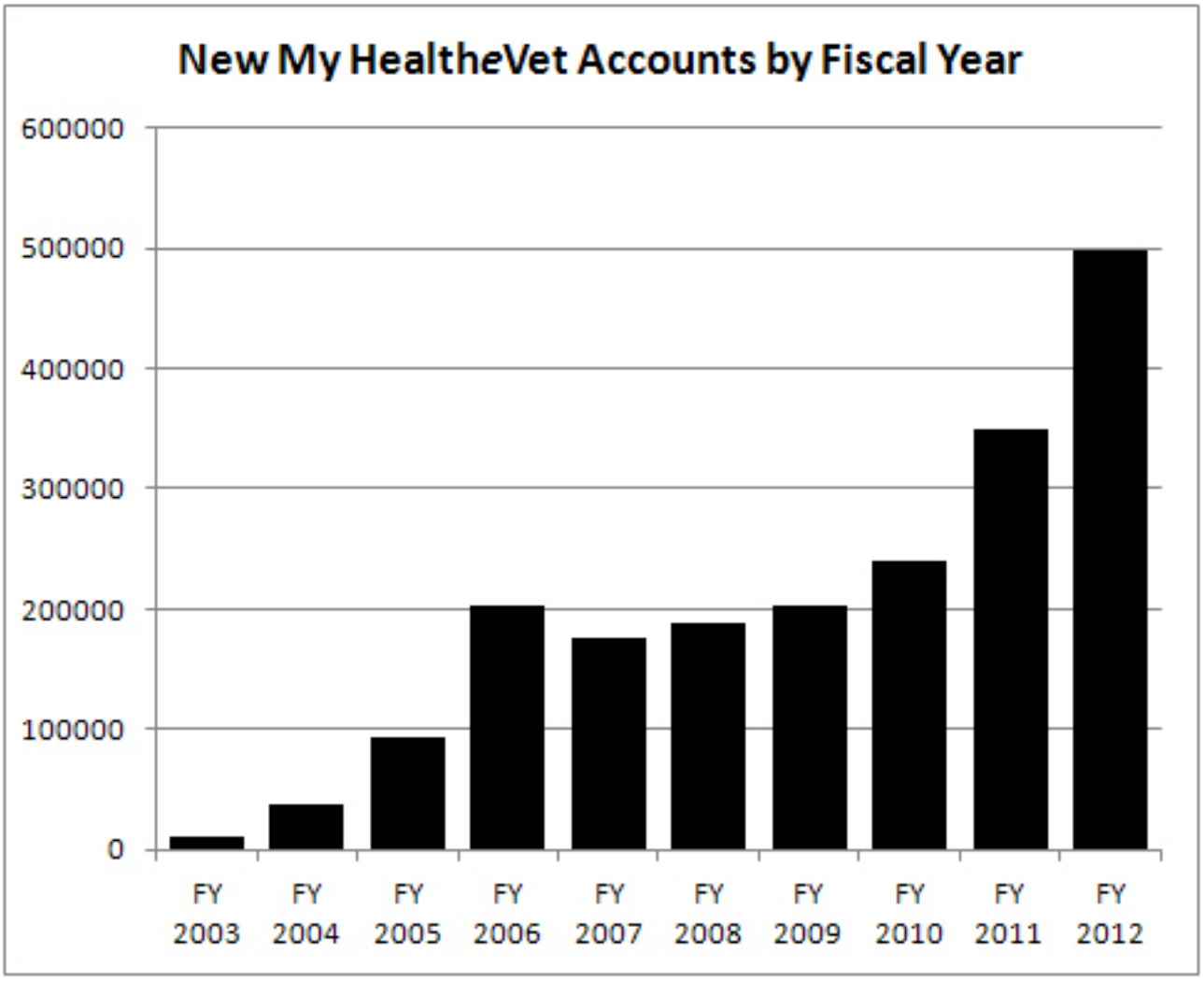



**Table 4.** Key factors in the implementation, adoption, and use of a new technology.

| Key factor | Description |
|---|---|
| Perceived relevance | In order to be adopted, the new technology must first be perceived by individuals as relevant to their work. |
| Perceived value | In order to be adopted and used, a new technology that has been deemed relevant must then be perceived as having greater value than the available alternatives for accomplishing work tasks. |
| Education and training | In order to be adopted and used, the new technology must be implemented with education and training opportunities targeted toward the intended user to ensure that they have the knowledge and skills needed to make effective use of the technology. |
| Integration with existing technology | In order to be adopted and used, the new technology must be integrated with the existing technology that is being used to accomplish work tasks. |
| Alignment with workflow | In order to be adopted and used, the new technology must be aligned with the workflow within the particular setting of use. |
| Incentives | If the implementation of a new technology is accompanied by incentives that affect intended users, the adoption and use of the technology will be facilitated. Incentives can operate at the organizational level or at the individual and/or team level. |
| Access to information | If the new technology is intended to support the accomplishment of work tasks that are dependent upon access to information entered by patients, it must enable health care professionals to have easy access to that information. |
| Communication | If the new technology is intended to support collaborative work tasks involving multiple participants, it must support asynchronous and bidirectional communication in order to be adopted and used effectively. |

### Education and Training

Since My HealtheVet was launched in 2003, education and training initiatives have primarily focused on veteran users. As a result, many health care professionals do not have adequate knowledge about specific features, and they have limited ability to make use of these features or to encourage and/or educate patients to make use of them. In contrast, secure messaging implementation has been accompanied by training and education programs for health care professionals, including opportunities for hands-on experiences. Training that was systematically being offered to VA health care team professionals across the country provided instruction on the use of the secure messaging system and also ways to align use of the system with the work tasks in a particular clinic, for example, by setting up the way that incoming secure messages are triaged and assigned for completion. In addition, health care professionals were also provided with ongoing opportunities to interact with other users, such as participation in weekly user-oriented conference calls. These forums help users to stay abreast of changes and updates to the system, and to learn from the experiences of their peers.

### Integration with Existing Technology

In the VA health care system, the technology used to accomplish and document the accomplishment of many work tasks is centralized within the VA EHR. The historical lack of integration of My HealtheVet tools with the VA EHR system has inhibited health care professionals' adoption and endorsement of these features. In contrast, secure messaging has been purposefully integrated within existing technology systems. The ability to save secure messaging interactions directly from the secure messaging system into the VA EHR as a clinical progress note connects use of the new technology with the existing technology system to accomplish work tasks and to document related information. In addition to the VA EHR, health care professionals emphasize the value of automatic notifications for new and escalated secure messages via the enterprise-wide email system; however, they also recommend that these alerts and notifications be more fully integrated directly into the VA EHR.

### Alignment With Workflow

In organizational settings, workflow represents a commonly understood set of procedures for and sequence of work tasks, along with the assignment of specific roles for individuals to accomplish these work tasks. Taken together, these comprise processes that organizations manage to accomplish work. In health care settings, clinical workflow is a description of how the work is done and by whom. If technology is intended to be used to enable the accomplishment of specific work tasks, alignment with the larger workflow is also needed for its use to be effective and efficient for the health care team as an organizational work unit.

In contrast to other My HealtheVet features, secure messaging has been designed and implemented in ways that purposefully align with clinical workflow. As part of the process of secure messaging implementation, health care teams have invested time in structuring their triage teams to process secure messages in ways that are aligned with the procedures for, and sequence of, work tasks. One strategy that was used effectively early in the implementation of the system was to conduct a simulation exercise in which teams were gathered around a table, given a piece of paper to represent a secure message, and instructed to pass the paper to experience the triage process and identify how it would work most effectively. This exercise often revealed that having a nurse serve as the triage person would enable many requests to be handled without further assignment. Within some clinical teams, health care providers have chosen to receive and respond to incoming messages directly. This ability of the system to support local adaptation makes it adjustable to the procedures for and sequence of work tasks, taking into account the individual preferences of health care team members within a particular clinic setting.

### Incentives

Incentives can operate at the organizational level or at the individual and/or team level. Within the VA system, as with many organizations, performance measures are established each fiscal year, and progress is measured and monitored closely via enterprise-wide reports. At the individual level, incentives can include remuneration for work efforts that can be either financial (eg, reimbursement for activity) or nonfinancial (eg, workload credit for activity). In the VA system, attribution of workload credit is seen as an important factor because it influences the number of patients empaneled to a particular provider. In addition, health care providers are employed by the agency, and organizational factors, such as performance measures, exert significant influence (eg, pay for performance), especially in comparison to settings in which health care providers are not employed. Although historically VA performance measures have been predominantly focused on clinical quality measures, the addition of measures related to technology use exemplifies the use of incentives at the organizational level to facilitate the adoption and use of the technology.

The My HealtheVet system was launched in 2003, but organizational incentives for increased patient adoption of the system were not formally instituted across the VA system until secure messaging was added to the system. As VA concurrently transformed primary care settings to the PACT model, secure messaging use became part of this organization-wide initiative, ultimately leading to the development of national mandates for staff use and national performance goals to incentivize increased patient use. Because My HealtheVet account registration and user authentication are requirements for patients to adopt secure messaging, national performance measures also targeted increased patient authentication in the My HealtheVet system. Incentives drive the prioritization of staff activities, the allocation of resources, and the continuous measurement and monitoring of progress. At the individual level, a workload code for secure messaging has been activated to enable workload credit for secure messaging activity. Implementation efforts for a new technology should address the facilitating effects of incentives to foster increased adoption and use of the technology.

### Access to Information

The inability of patients to share information from within the My HealtheVet system has made these tools less useful for

clinical care because information entered by patients is currently inaccessible to the health care team. In contrast, secure messaging has enabled patients to share information in a timely manner with their health care team. For example, management of diabetes often requires insulin titration based on a patient's blood glucose readings. By using secure messaging, patients have been able to share their blood glucose readings with their health care team in an efficient and timely way, enabling health care team members to complete titration without a face-to-face visit.

### Communication

Communication is a crucial requirement for accomplishing collaborative work tasks. Many work activities in health care are collaborative in nature; for example, the process of medication reconciliation requires direct interaction between patients and health care professionals. Bidirectional communication supports not only the conveyance of information, but also the interaction between participants that is needed for validation, clarification, feedback, and ultimately the accomplishment of collaborative work tasks. Secure messaging has enabled patients to provide timely updates about changes in their medication usage, to correct any observed inaccuracies or omissions in their VA prescription history, and to enable the interactive dialog that is needed to ensure understanding and feedback in between periodic face-to-face visits. In this way, secure messaging supports the accomplishment of the process of medication reconciliation. Members of the health care team can then document these interactions in the VA EHR and make updates to the medication list on record that is used to make decisions about clinical care and treatment regimens.

### From Diffusion to Assimilation and Routinization

As predicted by diffusion of innovations theory [64], study findings demonstrate that having adequate knowledge of the technology and its features is a prerequisite for adoption and assimilation. Given the prevalent lack of knowledge about My HealtheVet features among health care professionals and the perception of it as a tool solely for patients, little attention has been given historically to engaging patients about use of the tools in clinical settings. Other studies have found that health care professionals generally have limited knowledge about PHRs [65,66] and a relatively narrow view of PHR functions [67]. Extensions of diffusion of innovations theory also place strong emphasis on other organizational and social factors. Greenhalgh and colleagues [68] caution that the perceived attributes of technology are neither stable features nor sure determinants of adoption or assimilation. Instead, it is the interaction among the technology, actor, and a particular context that determines adoption and use. As initial adoption evolves into assimilation and routinization, organizational and social factors are increasingly influential factors, especially for complex process-based innovations. Often the unit of assimilation in organizations is the team or department, exemplified in this study by the secure messaging triage team. For these reasons, they emphasize that although the standard attributes of diffusion of innovations theory are important and relevant, they are insufficient by themselves to explain the adoption and assimilation of complex innovations in organizations. Other factors, such as social and organizational influences, must also be examined.

### Implications for the Evolution of Personal Health Records

The growing body of literature about PHRs and secure messaging is beginning to demonstrate that to be most effective for patients and their health care providers, PHRs should be combined with Web-based messaging tools to support information sharing and bidirectional communication. Several studies have begun to emphasize the important role of communication with patients as they make use of PHRs [69-71]. As Terry notes [72], "a PHR that doesn't connect to your doctor is like an ATM without any money in it." Systems that provide patients with access to their laboratory test results, for example, should also anticipate the need for additional communication by providing patients with the ability to ask questions about their results to "close the loop." These findings have important implications for the design and use of PHRs and PHR systems.

### Implications for Practitioners, Organizations, and Researchers

Study findings have important implications for individual practitioners, organizations, and researchers. These implications should be considered at the system level, the organizational level, and the individual level [48]. Findings in this study provide evidence that at the system level, the integration of secure messaging into the PACT model as a new model of care institutionalized use of the technology with a systemic implementation program, incentivizing performance measures, and an enterprise shift to patient-centered, team-based care. At the organizational level, integration within clinical practice settings has improved access, provided operational efficiencies, and enabled alignment of the technology within the clinical workflow. At the individual level, health care professionals report that secure messaging has improved patient engagement, and enhanced the relationships that patients have with their health care team. These findings suggest that secure messaging has promising potential to improve health care delivery by complementing traditional methods of communication. These findings also illustrate the complexity of implementing technology in health care organizations, and the need to examine the implementation and use of technology at multiple intersecting levels.

### The Personal Health Record Paradox: Looking Beyond Commonly Reported Barriers

Much of the literature about PHRs to date has been focused on an accounting of PHR features and consumer perspectives about their use. Although some authors speculate about the potential impact of use on medicine and the patient-provider relationship, less attention has been placed on understanding the actual experiences and perspectives of health care professionals with respect to use of PHRs and PHR systems. Given the persistent paradox between reported patient interest in PHRs and anticipation of benefits with relatively low adoption, this study examined PHR use within a particular organizational ecosystem as a component of health care work. Although the literature

generally highlights concerns about privacy and security as prominent barriers to the adoption and use of PHRs, this study provided a unique opportunity to look beyond these commonly reported issues to enable a deeper understanding of how patient PHR use may unfold within the context of the health care interaction and impact the provision of services by health care professionals in an organizational setting.

The VA health care system is an opportune environment to study the situated use of PHR features because veteran users express confidence in the system [9], health care professionals have embedded experiences working in an electronically mediated environment [73], and health care professionals are directly employed by the system. As a result, study findings highlight 4 implications for health care systems beyond the commonly reported barriers of privacy and security. First, health care professionals play a crucial role in the endorsement of PHRs to patients, and in subsequent engagement with patient use of PHR tools. Second, in order for health care professionals to adopt and use PHRs effectively, there are several key factors that must be present, including adequate education and training opportunities. Third, for technology to be effective in supporting and improving health care delivery, careful attention must be paid to align use of the technology with health care processes including incumbent work activities, information flow, and bidirectional communication when the process requires collaborative work. Fourth, increased patient use of PHRs and secure messaging may have significant workload implications for health care professionals that will need to be addressed. These implications raise important issues for organizations seeking to use technology such as PHRs to improve patient care.

### Clinician Endorsement and Engagement

Study findings reveal the important influence of clinician endorsement and engagement in patient use of PHR tools. Although PHRs are designed as consumer-oriented tools intended to engage and empower patients, study findings suggest that engagement must be a reciprocal process. This reciprocity has been represented in the chronic care model as productive interactions between the “informed activated patient” and the “prepared proactive practice team” [74]. Similarly, the Care Transitions Intervention model highlights clinician engagement as an important component of effective PHR use [75], emphasizing the role of the health care provider and other members of the health care team in fostering effective patient use. Based on a national survey of physicians about PHRs, Wynia and colleagues [66] concluded that to derive optimal benefit from using PHRs, patients and physicians should use these tools together as partners. Dunbrack [76] similarly emphasized that clinician endorsement is a primary motivator for patients to use a PHR. This assessment also predicts that the spread of patient-centered medical home models (such as VA’s PACT model) are likely to encourage clinicians to recommend PHRs to their patients to better manage their health and wellness.

Wynia and Dunn [42] caution that patients and providers have mixed views about PHRs that are not yet informed by direct experience. In a recent random national survey of US physicians, they found that 62% of physician respondents reported no previous experience with using electronic PHRs, although 42% said that they were willing to try using one with their patients. Similarly, consumer’s knowledge about PHRs continues to be limited, with more than half of consumers surveyed in February 2011 reporting that they were not familiar with the concept of a PHR [76]. Interestingly, 3 out of 4 consumers reported that they would start to use a PHR under certain circumstances, with 37% indicating that a clinician recommendation to use a PHR would be a primary motivator for them. The implications for health care systems are that better engagement of health care professionals may be needed to fully realize many of the broadly anticipated potential benefits of patient PHR use.

### Education and Training for Health Care Professionals

To engage clinicians and foster the adoption and use of PHR features, greater knowledge about PHRs and familiarity with their features is clearly needed. Study findings emphasize the importance of developing educational activities and promotional efforts aimed at health care professionals, taking into account their learning preferences and time constraints in the work environment. A great deal can be learned from implementation science about the most efficacious ways to meet this important need for education and training [77]. Further research is needed to develop and test interventions that will improve knowledge about a PHR system and its features. As the system is enhanced, it will also be important to include the ongoing dissemination of new information to health care professionals as part of an overall communication plan, and local champions may play an important role in this. Several health care professionals in this study were unaware of important changes that had been made to the system. The framework to support the implementation of secure messaging proposed by Wakefield et al [78] highlights these aspects of implementation as key factors.

### Health Care Processes, Information Flow, and Bidirectional Communication

Study findings suggest that to further examine how PHR tools can be integrated into health care work, a more holistic examination of the processes that embody work tasks and associated information flow is needed. For processes that require collaborative work, bidirectional communication is crucial. For example, an examination of the process of medication reconciliation revealed that sequential work tasks managed by health care professionals include a review of the medical list on record, interactive dialog with the patient to validate and identify changes including updates and amendments, and updating the list on record as an authoritative source. Simply providing tools for patients to document their medication information in their PHR was ineffective until this task was aligned with the larger process. Alignment with the process of medication reconciliation required asynchronous electronic communication that enabled the timely flow of information between patients and members of their health care team via bidirectional communication.

Findings from this study highlight the critical nature of information flow and bidirectional communication with 4 related considerations. PHR systems should enable patients to share information effectively with their health care team via tools such as delegation or the ability to authorize the addition of patient self-entered data to the official medical record [79].

Bidirectional communication tools should be implemented in tandem with PHRs to support the interactive dialog that optimal use of PHRs requires [23]. Integration with clinical workflow is a crucial determinant of use for health care professionals [80,81]. As use of PHRs increases, careful attention must be taken to address the potential for information overload [42], the need for complete and accurate information [82], and the potential for unintended consequences [83].

In this study, the organizational changes that occurred made it possible to witness the catalyzing impact of secure messaging related to information flow and bidirectional communication. Beyond the VA system, early standalone PHR models that lacked the capacity for information sharing and 2-way communication have given way to patient portals that include the ability to share information and support for electronic communication between patients and their health care team. Models, such as Kaiser Permanente's My Health Manager, that integrate PHRs or patient access to EHR data with communication via secure messaging are more effective because they address these needs. Nijland et al [70] caution us to apply technology in ways that foresee the patient's need for continuous and personalized feedback, especially for patients who have a greater need for care. These may be the same patients who are most highly motivated to utilize a PHR because of their need to manage a plethora of information related to their condition and/or their care.

### Implications Related to Workload

Although health care professionals have had concerns about the impact of secure messaging on workload, study findings provide evidence that workload to date has been manageable. Similarly, other studies have emphasized clinician's concerns about additional workload [5,84-86], whereas several studies have found these concerns to be unfounded [87-90]. Even if increased workload is balanced by workflow efficiency, as some studies suggest, remuneration of time devoted to these activities may be important, whether through financial reimbursement [91] or workload credit for panel management.

### Study Limitations

This study has several limitations. By design, this study focused on health care professionals within the VA as an integrated system. The degree to which findings are generalizable to other organizations and systems is an area that warrants further study. Study participants varied significantly in their role or position within the larger organizational structure, for example, whereas some health care professionals had a prominent role in the national scope of the system, others were in a rural community-based clinic. This presents challenges in characterizing "the organization" for individuals. Additional research is needed to that more closely examines role conception and organizational structure as contextualizing factors.

Another limitation is related to the degree to which study participants have practical experiences with use of the system. Although the aim of the study was to understand experiences with 4 specific features of the PHR, the scope of the study was inherently limited by the lack of adoption for some My HealtheVet features. Even with this constraint, however, a great deal was revealed about facilitating and inhibiting factors. The temporal changes in the organization also necessitating adding to the scope of the study as it became important to trace the trajectory of secure messaging implementation when it became evident that its availability had an important influence on other things.

Lastly, the potential for study participants to perceive the investigator as an advocate of the program could influence their willingness to report negative accounts. This potential bias was minimized by recruitment strategies that targeted health care professionals from across the country providing direct patient care in the field rather than relying on known sources. All study participants were also reminded at the beginning of their interview to be candid, and were assured that their perspectives would be reported with anonymity. The willingness of study participants to voice both positive and negative experiences instills confidence in the credibility of individual accounts. In-depth interviews were crucial in going beyond initial assumptions based on the background questionnaire. Member checking entailed dissemination of a summary of key findings to all participants in the study, inviting participants to clarify, elaborate, or amend. Feedback from participants further validated these findings. Final interpretation of specific findings should bear these potential sources of bias in mind.

### Areas for Future Research

Findings from this study point to a number of areas for future research. These areas can be generally categorized into 5 domains as shown in Table 5. These areas further expand the My HealtheVet PHR research agenda [10] and other calls for additional research about PHRs and their use [92]. Significant progress is being made within VA via a collaborative partnership with the VA eHealth Quality Enhancement Research Initiative (QUERI) Center established in 2010 [93], and a number of studies about My HealtheVet are currently underway. Although lack of evidence about the distinct effects of PHRs and other eHealth tools on health and other outcomes persists [94], this research will be more feasible as actual use increases.

**Table 5.** Areas for additional research.

| Domain | Description |
|---|---|
| Adoption | Further identify facilitators and inhibitors to adoption and use at multiple levels (system, organizational, individual) taking into account the various roles of health care professionals |
| Implementation | Develop approaches grounded in implementation science to measure the efficacy of implementation strategies |
| Education | Design and test interventions that will improve health care professionals' knowledge and familiarity with the system and its features |
| Information flow | Model information flow and map to health care processes and activities across the patient trajectory to identify optimal ways to apply technology |
| Communication | Apply communication theory to further examine the nuances of asynchronous electronic communication |

## Conclusions

The impetus for this study was the desire to deconstruct the PHR paradox: despite significant consumer interest and anticipated benefits, adoption and use of PHRs remains low, with some notable exceptions. After a decade of PHRs being promoted as independent tools to support a broad notion of consumer empowerment, a deeper understanding of these tools in the context of the social and organizational delivery of health care is needed to understand this paradox. Although there is anecdotal evidence that PHRs can improve health care, many of the benefits are presumptive and require further evaluation. Indeed, endorsement and engagement of health care professionals may be essential to fully realize the anticipated benefits of PHRs. As study findings demonstrate, patient use of PHRs also has broad implications for health care professionals and organizational delivery systems.

Although PHRs have been designed as consumer-oriented tools, health care professionals play a crucial role in the endorsement of PHRs to patients, and in subsequent engagement with patients' use of these systems. Patient engagement may be most effective as a reciprocal process. In order for health care professionals to adopt and support patient use of PHRs effectively, there are several key factors that must be present, including adequate education and training opportunities. In addition, for technology to be effective in supporting and improving health care delivery, careful attention must be paid to align use of the technology with health care processes and clinical workflow including incumbent work activities, information flow, and bidirectional communication when processes require collaborative work. These implications raise important issues for organizations seeking to support technologies such as PHRs to improve patient care.

Viewing PHR systems as an information ecology highlights the dynamic interplay among technologies, people, practices, and values [95]. This interplay, however, also occurs in the context of an ecosystem in which organizational and social factors influence technology use. Changes to the ecosystem, exemplified in this study by the implementation of secure messaging, occur along a trajectory with the adoption and use of technology followed by assimilation and routinization. Changes to the flow of information, exemplified in this study by the addition of secure messaging to the My HealtheVet portal, effectively alter the dynamic, recursively changing the ecosystem as a result.

Leveraging technology in new and transformative ways that are most meaningful for patients and health care professionals will require a more holistic approach to better understand the social and organizational context of technology use in clinical practice settings. Given the institutional context of most health care service delivery models, the broader notion of the organization setting as an ecosystem warrants further attention.

## Acknowledgments

The author thanks each of the VA health care professionals who participated in the study. The views expressed in this article are those of the author and do not necessarily represent the view of the Department of Veterans Affairs.

## Conflicts of Interest

None declared.

## References

1. Endsley S, Kibbe DC, Linares A, Colorafi K. An introduction to personal health records. Fam Pract Manag 2006 May;13(5):57-62 [FREE Full text] [Medline: 16736906]
2. Ferreira A, Correia A, Silva A, Corte A, Pinto A, Saavedra A, et al. Why facilitate patient access to medical records. Stud Health Technol Inform 2007;127:77-90. [Medline: 17901601]
3. Kim MI, Johnson KB. Personal health records: evaluation of functionality and utility. J Am Med Inform Assoc 2002 Apr;9(2):171-180 [FREE Full text] [Medline: 11861632]
4. National Committee on Vital and Health Statistics. Personal Health Records and Personal Health Record Systems: A Report and Recommendations from the National Committee on Vital and Health Statistics. Washington, DC: US Department of

Health and Human Services; 2006 Feb. URL: http://www.ncvhs.hhs.gov/0602nhiirpt.pdf [accessed 2008-02-20] [WebCite Cache ID 5VlINiXs3]
5. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006 Apr;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]
6. Archer N, Fevrier-Thomas U, Lokker C, McKibbon KA, Straus SE. Personal health records: a scoping review. J Am Med Inform Assoc 2011 Aug;18(4):515-522 [FREE Full text] [doi: 10.1136/amiajnl-2011-000105] [Medline: 21672914]
7. Detmer D, Bloomrosen M, Raymond B, Tang P. Integrated personal health records: transformative tools for consumer-centric care. BMC Med Inform Decis Mak 2008;8:45 [FREE Full text] [doi: 10.1186/1472-6947-8-45] [Medline: 18837999]
8. Mandl KD, Kohane IS. Tectonic shifts in the health information economy. N Engl J Med 2008 Apr 17;358(16):1732-1737. [doi: 10.1056/NEJMsb0800220] [Medline: 18420506]
9. Nazi KM. Veterans' voices: use of the American Customer Satisfaction Index (ACSI) Survey to identify My HealtheVet personal health record users' characteristics, needs, and preferences. J Am Med Inform Assoc 2010 Apr;17(2):203-211 [FREE Full text] [doi: 10.1136/jamia.2009.000240] [Medline: 20190065]
10. Nazi KM, Hogan TP, Wagner TH, McInnes DK, Smith BM, Haggstrom D, et al. Embracing a health services research perspective on personal health records: lessons learned from the VA My HealtheVet system. J Gen Intern Med 2010 Jan;25 Suppl 1:62-67 [FREE Full text] [doi: 10.1007/s11606-009-1114-6] [Medline: 20077154]
11. Pagliari C, Detmer D, Singleton P. Potential of electronic personal health records. BMJ 2007 Aug 18;335(7615):330-333 [FREE Full text] [doi: 10.1136/bmj.39279.482963.AD] [Medline: 17703042]
12. Halamka JD, Mandl KD, Tang PC. Early experiences with personal health records. J Am Med Inform Assoc 2008 Feb;15(1):1-7 [FREE Full text] [doi: 10.1197/jamia.M2562] [Medline: 17947615]
13. Kaelber D, Pan EC. The value of personal health record (PHR) systems. AMIA Annu Symp Proc 2008:343-347 [FREE Full text] [Medline: 18999276]
14. Ball MJ, Smith C, Bakalar RS. Personal health records: empowering consumers. J Healthc Inf Manag 2007;21(1):76-86. [Medline: 17299929]
15. Beun JG. Electronic healthcare record; a way to empower the patient. Int J Med Inform 2003 Mar;69(2-3):191-196. [Medline: 12810123]
16. Markle Foundation. Survey finds Americans want electronic personal health information to improve own health care. 2006 Nov 1. URL: http://www.markle.org/publications/1214-survey-finds-americans-want-electronic-personal-health-information-improve-own-hea [accessed 2012-11-12] [WebCite Cache ID 6C8C7CPUJ]
17. Markle Foundation. Americans overwhelmingly believe electronic personal health records could improve their health. 2008 Jun 01. URL: http://www.connectingforhealth.org/resources/ResearchBrief-200806.pdf [accessed 2009-05-03] [WebCite Cache ID 5gVcBRcqe]
18. Harris Interactive. Few patients use of have access to online services for communication with their doctors, but most would like to. 2006 Sep 22. URL: http://www.harrisinteractive.com/news/allnewsbydate.asp?NewsID=1096 [accessed 2012-11-12] [WebCite Cache ID 6C8Ceu9vc]
19. Health Industry Insights. Health Industry Insights Consumer Survey. 2006 May. URL: http://www.idc.com/downloads/HIIConsumersurveyePHRs_Q&A.pdf [accessed 2012-01-01] [WebCite Cache ID 6C8D9jZKi]
20. Deloitte Center for Health Solutions. 2009 Survey of health care consumers: key findings, strategic implications. 2009 Mar. URL: http://www.deloitte.com/assets/Dcom-UnitedStates/Local%20Assets/Documents/us_chs_2009SurveyHealthConsumers_March2009.pdf [accessed 2012-11-12] [WebCite Cache ID 6C8FiEwm7]
21. Noblin AM, Wan TH, Fottler M. The impact of health literacy on a patient's decision to adopt a personal health record. Perspect Health Inf Manag 2012;9:1-13 [FREE Full text] [Medline: 23209454]
22. Cronin C. AARP Public Policy Institute. Washington, DC: AARP; 2006 Apr. Personal health records: an overview of what is available to the public URL: http://assets.aarp.org/rgcenter/health/2006_11_phr.pdf [accessed 2012-11-12] [WebCite Cache ID 6C8G8yfAp]
23. Gearon C. California HealthCare Foundation. 2007 Jun. Perspectives on the future of personal health records URL: http://www.chcf.org/documents/chronicdisease/PHRPerspectives.pdf [accessed 2009-05-03] [WebCite Cache ID 5gVcBRcoV]
24. Reinke T. Rise of the PHR. Manag Care 2007 Jan;16(1):40-42. [Medline: 17285811]
25. Saparova D. Motivating, influencing, and persuading patients through personal health records: a scoping review. Perspect Health Inf Manag 2012 Apr 01;9:1f [FREE Full text] [Medline: 22783154]
26. Chilmark Research. Labs, PHRs, Platforms & Consumer Engagement: A Presentation. 2009 Mar 18. URL: http://chilmarkresearch.com/2009/03/18/labs-phrs-platforms-consumer-engagement-a-presentation/ [accessed 2012-11-13] [WebCite Cache ID 6C9Y8MVpQ]
27. Manhattan Research. Analyzing digital health by the numbers. 2009 Oct. URL: http://www.connected-health.org/media/262830/bard%20keynote%2010.21.09.pdf [accessed 2012-11-13] [WebCite Cache ID 6C9YLekWK]

28. California HealthCare Foundation. Consumers and Health Information Technology: A National Survey. 2010 Apr. URL: http://www.chcf.org/publications/2010/04/consumers-and-health-information-technology-a-national-survey [accessed 2010-06-05] [WebCite Cache ID 5qGWhzzFu]
29. Markle Foundation. The Public and Doctors Largely Agree Patients Should Be Able To View, Download and Share Their Health Info. 2011 Jan 31. URL: http://www.markle.org/publications/1460-public-and-doctors-largely-agree-patients-should-be-able-view-download-and-share-t [accessed 2012-11-13] [WebCite Cache ID 6C9Yyujs3]
30. Kaiser Permanente. 2012. News Center: Kaiser Permanente HealthConnect Electronic Health Record URL: http://xnet.kp.org/newscenter/aboutkp/healthconnect/faqs.html [accessed 2013-01-27] [WebCite Cache ID 6DzS4QrLa]
31. Tsai J, Rosenheck RA. Use of the internet and an online personal health record system by US veterans: comparison of Veterans Affairs mental health service users and other veterans nationally. J Am Med Inform Assoc 2012;19(6):1089-1094. [doi: 10.1136/amiajnl-2012-000971] [Medline: 22847305]
32. Fotsch E. AHIC Chronic Care Workgroup Testimony. 2006 Aug 16. Secure messaging and online services to enhance care and health URL: http://www.slidefinder.net/f/fotsch/fotsch/1323447 [accessed 2012-11-13] [WebCite Cache ID 6C9ZNgdMj]
33. Jimison H, Gorman P, Woods S, Nygren P, Walker M, Norris S, et al. Barriers and drivers of health information technology use for the elderly, chronically ill, and underserved. Evid Rep Technol Assess (Full Rep) 2008 Nov(175):1-1422. [Medline: 19408968]
34. Baur C. Limiting factors on the transformative powers of e-mail in patient-physician relationships: a critical analysis. Health Commun 2000;12(3):239-259. [doi: 10.1207/S15327027HC1203_02] [Medline: 10938915]
35. Ball MJ, Lillis J. E-health: transforming the physician/patient relationship. Int J Med Inform 2001 Apr;61(1):1-10. [Medline: 11248599]
36. Hagland M. Patient to partner. Will PHRs change the physician-patient relationship? J AHIMA 2005 Dec;76(10):38-40. [Medline: 16333942]
37. Scherger JE. Future vision: is family medicine ready for patient-directed care? Fam Med 2009 Apr;41(4):285-288 [FREE Full text] [Medline: 19343560]
38. Stone JH. Communication between physicians and patients in the era of E-medicine. N Engl J Med 2007 Jun 14;356(24):2451-2454. [doi: 10.1056/NEJMp068198] [Medline: 17568026]
39. Cushman R, Froomkin AM, Cava A, Abril P, Goodman KW. Ethical, legal and social issues for personal health records and applications. J Biomed Inform 2010 Oct;43(5 Suppl):S51-S55. [doi: 10.1016/j.jbi.2010.05.003] [Medline: 20937485]
40. Greeg C, Buck SL. PHRs and physician practices. J AHIMA 2007 Apr;78(4):71-75. [Medline: 17455852]
41. Hoffman S, Podgurski A. E-health hazards: provider liability and electronic health record systems. Berkeley Technology Law Journal 2010 Jul 06;24(4):1524 [FREE Full text]
42. Wynia M, Dunn K. Dreams and nightmares: practical and ethical issues for patients and physicians using personal health records. J Law Med Ethics 2010;38(1):64-73. [doi: 10.1111/j.1748-720X.2010.00467.x] [Medline: 20446985]
43. Berner ES, Moss J. Informatics challenges for the impending patient information explosion. J Am Med Inform Assoc 2005 Jul 27;12(6):614-617 [FREE Full text] [doi: 10.1197/jamia.M1873] [Medline: 16049224]
44. Greenhalgh T, Swinglehurst D. Studying technology use as social practice: the untapped potential of ethnography. BMC Med 2011;9:45 [FREE Full text] [doi: 10.1186/1741-7015-9-45] [Medline: 21521535]
45. Brown SH, Lincoln MJ, Groen PJ, Kolodner RM. VistA--U.S. Department of Veterans Affairs national-scale HIS. Int J Med Inform 2003 Mar;69(2-3):135-156. [Medline: 12810119]
46. Nazi KM, Hogan TP, McInnes DK, Woods SS, Graham G. Evaluating Patient Access to Electronic Health Records. Medical Care 2013 2013;51:S52-S56. [doi: 10.1097/MLR.0b013e31827808db]
47. United States Department of Veterans Affairs. My HealtheVet URL: http://www.myhealth.va.gov [accessed 2013-03-11] [WebCite Cache ID 6F2piwpXk]
48. Hogan TP, Wakefield B, Nazi KM, Houston TK, Weaver FM. Promoting access through complementary eHealth technologies: recommendations for VA's Home Telehealth and personal health record programs. J Gen Intern Med 2011 Nov;26 Suppl 2:628-635 [FREE Full text] [doi: 10.1007/s11606-011-1765-y] [Medline: 21989614]
49. Naditz A. Telemedicine at the VA: VistA, MyHealtheVet, and other VA programs. Telemed J E Health 2008 May;14(4):330-332. [Medline: 18584775]
50. Schneider J. My HealtheVet: fighting for health with information. Journal of Consumer Health On the Internet 2008 Jan;12(1):13-21. [doi: 10.1080/15398280802081410]
51. US Department of Veterans Affairs. VA Blue Button URL: http://www.va.gov/bluebutton [accessed 2012-11-14] [WebCite Cache ID 6CB2cRpUx]
52. Klein S. The Commonwealth Fund. 2011 Sep. The Veterans Health Administration: implementing patient-centered medical homes in the nation's largest integrated delivery system URL: http://www.commonwealthfund.org/Publications/Case-Studies/2011/Sep/VA-Medical-Homes.aspx [accessed 2012-11-14] [WebCite Cache ID 6CB2Upy6C]
53. Corbin J, Strauss A. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Thousand Oaks, CA: Sage Publications, Inc; 2008.

54. Green J, Thorogood N. In: Green DG, Thorogood DN, editors. Qualitative Methods for Health Research (Introducing Qualitative Methods series). Thousand Oaks, CA: Sage Publications Ltd; 2009.
55. Kaplan B, Maxwell JA. Qualitative research methods for evaluating computer information systems. In: Anderson JG, Aydin CE, Jay SJ, editors. Evaluating Health Care Information Systems: Methods and Applications. Thousand Oaks, CA: Sage Publications; 1994:45-68.
56. Neuman WL. Social Research Methods: Qualitative and Quantitative Approaches. Boston: Pearson/Allyn and Bacon; 2006.
57. Patton MQ. Qualitative Research and Evaluation Methods. Thousand Oaks, CA: Sage Publications; 2002.
58. Auerbach CF, Silverstein LB. Qualitative Data: An Introduction to Coding and Analysis. New York: New York University Press; 2003.
59. Emerson RM, Fretz RI, Shaw LL. Writing Ethnographic Fieldnotes. Chicago: University of Chicago Press; 1995.
60. Rubin H. Qualitative Interviewing: The Art of Hearing Data. Thousand Oaks: Sage Publications, Inc; 2005.
61. Glaser BG, Strauss AL. The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago: Aldine Pub Co; 1967.
62. Cohen DJ, Crabtree BF. Evaluative criteria for qualitative research in health care: controversies and recommendations. Ann Fam Med 2008 Aug;6(4):331-339 [FREE Full text] [doi: 10.1370/afm.818] [Medline: 18626033]
63. Lincoln YS, Guba EG. Naturalistic Inquiry. Beverly Hills, CA: Sage Publications; 1985.
64. Rogers E. Diffusion of Innovations. New York: Free Press; 2003.
65. Fortney JC, Burgess JF, Bosworth HB, Booth BM, Kaboli PJ. A re-conceptualization of access for 21st century healthcare. J Gen Intern Med 2011 Nov;26 Suppl 2:639-647 [FREE Full text] [doi: 10.1007/s11606-011-1806-6] [Medline: 21989616]
66. Wynia MK, Torres GW, Lemieux J. Many physicians are willing to use patients' electronic personal health records, but doctors differ by location, gender, and practice. Health Aff (Millwood) 2011 Feb;30(2):266-273. [doi: 10.1377/hlthaff.2010.0342] [Medline: 21289348]
67. Witry MJ, Doucette WR, Daly JM, Levy BT, Chrischilles EA. Family physician perceptions of personal health records. Perspect Health Inf Manag 2010;7:1d [FREE Full text] [Medline: 20697465]
68. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
69. Green BB, Cook AJ, Ralston JD, Fishman PA, Catz SL, Carlson J, et al. Effectiveness of home blood pressure monitoring, Web communication, and pharmacist care on hypertension control: a randomized controlled trial. JAMA 2008 Jun 25;299(24):2857-2867 [FREE Full text] [doi: 10.1001/jama.299.24.2857] [Medline: 18577730]
70. Nijland N, van Gemert-Pijnen JE, Kelders SM, Brandenburg BJ, Seydel ER. Factors influencing the use of a Web-based application for supporting the self-care of patients with type 2 diabetes: a longitudinal study. J Med Internet Res 2011;13(3):e71 [FREE Full text] [doi: 10.2196/jmir.1603] [Medline: 21959968]
71. Kahn JS, Hilton JF, Van Nunnery T, Leasure S, Bryant KM, Hare CB, et al. Personal health records in a public hospital: experience at the HIV/AIDS clinic at San Francisco General Hospital. J Am Med Inform Assoc 2010 Apr;17(2):224-228 [FREE Full text] [doi: 10.1136/jamia.2009.000315] [Medline: 20190069]
72. Terry K. Will PHRs rule the waves or roll out with the tide? Hosp Health Netw 2008 Aug;82(8):36-39. [Medline: 19062435]
73. Kizer KW, Dudley RA. Extreme makeover: Transformation of the veterans health care system. Annu Rev Public Health 2009;30:313-339. [doi: 10.1146/annurev.publhealth.29.020907.090940] [Medline: 19296778]
74. 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]
75. Coleman EA, Parry C, Chalmers S, Min SJ. The care transitions intervention: results of a randomized controlled trial. Arch Intern Med 2006 Sep 25;166(17):1822-1828. [doi: 10.1001/archinte.166.17.1822] [Medline: 17000937]
76. Dunbrack L. IDC Health Insights. 2011. Vendor assessment: when will PHR platforms gain consumer acceptance URL: http://www.idc-hi.com/getdoc.jsp?containerId=HI227550 [accessed 2013-03-11] [WebCite Cache ID 6F2svsn3Q]
77. Kitson AL, Rycroft-Malone J, Harvey G, McCormack B, Seers K, Titchen A. Evaluating the successful implementation of evidence into practice using the PARiHS framework: theoretical and practical challenges. Implement Sci 2008;3:1 [FREE Full text] [doi: 10.1186/1748-5908-3-1] [Medline: 18179688]
78. Wakefield DS, Mehr D, Keplinger L, Canfield S, Gopidi R, Wakefield BJ, et al. Issues and questions to consider in implementing secure electronic patient-provider web portal communications systems. Int J Med Inform 2010 Jul;79(7):469-477. [doi: 10.1016/j.ijmedinf.2010.04.005] [Medline: 20472495]
79. Reti SR, Feldman HJ, Ross SE, Safran C. Improving personal health records for patient-centered care. J Am Med Inform Assoc 2010 Apr;17(2):192-195 [FREE Full text] [doi: 10.1136/jamia.2009.000927] [Medline: 20190063]
80. Berg M. Patient care information systems and health care work: a sociotechnical approach. Int J Med Inform 1999 Aug;55(2):87-101. [Medline: 10530825]
81. Berg M. Implementing information systems in health care organizations: myths and challenges. Int J Med Inform 2001 Dec;64(2-3):143-156. [Medline: 11734382]
82. 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. [doi: 10.1056/NEJMp0801736] [Medline: 18420496]

83. Harrison MI, Koppel R, Bar-Lev S. Unintended consequences of information technologies in health care--an interactive sociotechnical analysis. J Am Med Inform Assoc 2007 Oct;14(5):542-549 [FREE Full text] [doi: 10.1197/jamia.M2384] [Medline: 17600093]
84. Kittler AF, Carlson GL, Harris C, Lippincott M, Pizziferri L, Volk LA, et al. Primary care physician attitudes towards using a secure web-based portal designed to facilitate electronic communication with patients. Inform Prim Care 2004;12(3):129-138. [Medline: 15606985]
85. Katz SJ, Moyer CA. The emerging role of online communication between patients and their providers. J Gen Intern Med 2004 Sep;19(9):978-983 [FREE Full text] [doi: 10.1111/j.1525-1497.2004.30432.x] [Medline: 15333064]
86. Patt MR, Houston TK, Jenckes MW, Sands DZ, Ford DE. Doctors who are using e-mail with their patients: a qualitative exploration. J Med Internet Res 2003 Jun;5(2):e9 [FREE Full text] [doi: 10.2196/jmir.5.2.e9] [Medline: 12857665]
87. Liederman EM, Morefield CS. Web messaging: a new tool for patient-physician communication. J Am Med Inform Assoc 2003 Jun;10(3):260-270 [FREE Full text] [doi: 10.1197/jamia.M1259] [Medline: 12626378]
88. White CB, Moyer CA, Stern DT, Katz SJ. A content analysis of e-mail communication between patients and their providers: patients get the message. J Am Med Inform Assoc 2004 Aug;11(4):260-267 [FREE Full text] [doi: 10.1197/jamia.M1445] [Medline: 15064295]
89. Zhou YY, Garrido T, Chin HL, Wiesenthal AM, Liang LL. Patient access to an electronic health record with secure messaging: impact on primary care utilization. Am J Manag Care 2007 Jul;13(7):418-424 [FREE Full text] [Medline: 17620037]
90. Zhou YY, Kanter MH, Wang JJ, Garrido T. Improved quality at Kaiser Permanente through e-mail between physicians and patients. Health Aff (Millwood) 2010 Jul;29(7):1370-1375. [doi: 10.1377/hlthaff.2010.0048] [Medline: 20606190]
91. Tang PC, Black W, Young CY. Proposed criteria for reimbursing eVisits: content analysis of secure patient messages in a personal health record system. AMIA Annu Symp Proc 2006:764-768 [FREE Full text] [Medline: 17238444]
92. Kaelber DC, Jha AK, Johnston D, Middleton B, Bates DW. A research agenda for personal health records (PHRs). J Am Med Inform Assoc 2008 Dec;15(6):729-736 [FREE Full text] [doi: 10.1197/jamia.M2547] [Medline: 18756002]
93. United States Department of Veterans Affairs. Quality Enhancement Research Initiative (QUERI): eHealth URL: http://www.queri.research.va.gov/ehealth/default.cfm [accessed 2012-11-14] [WebCite Cache ID 6CBDlhJUW]
94. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
95. Nardi BA, O'Day VL. Information Ecologies: Using Technology with Heart. Cambridge, MA: MIT Press; 1999.


## Abbreviations

**EHR:** electronic health record
**IRB:** Institutional Review Board
**PACT:** Patient Aligned Care Team
**PHR:** personal health record
**QUERI:** Quality Enhancement Research Initiative
**VA:** Veterans Affairs
**VHA:** Veterans Health Administration
**VISN:** Veteran Integrated Service Network


*Edited by G Eysenbach; submitted 15.11.12; peer-reviewed by T Eytan; comments to author 22.01.13; revised version received 30.01.13; accepted 31.01.13; published 04.04.13.*

*Please cite as:*
*Nazi KM*
*The Personal Health Record Paradox: Health Care Professionals' Perspectives and the Information Ecology of Personal Health Record Systems in Organizational and Clinical Settings*
*J Med Internet Res 2013;15(4):e70*
*URL: http://www.jmir.org/2013/4/e70/*
*doi:10.2196/jmir.2443*
*PMID:23557596*

Letter to the Editor

# A Systematic Self-Certification Model for Mobile Medical Apps

Thomas Lorchan Lewis[1], BSc (Hons)

Warwick Medical School, University of Warwick, Coventry, United Kingdom

**Corresponding Author:**
Thomas Lorchan Lewis, BSc (Hons)
Warwick Medical School
University of Warwick
Gibbet Hill Road
Coventry, CV4 7AL
United Kingdom
Phone: 44 7876453511
Fax: 44 2476528375
Email: t.lewis@warwick.ac.uk

**Related Articles:**

Comment on: http://www.jmir.org/2012/5/e128/

Letter: http://www.jmir.org/2013/12/e277/

***(J Med Internet Res 2013;15(4):e89)*** doi:10.2196/jmir.2446

**KEYWORDS**

smartphone; technology; education; medicine; app; health care; Android; iPhone; BlackBerry; mobile phone; standards

Errol Ozdalga and colleagues recently highlighted the impressive range of roles and uses of smartphones in the medical setting [1]. An important point highlighted and worth developing from this paper is the difficulties associated with accurately comparing and assessing different medical apps for smartphones. This is mainly due to the fact that medical apps are often designed with one particular focus and inherently different interfaces which often make a direct comparison between apps unfeasible. Furthermore, even apps that purport to complete the same task often include extra functionality or features that make direct evaluation impossible. One solution offered by Ozdalga et al is to survey doctors on the perceived impact of specific apps available. However I believe that this is unfeasible given the rate at which the medical app ecosystem is evolving in terms of number, range, and type of app. With thousands of medical apps available, it is highly improbable that a clinician has a working knowledge of the complete range available. As such, any surveys will be subjective depending on the target audience and consequently offer limited utility for physicians and medical students alike. Moreover, surveys regarding specific apps are usually out of date by the time they are published. What is more important, is establishing a systematic method by which medical apps can be compared and their utility for health care professionals validated.

One proposed method to solve this is to develop a set of standard criteria that can be used to systematically assess the utility of a medical app for a health care professional. I believe that the most efficient and effective method should be based on a self-certification system with key criteria that have been adapted from the Health on the Net foundation (HON, [2]). Table 1 shows potential self-certification criteria which medical apps could be reasonably expected to achieve in order to establish the validity of the information contained within the app. The Health on the Net Foundation Code of Conduct (HONcode) for medical and health websites addresses one of Internet's main health care issues: the reliability and credibility of information. It is therefore highly applicable to medical apps that are subject to the same issues.

Using this system, it would then possible to set up a self-certification process where registered developers could highlight the fact that their app conforms to these basic criteria. At the moment, no such organization exists although there is clearly scope for such an entity. With the impending launch of the United Kingdom National Health Service App Store, it appears that there has never been a better time to develop a self-certification model for medical apps.

**Table 1.** A list of potential criteria based on the HONcode to be used as the basis of a self-certification model for medical apps.

| Certification criteria | Detailed description |
|---|---|
| Information must be authoritative | • All medical information presented in a medical app must be attributed to an author and his/her training in the field must be mentioned. |
| Purpose of the website | • A statement clearly declaring that the information on the app is not meant to replace the advice of a health professional has to be provided.<br>• A brief description of the app's mission, purpose, and intended audience is necessary.<br>• Another brief description of the organization behind the app, its mission, and its purpose is also necessary. |
| Confidentiality | • This principle is applicable to all apps, even if it does not host patient records or store any medical or personal data.<br>• The app must describe a privacy policy regarding how confidential, private or semi-private information such as email addresses and the content of emails received from or sent to users is treated.<br>• Users must be informed whether their data will be recorded in your own database, who can access this database (others, only you, nobody), if this information is used for your own statistics (anonymous or not), or if these statistics are used by third party or other companies.<br>• Even if one or more of these points are not relevant to your app, you must state how you handle the following information sent to you by your visitors: (email addresses or/and contact information, names, personal, or medical data). |
| Information must be documented: referenced and dated | • All medical content (page or article) has to have a specific date of creation and a last modification date.<br>• All sources of the medical content must be clearly indicated the recognized, scientific, or official sources of health information quoted in the app. Ideally, a precise link to the source is provided whenever it is possible. |
| Justification of claims | • All information about the benefits or performance of any treatment (medical and/or surgical), commercial product, or service is considered as claims. All claims have to be backed up with scientific evidence (medical journals, reports, or others). |
| Contact details | • The app must be completely operational and the information must be accessible and clearly presented.<br>• There must be a way to contact the developer, such as a working email address or contact form, for users who would like to have more details or support. This contact must be easy to access from anywhere within the app. |
| Financial disclosure | • Each app must include a statement declaring its sources of funding. This is required for all apps, including those with no external sources of funds, and apps funded by government agencies, pharmaceutical companies, or other commercial entities.<br>• All funding must be declared: government agency, private companies, donations, etc.<br>• Developers also have to declare all conflicts of interest. |
| Advertising policy | • Conflicts of interest and external influences which could affect the objectivity of the editorial content must be clearly stated in a disclaimer.<br>• All apps displaying paying banners have to have an advertising policy. This policy must explain how the app distinguishes between editorial and advertising content and which advertisements are accepted. Any conflict of interest has to be explained. |

## Acknowledgments

The author of this article is an editor at iMedicalApps.com, an online publication focused on mobile health. This article received no sources of funding or sponsorship and there is no financial disclosure. There were no human subjects involved therefore it was not deemed necessary to seek ethical approval.

## Conflicts of Interest

None declared.

## References

1. Ozdalga E, Ozdalga A, Ahuja N. The smartphone in medicine: a review of current and potential use among physicians and students. J Med Internet Res 2012;14(5):e128 [FREE Full text] [doi: 10.2196/jmir.1994] [Medline: 23017375]
2. HONcode Operational definition of the HONcode principles. 2012 Nov 21. URL: http://www.hon.ch/HONcode/Webmasters/Guidelines/guidelines.html [accessed 2012-11-21] [WebCite Cache ID 6CLYzYQ4s]

*Edited by G Eysenbach; submitted 21.11.12; peer-reviewed by E Ozdalga; accepted 31.01.13; published 24.04.13.*

*Please cite as:*
*Lewis TL*
*A Systematic Self-Certification Model for Mobile Medical Apps*
*J Med Internet Res 2013;15(4):e89*
*URL: http://www.jmir.org/2013/4/e89/*
*doi:10.2196/jmir.2446*
*PMID:23615332*

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org